Chapter 4
Researching Surgical Education

Heather Fry, Nick Sevdalis, and Roger Kneebone

4.1 Introduction

Chapter 1 has already referred to education being a composite of many fields and disciplines. Surgical and medical education research can be viewed, arguably, as specialised branches of social science/education research; this view is contested, see for example Gill and Griffin (2009). The context of the problems to be researched is different, but many of the issues are essentially similar to those in other forms of education, for example, how do learners acquire knowledge or learn a particular skill, is assessment reliable, does training teachers make any difference, what is the optimal strategy for teaching and learning knowledge or a particular skill, what motivates learners, etc. There are particular parallels with other areas of professional education. A good introduction to various types of social science-based medical education research, by leading international experts, can be found in Swanwick (2010). A major exception to reliance on social science research methods is that neuroscience is increasingly drawn on by some branches of education.

However, a number of differences exist between surgical education and other forms of educational research. An obvious one is that surgeons do not just depend on factual knowledge, nor do they use technical skills on an inanimate object, nor do they just develop highly expert and sophisticated analytical and synthesising skills.

H. Fry (✉)
Higher Education Funding Council for England (HEFCE), Bristol BS16 1QD, UK
e-mail: h.fry@hefce.ac.uk

N. Sevdalis
Division of Surgery and Centre for Patient Safety and Service Quality,
Imperial College London, London, UK
e-mail: n.sevdalis@imperial.ac.uk

R. Kneebone
Surgical Education, Imperial College London, London, UK
e-mail: r.kneebone@imperial.ac.uk

H. Fry and R. Kneebone (eds.), *Surgical Education: Theorising an Emerging Domain*,
Advances in Medical Education 2, DOI 10.1007/978-94-007-1682-7_4,

They learn all these and more. Moreover, the focus of surgical practice is the patient, and the patient therefore has to be considered as part of the learning context; this adds to the uncertainty and variability of the surgical context, making it an even richer and more complex field for research. The patient's interests and needs will predominate over a focus on learning and training; the critical nature of surgical practice means the need to safeguard the patient and their health throughout the learning process is dominant. Patients, also, may act as part tutor; some patients' behaviours may distract the learner; some patients will have more complex problems to treat than others and the flow or the symptoms of the presenting patients cannot be controlled for educational purposes. A second difference is that the trainee surgeon or physician more often than not works and learns as part of a team, unlike the lone university student taking a mathematics examination. However, it could be argued that the 'patient, team and safety factors' are little different in essence from, say, the client and product team building a bridge in engineering, or from the client and barrister in practice of law – in other words the context of surgical practice is not similar in its essential nature from the context in which other professional education occurs.

The research tradition that surgical education draws on is more that of education and related social sciences, and less that of traditional surgical science. This means that lab-based testing, for example, is not something that can be used straightforwardly in surgical education. It also means that biomedical traditions of, for example, clinical trials can rarely be used because the component parts of the research can rarely be sufficiently controlled or isolated to render such an approach valid – and even if they can be, the conditions created are then far from the ones usually prevailing 'in reality', which will themselves be varied. A characteristic of surgical education research is therefore one in which, even when quantitative methods are used, the human factor and the difficulty of controlling variables are never far away. Another distinctive feature of social science research is its use of theory, for example, drawing on disciplinary knowledge to ensure that research problems contribute to the evolution of a field and forming theoretically informed research questions (Albert and Reeves 2010 and Teunissen 2010).

This chapter explores the two main paradigms used in surgical education research; they are often used together in mixed methods studies. The chapter is not sufficiently detailed to act as a manual of such methods. Its primary purpose is to introduce the non-social scientist to surgical educational research techniques, to tease out some key premises and features and to act as a useful backdrop to understanding some of the research studies that are developed/referred to in subsequent chapters.

