

CM20314 - Experimental Systems Project

Deliverable 4: Analytic and Empirical Evaluation Specification

Dr John Benardis and Dr Moody Alam

<u>Set:</u> Monday 20/02/2023 (week 21)
<u>Due:</u> Friday 05/05/2023 (week 31), 8pm
<u>Percentage of overall unit mark:</u> 30%
<u>Submission Location:</u> Group submission via Moodle
<u>Submission Components:</u> Report, Ethics documentation
<u>Files:</u> PDF named 'Group # - Analytic and Empirical Evaluation' PDF named 'Group # - Experiment Ethics Documentation'
<u>Anonymous Marking:</u> No

1 Overview

For CM20314 your overall project involves the proposal, specification, design, implementation, testing and evaluation of a novel interactive system. This deliverable focuses on the evaluation of your proposed solution in terms of usability as well as involves conducting an experiment on an important, observable aspect of your system (that is feasible to collect).

2 The Report

The deliverable is a group report, written in groups of 7-9. It must have between 4500 – 6000 words including figure captions – this is a hard limit (appendices are not to be used to extend this limit etc.). Tables of content, figures, diagrams, and references do not count towards the word limit. It must use the ACM Primary Article Template which is a single column format (use Word or LaTeX for editing; submission must be PDF):

<https://www.acm.org/publications/taps/word-template-workflow>

You should remove CCS Concepts, Additional Keywords and Phrases, and the ACM Reference format sections.

In addition to the report, the group must also submit a Group Contribution Form (see below) which does not count towards the word limit.

Structure: The following structure is mandatory. The report will be marked according to the criteria below. The report is worth 30% of the overall unit mark. Every point listed below must be adequately and concisely covered in the report.

Iterative Stakeholder Engagement – about 1 page

- Report the user/stakeholder involvement during the different stages of this project leading to your evaluation stage.
- It should cover stakeholder engagement (if applicable to your project) that was used to:
 - define and specify the problem and its proposed solution system.
 - specify requirements and offer formative feedback on architecture design.
 - translate requirements into system interactions and features during the design and prototyping phases.
 - iterate between different prototype versions (e.g. from low-fidelity sketches to higher fidelity digital prototypes).
 - inform and refine your testing and implementation approaches.

[This section can borrow from previous deliverables in addition to any further stakeholder engagement that was performed after them. Do not simply copy and paste, you should be aiming to illustrate how stakeholder involvement in your project defined, shaped, and refined it via iterative testing and evaluation]

Analytic Evaluation – about 3 pages

- Conduct a critical Heuristic Evaluation of your system's interface with an appropriate set of heuristics. Include a justification for choosing specific heuristics set.
- Conduct a Cognitive Walkthrough of a key task the user is expected to carry out. (You may want to refer to your use cases to determine important goal-oriented user actions or you may even want to perform new task analyses to identify complex tasks that would be suitable to illustrate this method)
- In both cases include a description of the methodology, the evaluation procedure, and the evaluation outcomes.

[Analytic evaluation is especially useful if performed before the empirical evaluation sessions since it could identify issues with the usability of the system that if not resolved could confound data collection towards the utility and usefulness of your system.]

Empirical Evaluation – about 6 pages

- Present evidence of conducting an experiment with your system. This should include:
 - Presenting the experiment/evaluation aim.
 - Illustrating your measurement concept and its suitability for your purposes¹.
 - Formulating a clear, testable hypothesis.
 - Providing information on the participants (you must use at least 10 participants)
 - Concisely describing the experiment design using scientific terminology.
 - Concisely describing the metrics¹ that will be used along with a justification of their adequacy.
 - Concisely describing the experiment procedure.
- Present your experiment data/results. This should include:
 - A summary table of quantitative results following scientific conventions
 - A figure illustrating quantitative results following scientific conventions
 - Statistical analysis of quantitative results to test your hypothesis, following scientific conventions. Report results using APA style²
 - A concise description of qualitative results
- Discuss your results. This should include:
 - An interpretation/explanation of the results (i.e. why do you think you got these results and not different ones)
 - An argument to support the interpretation/explanation based on your quantitative and qualitative results.
 - Limitations of your study (e.g. what might have negatively affected the validity of your study)
 - Brief discussion of impact of your results (e.g. implications for redesigning some aspect of your system, applications and implications for users and designers)

Further guidance on the experiment: As part of this deliverable, you will examine one observable aspect of your system as a means of testing a hypothesis. This will involve contrasting two versions of your system based on an independent variable of your choosing (e.g. two alternate

¹ Measurement v. metric: *Measurement* is the process of observing and recording the observations that are collected as part of a research effort. Measurement in research consists of assigning numbers to empirical events, objects or properties, or activities in compliance with a set of rules. On the other hand, a *metric* is a quantifiable measure that is used to track and assess the status of a specific process.

² e.g. see here <https://about.illinoisstate.edu/jhkahn/apastats/>

user interfaces, two alternate sorting algorithms). Your contrast will be accomplished by comparing observations you make of each version and performing a statistical analysis.

For the purpose of this coursework, empirical testing requires you to quantize your observations. That means defining metrics to represent an observable attribute that you believe is important for the effectiveness of your system as a whole and that is feasible for you to collect. Metrics are often a compromise based on the availability of data in a software system's environment of operation. Not all environments make the same data available. Comparing metrics should allow you to make an unbiased decision on the relative merits of one version versus another version of your system. You must state your measurement concept as well as the metrics in your analysis¹, with a critical discussion of both the adequacy of the metrics and suitability of the measurement concept for your purposes. The following example should help to illustrate what this means:

Example: Testing usability of user interfaces. You might assess user efficiency (a measurement concept) in carrying out specified tasks with your system (e.g. obtaining a summary of data for a particular day) by counting the number of operations they perform (a metric) or number of seconds taken (an alternative metric for the same measurement concept) with interface version one v. version two.

