Leading questions to help raise expectations about your BSc or MSc thesis

Prof. dr. Maarten Kleinhans, www.geo.uu.nl/fg/mkleinhans

Why do you want to read this thoroughly, and re-read it later?

The thesis you are about to write is the masterpiece of your Bachelor or Master of Science. Perhaps science can be summarised as follows:

- science revolves about curiosity, that is, asking questions
- science requires a positive critical attitude, that is, asking questions about answers without turning into a total sceptic, and abducing alternative, testable hypotheses
- science requires empirical research and the skills and experience to do investigations that test hypotheses, and to communicate the results with other scientists and with laymen

It is therefore quite essential that you master the basic skills of research, but perhaps even more important that you learn to communicate your results and ideas in writing and presentation. The masterpiece of your training is the final thesis and this is seen by your supervisor and by potential employers as a good, perhaps the best, indicator of your intelligence, skills and level of training.

Of course I could have written a simple checklist of hints for the thesis on the basis of my experience in supervision, but I don't want you to learn 'the trick', because there is no such thing. Rather, I want you to learn to think. This requires a process of conscious learning, that is, wherein you are conscious of what it is you are learning so that you learn it better. In this document I will therefore ask you a lot of questions that are intended to make you think and guide you through the process of writing a thesis.

This document may be a bit coloured by my field of research and my style of writing, but then you are not forbidden to look for inspiration elsewhere too. It may also come across as somewhat sarcastic and pedantic, but that is not directed at you. Just remember that I care about education as much as I care about science, because neither can live long and prosper without the other.

To boldly go where many, many others have gone before

What are the requirements for your thesis set by the BSc or MSc program? What is the time frame? How can you remind yourself of these during the process of writing?

Where can you find out what 'sort of' thesis is considered excellent by your supervisor?

When you come across a word or phrase in English that you do not understand in a research paper, a website, this document or an email from your supervisor, what (online) tools are at your disposal to find out the meaning, the double meaning, the synonyms and antenyms?

To twist the lion's tail

In earth science we have a few basically different approaches to unravel the past and the workings of the planet at our disposal: field data collection, experimentation, analytical and numerical modelling. As Lord Bacon once remarked, ideally one would like to twist the lion's tail, to intervene in the real world, in order to study the behaviour. However, that is rather dangerous, hence the lion. So instead of experimenting with the Earth we usually collect data on processes that are going on anyway, but we also do controlled laboratory experiments and modelling (Kleinhans et al. 2010). What are the basic benefits and problems of the approach you are employing, compared to all others? How must these problems affect your work, in particular the discussion and conclusions?

Structure out of chaos

Scientific research is a human enterprise and a difficult one (see http://en.wikipedia.org/wiki/What_Is_This_Thing_Called_Science%3F) so the process from question to answer is far from straightforward. However, a simplistic summary of the process, the empirical cycle, may help to structure your endeveours, and the same structure can be found in scientific publications. This model of how science works has been rather influential in the philosophy of science (Hempel & Oppenheim 1948). The empirical cycle and the structure of a thesis can be as follows:

Empirical cycle	**Chapter in paper / thesis**	**How to approach?**
Question	Introduction	See section below on questions. A literature review should make clear to what extent the question has been addressed by others, rather than summing up everything that is somehow connected to your topic. Are there any definitive answers in empirical science? Is there still ongoing discussion on (part of) your question?
Hypothesis	Introduction	Answers may come from divine inspiration. What are more practical ways to get a few alternative ideas of what the answer to the question could be?
Method	Methods & materials	If you wanted to repeat someone else's work to continue on that, what would you need to read in order to be able to do so? If applicable, example study areas can also fit in here.
Results	Results & interpretation	Here the data are presented that were collected and processed according to the methods section. Thus the results section 'mirrors' the methods section. If the data is entirely yours, this section does not need any references. See section below for the process of data reduction needed here.
Discussion and new questions	Discussion	Many scientific conclusions, when taken out of the paper, are meaningless without some context. Yet it must be a novel contribution to be a publishable conclusion. Thus the discussion mirrors the literature review. What did others find that relate to your subject? What are the consequences of your findings for the alternative hypotheses? It is rather unlikely that you resolved all questions definitively. What new questions arose from your work? This could restart the empirical cycle for you or someone else.
Conclusions	Conclusions	What conclusions would you give here, the specific conclusions regarding your data, the success of your field campaign or experiments, the wider conclusions regarding the hypotheses, new questions, of a combination of these? Look at other papers!