4.2 The Quantitative and Qualitative Educational Research Paradigms Compared and Contrasted

Quantitative research relies heavily upon deductive reasoning, where a hypothesis is formulated and then tested statistically. This type of approach to education research will be immediately recognisable to surgeons in that it may use similar

sampling techniques and statistical approaches to some biomedical research. Research investigations are designed which allow for a systematic approach, aiming for results that are generalisable beyond the limits of an individual investigation. Here attention is paid to random or representative sampling from large populations, adequate experimental control of extraneous variables, use of validated metrics, minimisation of investigator bias and other similar notions. But even here, a social science use of statistics is more likely to be in relation to something such as questionnaire responses or assessment scores rather than outcomes from randomised controlled trials. Such quantitative methods are extremely powerful, but there are limits to the kinds of questions they can address. This type of research tends to be published in the form of relatively short journal articles, with a consistent and highly formalised structure. The ordered sequence of Introduction, Methods, Results, and Discussion (IMRaD) provides a familiar framework, allowing key information to be conveyed quickly and compared with other articles. Critical appraisal involves being able to judge the quality of research design and methodological rigour, analyse statistical methods and weigh up the validity of conclusions.

Qualitative research tends to focus on capturing the perspectives of individuals, using methods such as interviewing and observation. Such methods make no claim to produce generalisable results, but instead aim to shed light on the area being investigated – often opinions, behaviour, interactions and processes. Qualitative research relies much more upon inductive than deductive reasoning. Understanding evolves through a process of observation and analysis, which looks to see what is there, without having a preconceived hypothesis. What is needed is a detailed description of what individuals think, feel or do. Research of this kind is likely to acknowledge that there may be many 'truths', and that each person's experience may be as valid as anyone else's. This kind of research tends to be published in the form of books or much longer articles, using a different set of conventions where the author often appears as a person, and his or her responses to the research are acknowledged as a factor. The research itself is framed within an extensive discussion of what has been written by other authors, and the findings are often interwoven with a critical discussion of what is already known. The criteria for assessing the quality of the research outcomes are very different from those of quantitative research, often depending on the transparency, thoroughness and suitability of the process, rather than an 'objective' evaluation of an impersonal method, and on how far findings are recognised by those in similar situations. 'Authenticity' is strived for. The results of this kind of work can rarely be exactly reproduced because every person observed or interviewed has a unique perspective, but the process by which the researcher has reached his or her conclusions must be laid open for scrutiny and critique.

Table 4.1 aims to capture some of the key differences between the two research paradigms, being a schematic representation of two broad approaches to research.

As mentioned above, in practice, educational research often uses a mixed method approach and none of the methods may be at the purest or most extreme end of either research paradigm.

Table 4.1 Characteristics of qualitative and quantitative research to illustrate their different approaches and emphasis

Qualitative research	Quantitative research
Acknowledges a degree of subjectivity and reflexivity	Claims objectivity
Focuses on individuals or groups	Focuses on populations
Exploration	Trends
Usually starts from an open question	Usually starts from a hypothesis
Inductive reasoning	Deductive reasoning
Analysis of text	Statistical techniques
Purposive sampling	Preference for random, or appropriately stratified sampling
Asks about why and how	Asks how many, in what proportion and is it significant
Findings often not transferable	Findings may be transferable
Rarely generalisable	Usually generalisable
Takes place in the real world, often uses observation, opinion and perception	Based on codified observations or responses
Analysis may start alongside on-going data collection, conclusions may start to emerge as data is being collected and influence subsequent data collection and analysis	Analysis will happen after all data has been collected
Both methods may generate models and theories	

4.3 Quantitative Surgical Education Research

As highlighted above, quantitative research approaches in the context of surgical education share numerous facets with similar approaches across the entire breath of biomedical and clinical sciences – from laboratory-based science to translational research aiming to take scientific and technological advances to the bedside and evaluate their impact on patient care. These approaches are also common in some social science disciplines. A key aspect of this type of research, which sometimes differentiates it from more qualitative approaches, is that decisions about *all* aspects of research have to be made at the early stages of the research and strictly adhered to whilst the research is being carried out. Such decisions include choice of research designs, endpoints and other metrics to be collected, and analytical (statistical) approaches. The stages of data collection and analysis are kept separate. In the light of this, adequate piloting of the scientific techniques, tools, and materials to be used is a key aspect of this type of research, which ensures feasibility and quality of the entire research endeavour.

