eTIPS
etextbook institutional publishing service

HOW TO SUPERVISE (AND BE SUPERVISED) ON A RESEARCH DEGREE

Tips and tools for supervisors and students

Frank Rennie and Michael Rayner

Chapter 4:
Implementing the data collection methods

Getting research ethics approval

Teaching someone about what we mean by the term 'research ethics' is one thing, but unless a person actually understands why ethical standards are essential, then everything else is fruitless. It is relatively straightforward to present examples of good ethical practice (and what happens when this practice is ignored) but this simply underpins the implementation of the ethical standards, not the need for them. Fortunately, there are lots of detailed guidelines and professional codes describing the expectations of ethical behaviour, many of which are readily available on the web. We say 'lots' because the ethical standards vary widely in content and detail, dependent on the subject discipline, the research methods employed, the level of study, and several other factors. This might sound

vague, but think about it. There will be a different level of scrutiny required if a researcher seeks access to the confidential medical files of patients, rather than simply asking patients to respond to a few verbal questions. There will be different standards again if the researcher plans to work with animals, or children, or vulnerable adults with diminished responsibility. There is also an ethical code for internet-mediated research, although this is new, variable, and highly contextual, so it is an evolving set of guidelines. The 'copy-and-paste' command has opened the door to a whole new set of rules and responsibilities for working in the digital environment. Despite these various differences, the purpose of research ethics is the same in each case – namely to prevent causing harm to the participants, to preserve their dignity (for example their right to anonymity) and to enable them to withdraw from the study without any undue pressure or penalty.

The supervisor has a crucial role in talking through with the student all of the possible ethical issues which might impact upon the research. Frequently, the ethical issues can be resolved very easily, and the research can proceed, but the

simple fact of working through the range of possible issues which might arise, and sharing experience on how they can be satisfactorily dealt with, is an important part of the professional training required by the student.

Every university, and most professional associations which come into contact with research, will have codes of conduct and formal procedures for scrutinising and approving research ethics proposals for research projects with which they are involved. This will require specific application forms, a scrutinising committee, and a formal code of research behaviour with which researchers are required to conform. Some procedures to gain ethical approval for research are particularly detailed, for example anything requiring contact with the health service, such as patient notes, contact with patients, or any engagement with either health staff or medical procedures, is likely to involve substantial detail and very careful research design.

For these reasons, there is a crucial stage between deciding on what research methods are to be adopted for a study, and the commencement of data collection. This

crucial stage is where the researcher submits the details of the design, methodology, and any issues relating to the collection and storage of data, for approval by the university ethics committee. Only after ethical clearance has been approved can the student begin to collect data. Failure to obtain approval before data is collected may result in the university deciding that this data is not admissible for inclusion in the study. If there have been any severe breaches of ethical responsibility, the study may be terminated or the student de-registered. For this reason, the ethical approval of a student research project is a gate-keeper stage of every study.

Key concerns in all research ethics matters are to avoid causing harm, to respect confidentiality, and to maintain high standards of moral integrity. The latter, for instance, might refer to a very wide range of 'common sense' standards such as to refrain from cheating, plagiarism, falsifying results, vested interests, and so on. Though they may seem 'common sense' to most of us, we tend to forget that many of these issues are perceived differently by different cultures, and influenced by pressures which might

be applied – internal and external to the institution – to 'encourage' researchers to produce favourable results one way or another. It is for these reasons that gaining the approval of the research ethics committee is a fundamental gateway for any research project before it can be seriously undertaken.

Fortunately, most research projects have fairly straightforward ethical requirements which are easily satisfied in full. A lot of the ethical safeguards might be regarded as 'simply common sense' (and so they are) but you might be surprised by the number of times people say 'Oh, there are no ethical issues with my research!' This is almost certainly wrong. Even the issue of whether the researcher with half-formed ideas should be 'wasting' the time of an interviewee who almost certainly has something better, perhaps crucial, to do, is an ethical issue. For these reasons, seeking ethical approval for research should be a serious matter, but not something to fret unduly about, if the researcher has properly thought through the research design. Once the ethical approval has been obtained, the researcher is able to jump out of the starting blocks to

engage with data collection, and this is where the real fun part starts.

Pilot studies

Before rushing off to take the final leap into the swimming-pool of the main data gathering exercise, it is usually advisable to conduct a quick reality-check. In some form or other, a short pilot study, which samples just a small part of each data gathering method, is a useful activity at this stage. Depending on the diversity of the selected data-gathering methods to be used in the main study, it could mean asking 3 or 4 people to complete a questionnaire, or trying out the interview questions on a few 'volunteers', or perhaps conducting a trial run of a bench experiment, just to make sure that things progress in practice as smoothly as they have been envisioned in theory. A pilot study can do several things. In the first place, it allows the supervisor to observe just how much thought, care, and background research has been already conducted in the formulation of the research methodology of this study. There may be some opportunity for improving the methods, or there might

simply be a reassurance that things have been well-planned… so far. Feedback at this design stage may avoid making elementary mistakes or designing a method which will lead to incorrect or misleading results.