The empirical cycle can have a large range of dimensions. For example, the search for the human genome as reported in the famous human genome issue of Science is one large cycle. However, finding the answer to the question why your lunch box doesn't close or your matlab code does not run can also follow the empirical cycle. Can you think of other examples in daily life? And, as you proceed in your research project, can you think of where you went through small cycles embedded in the larger cycle?

Begin at the beginning

Another essential element in a paper and a thesis is the summary. A paper for a journal will have a shorter abstract than an entire MSc thesis. You can be more complete in the latter, Furthermore, you may wonder whether abstracts written by experienced writers are perhaps more concise (i.e. same amount of information but more densely written) because they are better writers. Which steps of the empirical cycle do you recognise in the abstracts of well-written papers and the (larger) summaries of theses?

It may help to start your writing by drafting a summary of the entire work. However, the summary is also the part that is rewritten after everything else has been done. If you start reading someone else's work, where do you start? It has to be crystal clear.

All is fair in love and war

How would you feel if you contributed to some scientific study with data and ideas, and this is not somehow obvious in the paper? What does one have to contribute to a paper to earn co-authorship? Can a thesis have more than one author? How can you acknowledge contributions that were not large enough to warrant co-authorship but were important to the study and you as a scientist (not you as a person)? How do you refer to work and data published in papers or online in libraries (such as theses or websites)? These were simple questions with, to a large degree, simple answers, but there are gray areas and people have become enemies over these issues. Ideas and data are the currency of science and using other people's work without proper reference is felt as theft. See www.plagiarism.org.

Reducto!

Raw data is not clear evidence for a hypothesis, because it is usually far too much to comprehend the trends and basic interpretation. It first must be reduced in steps. How many steps were needed in the data reduction in Kleinhans et al. (2011) from 8GB of raw video and acceleration recorder data to the single conclusion? How is this presented such that it can in principle be repeated by you? How are the data presented in steps such that you understand the story?

What different types of analyses (e.g. statistics, spectral, fitting, comparison to models, dimensional analysis) could you apply to your data and how would that help answering the research question convincingly? What types of graphs and maps can you think of that would present the analyses in a self-explanatory way?

What types of analyses could you do on your data that would strengthen the conclusions you believe to be correct? What reasons can you think of to remove certain bits of data that weaken the trends that you see? How would you feel if this had been done but was not reported in a study of positive effects and side effects in medication that was prescribed to you by a medical doctor from a university hospital?

Language: a mental faculty to communicate

Surely you read literature relating to the subject of your thesis. Which papers or reports did you find particularly clear in writing and presentation? Why?

Do you think that good papers and theses are written in one go? What style of working do you have: do you first put a structure in the text, or do you first throw out a brain dump and later shuffle this text, or what do you do? And did you like writing as a creative process like that?

Now I will give some explicit hints. Make the graphs and/or figures that answer your question. Thereafter, start sorting your graphs. If you have done the data reduction well, they tell the story for you. Include photographs and drawings of the setup and figures that explain the methods. If you are something of an artist, do produce a sketch that graphically explains your work, hypothesis, conclusions, etc. Then tell the story in words by describing the figures.

Don't try to get everything right in one go. Do several separate editorial cycles to check references to literature and figures, to check correctness of the content in relation to the data, to check whether every first sentence of every alinea contains the subject and message of that alinea, to check whether each chapter takes the reader by the hand, to check whether all material is in the right chapter (e.g. no discussion in the results), check spelling and grammar, check whether you can read the story entirely by only looking at your figures and abstract and conclusions, and so on.

Humans are fallible, and science is a community enterprise. So ask other students you trust to read your first version carefully and comment on the larger issues, not the spelling (and offer to read his/hers in return). Also, take time for this stage: put the thing away and do something else, so you can attack it with fresh views.

To ask or not to ask, that's the question

Questions arise from your curiosity and from the assignment by the supervisor. There are various sorts of questions: about causes (e.g. past climate change), about effects (e.g. storm erosion of the coast) and about mechanisms (e.g. debris flow mechanics) (Kleinhans et al. 2010). Why are these fundamentally different types of questions?

It is useful to distinguish three domains of questions: the a priori, the normative and the empirical. What does this mean? Which domain is that of the natural scientists?

