Principles for Term Paper Grading

CSCI 6606: Human Factors in On-Line Information Systems

02 January 2012

1 Basic Expectations

- length (your essay must not be much shorter or longer than the specified length)
- format (must be appropriate for an academic essay, and conform the specification in the syllabus)
- structure (similar to the articles studied in class)
- spelling & grammar

Violations of any of these expectations will result in deductions or outright failure.

Use Of References

Your essay should be self-contained. References to any external sources should be for additional information and sources of evidence rather than being necessary for the reader to understand the essay.

Plagiarism will not be tolerated. You must give proper credit when referring to or using work by others. If you require guidance about the mechanics of citation and what constitutes plagiarism, consult with the Computer Science librarian, or staff at the Killam Library reference desk.

2 Content

According to Dalhousie's undergraduate calendar [1]:

> **$\mathcal{A}$-level grades** reflect 'considerable evidence of original thinking; demonstrated outstanding capacity to analyze and synthesize; outstanding grasp of subject matter; evidence of extensive knowledge base.'
>
> **$\mathcal{B}$-level grades** reflect 'evidence of grasp of subject matter, some evidence of critical capacity and analytical ability; reasonable understanding of relevant issues; evidence of familiarity with the literature.'

I evaluate your essay for the demonstrated understanding of underlying issues and terminology.

In particular, the most important aspects of your essay that I evaluate when grading content are:

- quality of research (how well you selected appropriate high-quality materials by others);
- quality of analysis (your contribution in making sense of the material you have selected);
- demonstrated understanding of underlying issues and terminology; and
- style (how the essay is organized, and how well the text flows).

Reference

[1] Dalhousie University. Section 17.1 Grade Scale and Definitions In Undergraduate calendar. ⟨URL: http://www.registrar.dal.ca/calendar/ugrad/acaregs.htm⟩. Downloaded 14 Dec. 2002.

[2] Mary-Claire van Leunen. *A Handbook for Scholars* (Revised edition). Oxford University Press, 1992. ISBN 0-19-506953-5 (cloth) / 0-19-506954-4 (paper).

A Examples* of Grade Levels

$\mathcal{A}^+$ Discussion of a research problem with a plan of how to use the research

- begins with an introduction to the area and explanation of why the issues are important
- all relevant technical terms are clearly defined without interrupting the flow of the essay
- comprehensive survey (within reasonably defined limits) of articles within an area
- a proposal that incorporates the research into a plan of action (e.g. an experiment, a new procedure, a strategy for the redesign of some procedure or procedures)
- if the use is to be in a research study or experiment then
  - a statement of hypotheses justified by evidence in the survey
  - plan for an experiment to test the hypotheses including descriptions of dependent and independent variables, experimental design, and planned statistical tests (see Appendix D, on page 4)
- concluding summary that serves as a short, complete précis of the essay

$\mathcal{A}$ Comprehensive survey intended to lead to an experiment or study

- clear structure of the essay follows from the goals in the introduction
- survey is comprehensive within reasonably defined limits
- insights throughout
- conclusion that summarizes the essay and points the way for future work and outstanding issues

$\mathcal{B}^+$ Insightful and purposeful survey

- survey beginning with motivation for study
- examination of relevant articles ordered by issue and with a clear, rational connection between sections
- insights throughout
- conclusion that summaries key points

$\mathcal{B}^-$ Topic-based survey with little coherence

- survey of articles within an area, with one section about each article
- little or no clear connection between sections
- includes introduction about importance of topic and motivation for its study
- concluding summary compares and contrasts the various articles

$\mathcal{C}$ Survey with no summary

- same as $\mathcal{B}^-$ but without concluding summary
- does not violate basic expectations

*These are examples only. This is neither an exhaustive nor a prescriptive list of ways to achieve these grade levels.

B Common Problem: Insufficient Justification

The notes in this appendix are intended to help you to prepare better essays and theses. Most of the students whose essays I have read have no problem writing interesting things, although a few exaggerate beyond the evidence they present. You must express interesting things in your essay, however the reader should never have reason to doubt the truth of anything you claim.

B.1 Special Considerations for Experiments

If you are writing up an experiment, remember that your are describing an investigation, not attempting to obtain any specific results. In the case of experiments, you must justify that the experiment is (or was) worth conducting, and that the results will be (or are) credible.

B.2 Justifying Claims — Avoiding unsupported assertions

Following standard scholarly practice, you must justify all of the factual claims in your essays, and be careful not to exaggerate. The strongest way is to provide an unassailable argument based on evidence. Since there is not enough time to conduct an experiment during this course you will need to rely on the four other general ways of justifying your claims, which are detailed in the sections immediately below.

B.2.1 Reference to Credible Sources

The strongest way is to provide evidence (either the results of experiments that you are presenting in the essay itself, or by reference to work by others). Any reference to work by others must include a formal citation to the complete record in the References section at the end of your essay.