You are expected to make defensible claims about the quality of your software system or process supported by the evidence produced by your empirical testing. That means you must make a clear argument for the meaning of the measurements you have decided upon in relation to the stakeholder viewpoints you have identified.

The marking criteria for the deliverable are as follows:

Criteria	Percentage of Presentation component
Iterative Stakeholder Engagement	10%
Analytic Evaluation	30%
Heuristic Evaluation	15%
Cognitive Walkthrough	15%
Empirical Evaluation	60%
Experiment Design	25%
Experiment Data/Results	20%
Discussion	15%
Ethics Documentation	-9% max loss
Information Sheet	-3% if missing or -1.5% if not appropriate
Consent Form	-3% if missing or -1.5% if not appropriate
Checklist for Full Sub-Committee Review	-3% if missing or -1.5% if not appropriate

3 Ethics Documentation

When you perform research as an undergraduate student, it is essential that you consider the ethical implications of your work, and that you work according to established processes that manage:

- risk of harm to investigators, research participants, or the wider community
- privacy, and its ethical and legal basis (e.g. GDPR)

In most cases ethical approval of a study will depend on appropriate materials and processing to ensure:

- Informed consent of participants is properly obtained, recorded, and stored.
- Participant data (including a record of consent) is properly managed, e.g., anonymised and securely stored.
- Participants are debriefed on request.

For the purposes of this deliverable you need to prepare and submit a collated PDF file that must include:

[**Note:** Templates for all ethics documentation required are available on Moodle].

1. Information or 'briefing' sheet for participants.

This provides the details required for a participant to provide informed consent to participate in the study. Participants need to be given enough information to decide about whether they wish to participate. What is the study about, what is involved, what risks are there etc.

2. Consent Form for the study.

This is the form that participants must sign, or otherwise explicitly approve before they begin the study. All participants must be informed that they have the right to withdraw at any stage, and likewise to request that any data is deleted.

3. A completed Checklist for Full University Ethics Sub-Committee Review

This is to ensure that all necessary considerations have been made with regards to human participation in your experiment. You are as well reminded of the following in planning your experiment and completing your checklist:

- Participation is voluntary - no-one should be forced/coerced to take part;
- Informed consent - participants should understand what the study or interview/questionnaire is about and what their data will be used for. They should explicitly give consent to take part (can be verbal doesn't have to be written);
- Do not use participants who are vulnerable or unable to give informed consent (e.g. children under the age of 16 years, people with learning difficulties, patients, people receiving counselling, people living in care or nursing homes, people recruited through

self-help groups);

- No deceiving participants;
- No personal data - there should be no reason for students to be collecting or processing personal data for the purposes of the experiment. This includes no voice/video recordings;
- No sensitive data - students shouldn't be asking people about sensitive topics (e.g. sexual activity, drug use, mental health);
- Anonymisation - based on responses students may need to anonymise so that participants aren't identifiable;
- Only collect what you need – be mindful of participants' time.

4. A debriefing sheet [optional]

Which provides brief additional information about the study, perhaps its hypotheses or findings, to be made available to participants after the study on request.

The required ethics documentation is not explicitly awarded a mark percentage for completion however failure to consider and submit the mandatory items will result to a deduction of marks as listed in the criteria breakdown.

Compliance with the current UK Data Protection Legislation

It is highly unlikely that you will need to collect personal contact details of participants - and this should only be done so if necessary although it is not encouraged for the experiment. However, if you need to keep a list of personal contact details for participants:

- a) they need to be kept securely for at least 10 years. A password-access folder on the x-drive is secure but if any details are kept on a personal device, then this needs to be encrypted as well (the library can help with this)
- b) the participants need to give explicit consent for their contact details to be kept, for what purpose, for how long and who will have access to them. They also need to be told who to contact if they want their details removed at any point in the future. They have to take an active step to do this, i.e. completing the original survey or whatever can no longer be counted as implied consent. They must be given all the information about where, what and why etc. their information is being kept and then click something or sign something to say they're happy for that to be the case.

4 Group Contribution Form

In addition to your report, please submit a separate PDF with a Group Contribution Form as described in a separate document “**CM20314 - Group Contributions and the Group Contribution Form (GCF)**”.

5 Academic Integrity

Your work will be checked to ensure that you have not plagiarised. For more information about the plagiarism policy at the University see: <https://library.bath.ac.uk/referencing/plagiarism>

Remember that published work that you refer to in your report should be clearly referenced in your text and listed in a bibliography section given at the end of your report. For more information see, <https://library.bath.ac.uk/referencing/new-to-referencing>

6 Feedback

Formative feedback on your work will be offered throughout the duration of the coursework:

- During your tutorials the tutors will be available to answer questions and offer guidance. Please note that tutors will not be able to make decisions on behalf of your group about the course of the project. They are there to discuss your ideas and offer advice.
- Use Moodle forums to post general questions or questions specific to your project. The unit convener will respond to these as well as your peers. This way we will create a repository of knowledge that will be available to all.

You will receive **summative feedback** on your work within 3 semester weeks of the submission deadline. The feedback will discuss your performance based on the criteria for marking, including what you did well and how specific sections could have been improved.