For the research student, the pilot study can have multiple benefits. The reassurance of the supervisor is useful, but the feedback from the pilot participants can be even more critical. This is the time when slightly ambiguous questions can be reworded, and research methods can then be tweaked to make sure that they do what it is hoped that they should do. If a participant reports that the wording of a question is difficult to understand, or that there is no relevant category of response, this suggests that other people in the larger study will encounter the same difficulties. The error created will become multiplied when the full study progresses, and may become significant. The fault in the misunderstanding lies with the researcher, not with the participants being questioned. It is up to the researcher to construct questions which are unbiased, not leading towards a particular response, and are clearly understandable by participants in the sample population.

Similarly, with experimental design, if the experiment has a fault in its design, it is much better to find out at this stage through a short pilot study, than to run the experiments several hundred times before finding out that there is a fundamental problem.

Writing up a description of the pilot study is an integral part of the methodology chapter in the dissertation. If there were changes made to improve the design of the main research survey, (and even if not) then this is a good place to note the changes, justify them, and demonstrate that the researcher has not simply woken up one morning and plucked a research design idea from thin air. Demonstrate that thought and care has been invested in this. Even the experience of codifying and analysing a few results from the pilot study might give the researcher (and the supervisor) a good sense of the ease (or difficulty) which the final main data-set will present, and allow for a simplification or clarification as appropriate. It is a huge mistake to seek a 'short-cut' by avoiding pilot studies!

Recording data

Starting to record the new data which is being gathered as part of a research project, whether a long-term study like a PhD, or a quick toe-in-the-water project, is the most crucial, but perhaps the subtlest stage of the research. If you gather too little data, the project may flounder even before it gets started; too much data, and a metaphoric mountain of results can be generated by cross-correlation and individual analysis, which can paralyse a project almost as quickly as having no data at all. Then there is the question of what is the 'right' data? 'How will I know it when I see it?' In reality, it is as likely to be different for every individual project as the diversity of methods of data gathering. The correct procedure, of course, is to recognise that recording the correct data is integrally dependent on selecting the correct research methodology, and in carefully selecting how the data will be collected, coded, and stored in the future.

One of the most impressive records of research data that we can remember, is from a scientist who was studying birds of prey, and his handwriting in an old notebook

recorded what seemed to us to be almost every conceivable factor which might influence nesting success, including several factors that most other researchers would never have begun to consider relevant. He was of course correct, for it is often the correlations with hidden, and often apparently spurious, information which leads to the really stunning breakthroughs in research projects. There are many different ways of recording research data that you might collect, and there is no one-size-fits-all solution. If you are interviewing people, there is a choice between taking notes, audio recording, or video recording; all these methods have their advantages and disbenefits. Taking notes is less obtrusive, but also can be distracting for the researcher. Audio recording can be done easily with a digital recorder, or a suitable app on a smart-phone, but some people may be more guarded in their responses when they are being recorded (at least until they begin to relax into their conversation). There is also the problematic issue of what to do with all the data you have gathered. Gathering a huge mass of data can be attractive, but it needs to be proportionate to the scale of the project, because there is little point in generating a mountain of data

if 80% is left unanalysed and unused. Great care needs to be taken to strike a balance between collecting a good data-set which provides rich possibilities for future analysis, against de-motivating your participants by presenting them with huge questionnaires or over-long interviews. Similar constraints apply when conducting laboratory experiments, fieldwork, or desk-top studies.

Finally, in addition to having to consider your recording requirements in terms of how you propose to codify and analyse the potential results (there is little point in collecting data so randomly that it cannot be interrogated effectively) there are the issues of long-term storage and access to the data. The research supervisor has a crucial role here, not simply in helping to shape what the research student proposes to gather, or how that might be analysed and interpreted, but in providing the continuity which may extend over several decades and overlap with numerous related research projects. In an increasingly digital and open educational society, not simply the research results, but also the raw research data is also becoming more open and accessible. It is becoming more possible and more

likely that scholars coming after you will read not just your conclusions, but also your original data recording notes, so think carefully about what you collect and how you record it!

Storing and archiving data

When doing a PhD, it is possible to have hundreds of photocopies of academic papers stacked in small piles on according to their theme and relevance to the different sections being written about. The raw research data, however, is likely to be compactly contained in electronic format in the form of tables and graphs; row after row of numbers on spreadsheets which could be tabulated and correlated in any format desired.