In the remainder of this section, we consider the way quantitative research questions and hypotheses are set, the range of quantitative research methods used and the range of analytical tools that are available in the context of surgical education. These aspects of quantitative educational research will be highlighted in order for comparison with qualitative approaches to become apparent; it does

not provide a comprehensive overview. Numerous research methods textbooks are available to surgical education researchers, which should be consulted for in-depth coverage of the issues discussed below (Blalock 1979; Cohen et al. 2007; Shadish et al. 2002; Zeller and Carmines 1980).

4.3.1 Research Questions and Hypotheses

Quantitative research in surgical education is typically driven by well-defined questions and hypotheses set prior to the commencement of the research. Research questions tend to be broad in nature. Recent examples of research questions tackled within the quantitative research paradigm include 'what is the impact of stress on surgeons' performance in the operating theatre?' (Arora et al. 2010a) and 'can behavioral and cognitive skills of surgeons be captured reliably and validly in the operating theatre?' (Sevdalis et al. 2009). These questions are in effect open-ended: there is no preconceived direction in the possible answer that will emerge from the data. Research hypotheses tend to be narrower than questions, typically specifying a direction in the anticipated research outcomes. For instance, the research question about surgeons' stress could be 'translated' into a research hypothesis as follows: 'Stress impairs surgeons' performance in the operating theatre'. Analysis of the data that the researcher collects should subsequently be able to determine whether the research hypothesis is supported or not, and how strongly.

Because of the specific, pre-specified nature of research questions and, particularly, research hypotheses, successful quantitative research in surgical education tends to be theory or evidence driven. Research questions and hypotheses do not arise in a vacuum; rather, they emerge as the 'where next' questions of empirical research that has already been carried out. Previous research within a certain topic thus drives quantitative approaches, so that forthcoming research builds onto the empirical evidence already accumulated and takes it to the next scientifically logical step. For example, research on the efficacy of surgical simulators as training and education tools for technical skills in novice surgeons started with providing evidence that such simulators can differentiate between novice and expert performance, went on to ask whether trainees' skills improve over time following simulator-based training, followed this up by asking whether the skills acquired on a simulator transfer to performance in the real environment of an operating theatre, and is now addressing the question why skills under certain conditions show transferability to real-life performance, whereas under other conditions they do not (for relevant reviews, see Issenberg et al. 2005; Lynagh et al. 2007; Sutherland et al. 2006; Sturm et al. 2008). Adequate reviews and/or meta-analyses of past research findings within the domain of interest are invaluable in shaping the direction of quantitative educational research and the specific content of the questions asked and hypotheses tested.

4.3.2 Research Methods

A key aim of quantitative educational research is robustness and generalisability of findings. The key idea here is that all tested participants are sampled from more or less similar populations. By collecting quantitative data and submitting these data to statistical analyses, quantitative researchers use the data they collect from their samples to make robust inferences about the populations from which the samples have been drawn. A key aspect of this process is that in order for an inference to be accurate, the sample that is being tested should be as similar as possible to the target population. For instance, if a researcher is interested in teaching and learning opportunities in Emergency Departments (EDs), a range of EDs need to be sampled to ensure that factors like whether the ED belongs to teaching or non-teaching hospital, staffing levels (low or high), and volumes of patients seen by ED personnel (low or high) are taken into account. Deciding which factors should be taken into account is driven by the existing evidence base, which reveals whether some factors are likely to affect the measure(s) of interest. Sampling considerations are therefore of critical importance in quantitative research and care should be taken in choosing a sampling strategy.

For our purposes here, it is important to highlight the relevance of random sampling and stratified sampling for quantitative research. In the former, all members of a population are equally likely to be drawn into a researcher's sample; in the latter, the sample is constructed in such a way that important population strata (e.g., gender, age and level of expertise) are accurately represented in the sample. True random sampling is often difficult to achieve in surgical education research. Instead, evidence-driven stratification is a viable alternative, which ensures that potentially confounding variables are built into the design of the research – and therefore can be taken into account when the data are being analysed.