To raise questions is one thing, but how do we arrive at a question that can empirically be researched in one project? Limit thyself… but how? A useful practical tool to help you think about this was developed by the Science Hub of Radboud University Nijmegen (Vragenmachientje van het Wetenschapsknooppunt, my questions in *italics*):

1. The question is specific: it is entirely clear what subject exactly you will be studying. (*Why would that work better?*)
2. The question is simple: there is only one variable, rather than studying four variables at the same time. You have to 'unpeel' your question by asking questions about it, until you end up with the narrowest question possible. (*Why would this be?*)

3. The question is measurable: the answer can be found in measuring something. Usually that measurement is done relative to something else (e.g. 'larger than') in the context of literature. (*Why is that?*)
4. The question is feasible: you can do it in the given time frame with your resources and the facilities offered by the BSc or MSc program, in particular the supervisor.

One way to gain more understanding of these four characteristics is to ask the question what your project would look like if the questions did not adhere to all of these. So, what can you do when your question is not specific, not simple, not measurable and/or not feasible?

Close encounters of the third kind

A public image of science is that scientists spend their lifes in isolation—the ivory tower—to produce explosions, brilliant ideas and useless science on their own. However, science has always been a community enterprise with much communication going on through personal contacts, coffee corner encounters, telephone, internet, conferences, project meetings and so on. What reasons can you think of?

Don't ever believe the myth that YOU should do your thesis on your own. But what does it mean then, a thesis written individually? When do you go and meet the supervisor? How often do you ask for help? What types of questions do you pose to the professor, which to the PhD candidate, and which to your fellow students?

To help you with the above I would like to make three points.

1. every single brain is different and has different experiences, so every different brain may think of something new and useful that others have not yet thought of (including the supervisor). Open-minded scientists love that kind of interaction and the ferocious (but impersonal!) discussions about new ideas, because it spawns even more new ideas!
2. every code, every lab and every piece of data has its own unwritten context and history. This is likely well-known by the supervisor and not by you, so you should ask questions about this, for example about the intricacies of a piece of equipment or whether you are allowed to use that bag of sand of thousands of euros…
3. many other questions can be answered by looking up in a manual or paper, and the official abbreviations that are globally used to answer such questions is RTFM ('read the * manual') and RTFL ('read the * literature').

Excellent ways to annoy your supervisor (and get a lower grade)

You are doing clever things at this level of education. Well, you are expected to do clever things at this level of education. This means that it should go without saying that the more mondane and menial aspects of science and communication have your full attention and do not require much attention from the supervisors. So here are some excellent ideas to annoy your supervisor:

- leave in plenty of spelling and grammatical errors
- don't bother to finish the work with a front page with all necessary information, a list of contents and so on

- write each alinea and each section as a horror story, working towards the surprise and leaving the reader in suspense what the alinea is about until the very end of the sentence where you come to the point (K.M. Cohen, pers. comm. Fac. Geo. UU 2005)
- don't bother the supervisor with questions if your fellow students have an opinion about it
- give your supervisor your first version of the thesis, as it is his/her job to teach you how to write and indicate what is essential for a good grade.
- don't bother to backup your data, code and writing. Computer storage is so reliable these days and people have completely stopped stealing the hardware, and fires are forbidden
- don't ever, ever give your data to the supervisor, or make a nice labyrinth of the data structure, spreadsheets and code
- don't waste your time on cleaning and clearing the lab
- the technicians are there to help you immediately and, as is quite clear from their grumpy appearance, they do not appreciate 'please' and 'thank you' and all that
- never mind deadlines… we all die in the end and the supervisor, who will die before you anyway, has nothing else to do and will correct your work overnight

Good luck, and forge a master piece!

Acknowledgements

Comments by Tjalling de Haas and Jasper Leuven were helpful to improve this document.

References

Hempel, C.G. and P. Oppenheim (1948). Studies in the logic of explanation. Philosophy of Science, 15(2), 135-175.

Kleinhans, M.G., M. F. P. Bierkens and M. van der Perk (2010), On the use of laboratory experimentation: "Hydrologists, bring out shovels and garden hoses and hit the dirt", Hydrol. Earth Syst. Sci. 14, 369-382, http://www.hydrol-earth-syst-sci.net/14/369/2010/.

Kleinhans, M.G., H. Markies, S.J. de Vet, A.C. in 't Veld and F.N. Postema (2011), Static and dynamic angles of repose in loose granular materials under reduced gravity, J. Geophys. Res. Planets 116, E11004, doi:10.1029/2011JE003865.