

How to prepare a theory-driven, quantitative empirical study?

Eldad Davidov

Institute of Sociology, University of Zurich

First and foremost, be clear about your research question. To do this, prior reading and investment of time are a necessity in the question development process. Research question(s) may evolve, for instance, from previous seminars or courses that you attended or from personal interest. This is your opportunity to influence the topic in such a way to ensure that working on your Forschungs- and/or Masterarbeit will be an enjoyable process. During your initial consideration, you may find several topics that are potentially interesting for you to deepen your knowledge about. Finding the right question(s) for you and trying to provide answer(s) in a theoretically driven empirical way may involve the following steps:

- 1) Conduct a literature search about the topic that interests you.
- 2) List various research questions that have been addressed in the studies you find on the topic.
- 3) List hypotheses that have been tested. Research questions are different from hypotheses. Whereas research questions tend to be somewhat more general (although this is not necessary, some questions may be very concrete), ideally the hypotheses are clearly formulated, concrete and testable statements.
- 4) Identify a gap in the literature: It may be an aspect which has not been addressed or one that recent studies have proposed for future research. It could also be the absence of an empirical test of a research question in a certain context: You may propose a replication of an existing study in a different setting (countries or cultural groups), where this aspect has not yet been tested and explain why it is important to conduct it. Such replications may also be potentially valuable contributions to the existing literature and a test of the robustness of theories in different contexts.
- 5) Make sure that the aspect you identified is really a gap: Verify that it has not been dealt with before in the literature.
- 6) Once you have identified the gap, try to formulate your research question(s).
- 7) Derive concrete and testable hypotheses based on the literature you acquainted yourself with. Think about the mechanism that underlies each of your hypotheses.
- 8) Decide where or in which context you would like to test your hypotheses: Are there certain cultural groups or countries where this aspect would be particularly interesting to study? Are there any time points in which it would be advantageous to investigate your question? Would you like to conduct a comparative study across these contexts (e.g., countries, time points)?
- 9) Think about the contexts (culture/country and/or time) and whether they have a possible effect on your findings. Will the contexts play any role in your comparison (moderating your postulated individual-level relations)?

- 10) Identify data sources that you could use to answer your questions and test your hypotheses. Secondary data analysis and large-scale data are valuable sources, and such data are often freely available on the Internet (e.g., the European Social Survey, the International Social Survey Program, the European and the World Value Surveys, etc.). If you identified a usable preexisting data set, check if all the variables you need are available in the data set. If only some of them are available, consider modifying your hypotheses. You might have to give up testing hypotheses for which no data are available. Such hypotheses that you cannot test will be a matter of future research that you could address in the discussion/conclusions section when you finalize your Forschungs- or Masterarbeit. It is advisable to look for potential data sets right from the beginning of your search for a research question. It is not productive to invest so much time on developing a research question, which is impossible to test due to the absence of data. Obviously, it would also be possible for you to collect your own data for your study although this may involve much work. For instance, if you wish to use survey data, you may have to deal with the formulation of the questionnaire items, the preparation of a questionnaire, collection of data (e.g., using face-to-face interviews, paper-and-pencil, online tools, etc.), and the preparation of a data set that contains the data you collected. As an individual researcher you will have, in most cases, very limited resources for data collection. As a result, the quality of the data might be lower than that of international projects (e.g., in terms of reliability, validity, sample size, and representativeness of your target population). However, this should not discourage you from collecting your own data: On the contrary, you will be much more flexible in the formulation of your research questions, as you will be less dependent on external data sources and on the questions that were included in secondary data sets. In any case, you will be able to decide on your own about which data to collect. Another possibility to produce your own data may be to conduct experiments. Experiments are another exciting and powerful tool to collect data, test for causality, and answer various research questions. You may also try to collect data from other sources, such as newspapers, media, and various published or unpublished reports. These data may be on the individual or the contextual (group) levels of analysis (e.g., regions, countries). Finally, you may combine data from various sources. When you collect your own data, make sure that your sample size is large enough to enable you to conduct meaningful quantitative multivariate analysis.
- 11) The next step involves data analysis. This is where you may apply and further develop and expand the knowledge you have gained during your studies of data analysis techniques. Get to know your data before you begin to analyze it! Check for data entry errors, label your variables, and check for missing values. Always begin with descriptive analyses of your variables (e.g., distributions, means and standard deviations, observation of the labels and scales used, correlations, sample size, imputation of missing values, if necessary) before turning to the multivariate

analysis. The multivariate analysis may involve, for example, the construction of indices or latent variables, testing for the measurement quality of your data, testing for violations of assumptions about the data, investigating invariance properties across groups (such as cultures, countries, or time points), estimation of causal effects and relationships between your variables of interest, improvement of your models, and evaluation of their fit to the data. Data analysis is generally a very time-consuming process; plan about one quarter up to one third of the time invested on your research for this step.