Once the sampling strategy has been decided, quantitative educational researchers are faced with a range of approaches to the design of the research. A number of options are available here, with the 'true experiment' (randomised controlled study) being thought to offer the highest level of unbiased empirical evidence. Such studies offer truly random allocation of research participants (individual subjects, clinical units or entire institutions) to the various conditions of the researcher's design and include a control group, in which no educational intervention/treatment is carried out. In addition, ideally both the research and the participants should be blinded regarding which condition a participant has been allocated to, to prevent bias from the research as well as from the participant. For practical reasons, however, such studies are often not feasible in educational research: for example, it is not feasible for a researcher to be blinded to the treatment condition of a group of participants, if s/he is the one to deliver the training intervention to them (a problem often encountered in clinical surgical research too). Apart from true experimental studies, other design possibilities include quasi-experiments, with pre- and post-intervention treatment groups, or even only post-intervention studies, when a pre-intervention baseline is not available.

It is important to stress here that, because surgical education research is typically carried out with busy clinicians, often in highly pressurised environments, researchers often find themselves having to make informed trade-offs between what would be scientifically desirable versus what is practically achievable. Practical considerations should be taken into account early on in educational research designs, as they are often important determinants of how many participants will be available and when, how much time is available for an intervention to be delivered, whether the same participants will be traceable for a post-interventions assessment, and similar considerations. Whereas a laboratory scientist exerts adequate control over such issues, surgical education researchers often have to 'make do' within the constraints of service and education delivery as they currently stand. Although this does not mean that the research will be of a lower standard, practical considerations very often feature in the relevant 'Methods' sections of educational research articles.

The final step in the research method process is the choice of outcomes, or endpoints and the tools that will be used to capture them. As in all quantitative research, in surgical education too measures should be reliable (i.e. they should be measuring educational and learning constructs consistently) and valid (i.e. they should capture the constructs they purport to capture). A significant body of quantitative educational research in surgery focuses on the development and validation of assessment tools, which can then be used in further studies. Recent examples of such research include the development of tools to capture stress experienced by surgeons in real time in the operating theatre (Arora et al. 2010b), and tools to capture the quality of team working behaviours exhibited by surgeons (Sevdalis et al. 2008, 2009). Once such tools have been developed and evidence exists on their reliability and validity, they can be used to capture factors that affect learning outcomes (e.g. the impact of stress on the transfer of skill from simulation-based training to real operating theatre) or to assess directly the efficacy of training modules for surgical trainees (e.g., to capture improvements in team skills and behaviours following simulation-based team training).

Regarding types of data collection tools used in surgical education research, tools tend to capture research participants' knowledge, their attitudes/views, and their skills/behaviours. Paper and pencil, or electronically administered knowledge tests are typically used to capture knowledge. Attitudes or perceptions are typically assessed via self-report questionnaires, ideally validated (i.e. that have been shown to adequately capture the attitudes, beliefs and perceptions of interest in a manner that allows scores derived from such questionnaires to be correlated with other measures). Such tools offer robust measurement of participants' subjective views. Finally, skills and behaviours can be captured via a variety of more objective metrics, which are typically obtained from direct observation of research participants and scoring of relevant scales. Observation can be carried out in real time, or retrospectively using audiovisual recording equipment and video/audio analysis. In recent years, with the advent of simulators and virtual reality technologies into surgical education, simulator-derived metrics are also used to capture technical or procedural skills. Finally, in the last 5 years or so, there has been a significant expansion of the evidence base on observational tools to capture behaviour in

surgical contexts – with validated tools such as the Observational Teamwork Assessment for Surgery (OTAS) (Sevdalis et al. 2009; Undre et al. 2007a) and Non-Technical Skills for Surgeons (NOTSS) (Yule et al. 2008) being increasingly used in research.