Credibility of Sources Your readers will need the full reference to determine the credibility of the evidence. While refereed works (articles, certain books, etc.) are usually considered to be high quality sources, unrefereed works (such as technical reports, personal websites, Wikipedia, etc.) have little or no credibility. If your reader does not recognize the author's name, then they will likely look to see where the work appeared as a way to judge its credibility. Works appearing in ACM Transactions or from major academic presses (the Oxford Univ., Yale Univ., and MIT Presses for example) are much more credible than those appearing from low-quality journals, technical magazines, and unrefereed workshops.

Some readers often depend on only the citation to judge the credibility of the evidence you provide. However the **real purposes of the citation** are to enable readers to (a) verify that what your cited actually appears where you said it did, and (b) to find out more about the topic [2].

B.2.2 Expert Opinion

Where evidence is not available, you can cite the opinion (of one or more established experts, or yourself*). Such opinions are easiest to believe if some of the reasons for them are described too. In a thesis or published article you must provide a full citation for all non-print sources, e.g. telephone conversations. The CS Librarian can show you how to cite such 'personal communication'. The most important aspect for these notes is that you provide a reason for the reader to understand why you are citing someone else's opinion. For example: 'Experimental participants were placed into one of two groups depending on their ABC score. Following the accepted practice in $\overline{\text{(a relevant field)}}$ [cite an experiment published in a respected venue†], partitioning was based on the median score.'

B.2.3 Pure Reason

Sometimes it is appropriate or provide reasoning without hard evidence. For example: 'The groups were partitioned into three groups (referred to as "low", "middle', and "high") based on their XYZ scores. Scores in the two extreme groups ("low" and "high") differed by at least one standard deviation but there was some overlap between each of those groups and the "middle" group, as can be seen in Fig. X.'

B.2.4 Appeal to Common Knowledge

Finally, the weakest justification (which nonetheless may be sufficient) is your own impression of generally held views. You can write something like 'As it is well known that the sky is usually blue, . . . '.

*Even if you are an expert about what you are discussing, you must justify your reputation in the essay, otherwise you are begging the question of your expertise: 'I am an expert because I wrote this essay which made claims based on my opinions as an expert'.

†*Venues* are where articles are published, e.g. journals.

C Occasional Problem: Lack of Foundation

When proposing a technology (tool, system, etc.) you must explicitly address the following three issues:

1. What are the relevant characteristics of the target user population?
2. What goals, needs, or necessary tasks will your proposed technology be in aid of?
3. What measures of success will you apply, e.g. measures of satisfaction, speed, accuracy, ease of learning? Specific measures are better than generalities but you must always tie your measurement instruments to the fundamental concepts that you are measuring. See §D.3 for a lengthier description of the factors in the proposal for your technology.

D What Is A 'Plan For An Experiment'?

We will have studied several articles describing experiments before your term paper is due. In addition to the lessons you should have learned from such articles, here are some specific points that should be explicit parts of your plan for an experiment.

D.1 Motivation & Expected Importance

Always remember that an experiment is an investigation and never an attempt to prove anything. You must explain what you expect to learn from the experiment and why that information will be useful or interesting.

D.2 Hypotheses

Clearly state what you are planning to investigate with your experiment. A simple example could be that your new interface will be more effective than another interface.

Traditionally, hypotheses come in pairs: the *null hypothesis* which states that there is no difference between the values you are measuring, and the *alternative hypothesis*, (which states that there is a difference). The alternative hypothesis can be one-tailed or two-tailed. *One-tailed hypotheses* are directional (the result of the treatment will be strictly greater than or less-than the control condition). The control is a condition in which everything is exactly the same as in the treatment except that the factors are missing. For example: if one of the factors is that users are predicted to improve with practice, then a control condition could be that you measure before and after using your system / software /whatever. The measurements before use are a control for the practice factor in the post-use condition. It can be appropriate to control for several factors with separate conditions.

The values you measure are described as Factors in §D.3.

D.3 Factors

List and describe the independent and dependent variables.

Independent variables are the factors that you hypothesize as causes. Sometimes these can be measured directly (such as time to complete a task), but sometimes they must be operationalized as dependent variables instead. For example you might expect that users' understanding of a topic will be improved by using the interface you propose to create. However, you cannot measure understanding directly, so you measure something that depends on understanding: how much users' scores in comprehension tests improve after using your interface, for instance.

You must account for all factors that could influence your experiment so that you can design adequate controls. For instance, if you use a repeated-measures design, then you will need to control for order effects, even if you do not expect them to be significant.

D.4 Statistical Investigation

- What descriptive statistics will you present?
- What statistical tests will you perform?
- What standards will you use to decide whether to accept or reject the null hypothesis? (In particular, is a one- or two-tailed test most appropriate?)

D.5 Possible Implications

A well-planned experiment can often make use of so-called negative (or unexpected) results as well as results which confirm the alternative hypothesis. You must foresee the implications of both positive and negative results as part of the design of your experiment. Include a brief description of what accepting or rejecting each of your alternative hypotheses would mean for further experiments and for practice, if it is appropriate to your experiment.

E Copyright