- 12) Now you are ready to reflect on your findings. Did the data provide full or partial support for your hypotheses? Which research questions could be answered, and – maybe even more interesting – which questions could not be answered or addressed by the analysis? Are there any further analyses necessary? What can you learn from your findings? Do they provide a significant improvement of our knowledge with respect to your research question? Below you will find some suggestions about what to keep in mind while writing your report. The suggestions below are provided only as an orientation, as they may not always perfectly apply to all types of theory-driven quantitative empirical studies.
- 13) a) Your report should contain a description of the steps you followed during the time you actually conducted the study. In other words, a typical theoretically driven quantitative empirical study will have an introductory part, a theoretical part, a description of empirical findings in previous research, a summary of the hypotheses and the mechanism behind them, an empirical part, and a discussion/conclusions section. During the preparation of your theoretical part, it is very important to invest time and energy in thinking about and explaining the mechanisms, that is, the logic behind the postulated relations between the theoretical variables of interest. This is the main bridge between your theory and the data you use to assess it. Provide the necessary theoretical background for understanding these mechanisms and include relevant references. Avoid discussing theoretical aspects that are not relevant for the study, and do not forget to discuss those aspects that are central for your empirical investigation. Providing theoretical background for the effect of control variables is useful, but do not invest too much time and space for this part, because the control variables are not the focus of interest. If your model is complex, include figures to illustrate your ideas. After your theoretical discussion, provide findings from past research, and discuss whether such previous findings support your expectations. This theoretical part should typically not be longer than about one third of your report. After summarizing your hypotheses, you are ready to begin with the empirical part.
- b) The empirical part should be the main part of your report. Describe all the data sets used in your study, the place and time of data collection, and the procedures and techniques used. If you conducted secondary data analysis, provide a link for further documentation of the data collection. In the next part, describe your variables. Provide a table that lists your variables, that includes their names,

question formulations, scale used, and some information about their distribution (range, mean, standard deviation). Discuss the procedures you used to deal with missing values. Next, report your findings from the descriptive analysis. If you identify problems or mistakes in these analyses, you may need to repeat them. Include figures that display relations between your main variables of interest. These figures might already give you and your readers a first hint about whether the relations you expected to find in your hypotheses are present in the data. Finally, in order to examine whether these relations also hold after controlling for all the other relevant variables that are involved in your analysis, you will turn to the last and main empirical part, in which you report the findings of your multivariate analyses. During the preparation of this section you may reveal problems or mistakes in your analysis. In such a case, you may have to repeat your analysis. If there are any assumptions you made about your data, report them, and describe whether your findings are robust to violations of these assumptions. You may be able to solve some of the problems you discover in your analysis, but some problems might remain unresolved. This is absolutely okay as long as you discuss these problems and explain why they cannot be resolved.

c) Finally, you may turn to the discussion and conclusions part. Do not underestimate the importance of and challenges related to this section. This is where you have an opportunity to reflect on the whole study, summarize what you found, why it is important, and what still remains a topic for future research. If there are any problems or limitations you encountered during your work, this is the place to discuss them. Limitations are an integral part of theoretically driven empirical studies; do not hide them, instead – discuss them offensively, and explain why they could not be addressed in your particular study. Finalize this section with what you have achieved in your study.

d) The report should finalize with a references list followed by any appendices you may like to include, such as output data, questionnaire items, etc. Choose some style for citations and stick to it (see, e.g., <http://www.apastyle.org/> or <http://www.chicagomanualofstyle.org/home.html>). Do not forget to refer in the text to any figure or tables you include in its body. Read your report several times, with a time gap between each inspection of your text, and if possible, allow someone else to proofread the text, and revise it wherever necessary.