4.3.3 Quantitative Approaches to Data Analysis

Quantitative research typically utilises quantitative endpoints (see previous section), which therefore require quantitative handling and analyses. Descriptive statistics are typically used to summarise findings across groups, or interventions in this type of research. These statistics provide numerical information for the central tendency of the responses (e.g. mean, median and mode), as well as the dispersion or 'noise' in the collected data (e.g. variance and range). Graphical representations of such findings (including frequency plots, box plots, bar graphs, and scatterplots, amongst other options) often accompany numerical indices to visually illustrate findings. Statistical description is typically followed by statistical inference, in which statistical tests are applied to the data to discover quantitative patterns and trends in them and to determine whether such patterns are likely to hold for the populations from which the samples have been drawn (i.e. whether they are 'statistically significant'), or can be attributed to chance (i.e., random, non-systematic variations). Well known and widely used such tests include the following:

- *t*-test: a parametric test used to compare two independent or paired groups of subjects, or observations (parametric test make assumptions that may or may not be met by the data to hand)
- Analysis of variance: a parametric test used to compare three of more groups of subjects or observations, independent or paired
- Wilcoxon's test: a non-parametric equivalent to the paired samples *t*-test (non-parametric texts attempt to relax the assumptions about data and therefore have wider validity in dubious cases)
- Mann–Whitney test: a non-parametric equivalent to the independent samples *t*-test
- Correlation coefficients and regression analysis: a range of tests (Pearson *r* or Spearman's rho coefficients, simple or multiple regression, linear or logistic regression, etc.) that allow the researcher to establish whether there are significant relationships between different variables or measures of interest.

It is important to stress here that in quantitative research, statistical handling of the data and the relevant analyses should be thought through and specified at the early stages of a research project – with the design of the research. The choice of endpoints and outcome measures (e.g. multiple choice knowledge tests, standardised validated surveys or quantitative observation protocols) predetermines to a large extent the range of statistical analyses that can be used once the data have been collected.

4.4 Qualitative Surgical Education Research

Often a study may be designed from the outset to use both qualitative methods and statistical approaches – for different types of data related to the same phenomena. Qualitative research is also often thought of as being used in the preliminary stages of research, for example, to interview a group of trainee surgeons to seek opinion about the usefulness of the lecture teaching method for learning. The findings could help inform the topics and nature of questions to ask on a questionnaire. Alternatively, interviews can be used to explore the reasons and perceptions behind a finding from quantitative data. But qualitative approaches are also often used in their own right, simply to generate better understanding of phenomena that are suited to their approach (see Sect. 4.5.2).

Qualitative research is often lengthy in all its stages. Determining how to best collect the data and how to analyse it is vital. As with quantitatively based approaches, failure to consider these aspects at the start will often limit the type of analysis that can be conducted later. Collecting data can be time consuming, as can processing it. The amount of data generated is often enormous. Analysis is complex and can rarely rely so much upon a computer as is the case for statistics. Such research often requires comparatively lengthy writing up to ensure methods, analytical processes and how conclusions have been derived, are transparent; and as findings are usually discussed alongside preceding relevant literature, length is again likely. Discussions will tend to be quite discursive.

4.4.1 Key Approaches in Qualitative Research

Table 4.1 has already indicated many of the key characteristics of qualitative research. There are probably three major types of activity for which it is best used: the understanding of processes (e.g. how a trainee learns to make a particular type of flap, which might use observation and the interviewing of all parties involved, this might lead to improvements in training); exploring people's reactions, perceptions and opinions (e.g. 'what in your experience are the biggest challenges that learning minimally invasive techniques pose?', with a view to minimising these in training) and analysing documents (e.g. a report on future training requirements for surgeons, for example such a study might take a historical approach to such documents, tracing policy and practice shifts). Other key areas for qualitative research are as follows: evaluation studies that have an emphasis which qualitative research can address, and which typically draw conclusions about the impact of something; in depth comparative write-ups of circumstances/contexts (case studies); and action research in which the researcher investigates their own practice, usually through cycles of evaluation and amendment to practice. Words and what is seen are the focus of qualitative research, not numbers or precise measurement. There are innumerable general works that consider in some detail the main methods and approaches to qualitative research in education and the social sciences (e.g. Cohen et al. 2007;

Miles and Huberman 1994; Pope and Mays 2006; Robson 2002; Seale et al. 2006; Somekh and Lewin 2005); Kneebone and Fry (2009) consider how these methods may be used in surgical research and have a useful glossary explanation of key concepts in qualitative research.