Now, when we generate research data, it is almost entirely in electronic format, and it is automatically stored in several separate places. We will probably have a personal space in the memory banks of the university computing system, and this space is automatically backed-up overnight to an institutional cloud-space, so that the information can be accessed from different devices and geographical

locations. Usually, we also store data for individual projects on a separate, encrypted, memory stick or portable hard-drive. The digital age means that after two or three clicks, we can be assured that copies of our data are safely held in four or five independent locations. Research students can simultaneously share data with a colleague or supervisor in a different part of the world without even leaving their own desk. A key responsibility of a supervisor is to ensure that a research student is adequately trained and aware of the various ways to ensure that important (often confidential) data is securely stored and accessible only to approved people.

This is only the tip of the iceberg, however, because the production of digital data raises almost as many questions as it provides innovative opportunities. There needs to be an early discussion in the supervisory team, for instance, about not simply which data will be stored, but where will it be stored, for how long, and who will have access to it? This is not simply an issue of security, although security, confidentiality, and appropriate use of the data will certainly figure in the discussion. There is a growing awareness that

when public money is used to fund research, there needs to be a transparent return on public interest. Initially this has meant that research results, reports, and journal articles, should be made freely available to the public. This is being extended in the next Research Excellence Framework in the UK to insist that if the journal article is not already published as an open resource, it needs to be added as an open source on the digital repository of the relevant institution. But there's more.

The argument for openness has been extended to include the research data generated by the public funding, so the datasets themselves are trending to become open and shared property. Whether the raw data is numbers, interviews, audio recordings, photographs, or other recordable results, the likelihood is that the data being gathered by a researcher today, is probably going to be a shared resource tomorrow. It will be possible for other researchers, in subsequent years, to access your raw data, perhaps combine it with other sources of raw data, and re-analyse, re-interpret, and publish their conclusions. It now begins to matter a great deal more seriously exactly who

can gain access to your research data, and for what purposes. As the law currently stands, a bona fide researcher can have access to open datasets for up to ten years after they have been deposited. But here is the catch – if any researcher accesses this data after nine years, the open-access clock is automatically reset for a further ten years. This ensures the certainty that data which is being collected and digitally stored just now, might be still openly available long after the initial researcher has moved on from that research topic, perhaps changed institutions, changed careers, maybe even passed away. The raw data of open access digital resources is now guaranteed to have a lifetime longer than the career-span of many individual researchers. So think carefully about what you gather, how you organise and store it, and what your legacy of research data will be!

Data analysis

The data analysis stage is one of the main research phases where the supervisory team can really make a significant contribution to assisting the research student. Obviously,

the student still needs to do the work for themselves, but this is a stage in the PhD process where the depth and breadth of experience of the supervisors should shine through and help the student to make sense of a complex set of tasks, and make them a bit simpler to complete. Having collected a mass of data, perhaps even coded and categorised this data according to exacting and laborious protocols and methods of analysis, the student needs to understand what this data is actually saying. This might be a simpler task for some projects than for others, according to the amount of data collected, the form in which it was collected, how detailed or exact the observations or calculations are, or what methods for codifying or interpreting the raw data have been employed during the research.

To the beginner, this might seem a straightforward task, but there is no 'one-way' to analyse data, because there are many, many different forms of data. That data might be collected at different levels of granularity, different levels of accuracy, and embody different assumptions and methodological approaches. At some early stage of the

analysis it is usually a good idea for the research student to sit down with the supervisory team, spread the collected data out on a table, and look at it together. The student needs to identify a number of key attributes of this data, such as what does this data actually indicate? How robust is the data? What is its accuracy and what are its limitations? Are there any correlations (positive or negative) and so on, leading to the penultimate question, which is what does this data actually tell us? Hopefully there will be a new insight, or a discovery which, at least in part, will address the initial research question(s). The final challenges are then, how should this information be presented so that it is understandable to others who have not been involved in this research, and how transferable is this knowledge to other (present and future) researchers? The ability to repeat certain methods, or to re-combine and re-use data (and the results of its analysis) are particularly useful features to enable contrast and comparison with other projects in similar or related subject areas.

It may seem a bit ironic, or a perhaps a bit of a paradox, given that we are often seeking to reveal hitherto unknown

facts, and 'answers' to high-level research questions, but usually it is better to be less ambitious in the interpretation of the conclusions, but to be absolutely sure of the reliability and credibility of our data, rather than to propose over-ambitious conclusions (exciting though they may seem) which are based on assumptions, sketchy evidence, and correlations which are skating on thin ice. If the research can be shown to produce solid, incontrovertible evidence, regardless of how large or small the breakthrough, then this small advance can be built-upon by subsequent researchers. If the conclusions are like a house built on thin ice, then there is always a doubt about the stability and credibility of the data, or its interpretation, and therefore the value of the research output is devalued. With data analysis, it always pays to check, cross-check, consider the constraints and limitations, then double-check each stage before you venture to draw conclusions to be shared with a wider audience.