Setting aside the analysis of documents, people, their actions and opinions are usually at the centre of qualitative research. For this reason, qualitative research has a strong consideration of ethical factors, including obtaining appropriate ethical permission to start the study, obtaining individual consent to take part, and with clear agreement about anonymity (or otherwise) and how data will be used, stored and published.

Rigour in qualitative research in best achieved and demonstrated by careful and appropriate selection of methods and modes of analysis; declaration of the position/possible influence on outcome of the researcher; triangulation of methods and findings by the use of multi-methods and perspectives; detailed description of the use of methods and analytical techniques; critical analysis of the strengths and weaknesses of the study design; the drawing together of ideas and conclusions into a model or theory with explanatory power that others recognise as an appropriate abstraction that also represents and aids the understanding of what was being researched.

4.4.2 Design and Methods

Selection of subjects or episodes for study is a tricky part of qualitative research. As with quantitative research, rarely can all examples of something be researched. Moreover it can be extremely difficult to create a truly random or representative sample, not least because such a sample might be too large for most qualitative studies to tackle; sometimes an opportunistic sample is all that can be achieved, in this case its likely representativeness or otherwise will typically be commented on when the researchers discuss the strengths and weaknesses of their study design. Some might even argue that it is pointless to try to create or hold up as desirable a representative sample, as all teachers and learners are different, patients and contexts also vary and trying to control for factors such as age, sex or length of training in one group of subjects, or interviewing every tenth trainee or trainer will not address the multiple layers of sampling issues. For many qualitative researchers, true sampling is therefore anathema; to undertake it is to misunderstand the strengths of what qualitative research can offer, and seeking to ape the quantitative paradigm risks losing the power, consistency and authenticity of the approach in use. However, many qualitative researchers do undertake purposive sampling, that is, they will seek out persons and situations so as to intentionally include a wide spectrum of types or likely variation across a range of parameters, based on previous literature and studies or personal knowledge of context.

Although there is probably wide recognition that face-to-face or telephone interviewing is the most commonly used method in qualitative research, there is

less recognition of the variety of types of interview that can be conducted. These may be not only individual or group but also tightly structured or much more free flowing. This means that there could be a set of fixed questions used consistently with all subjects or there could be questions that are little more than general starting points and which are followed up as seems sensible depending upon the responses, to try and elicit very detailed and thoughtful answers. Although questions should always be open and non-leading, more 'free' approaches will likely lead to more complex and less focused data – with obvious benefits and drawbacks. (Free response questions on questionnaires also produce qualitative data and in some cases may be little different from other types of interview – but they lack the opportunity for further exploration and follow up. Comment below about analysing qualitative data also applies to free response sections of questionnaires.) In this book, Chaps. 6, 10 and 12 especially draw substantially on interview techniques.

Interview data recorded as notes loses much that makes for the strengths of interviews – it loses complexity and much of the actual wording used, and may be less useful than written answers on a questionnaire, even when these are of limited length. Most qualitative researchers will wish if at all possible to record and transcribe interviews for these reasons, perhaps keeping notes as well about matters such as the body language of the interviewee. Transcription enables close study and coding of the text into themes, which can be recorded and ordered using proprietary software. Transcription will also seek to note hesitancy, tone of voice, etc. to capture as much as possible of the nuance of the response. Audio or video recording of interviews or other situations that may be observed, such as consultations, is often considered intrusive, and much is written about the possible impact this has on detracting from the 'naturalness' of the situation. Such intrusion always has to be commented on when the research is written up.

There are various specific techniques for eliciting opinion about carefully defined topics, which are quite different from the interview. The term focus group is often used as if it were synonymous with interviewing a group of people, this is not so. The focus group usually considers a single topic or issue and is heavily structured to reduce the influence of any dominant voices; moreover, its aim is to gradually sort and reduce views towards a single consensus output by means of a staged process that controls against the view emerging being that of the strongest personalities present (Moore 1987). The nominal group technique is a similar method but does not seek a consensus view from the group. The Delphi technique (conducted through written iterations that downplay minority views and rewrite statements in forms that seem to reflect a more consensual view) is a comparable technique that is not conducted face to face. Its purpose is to determine some form of majority view among the participants. This process has, for example, been used as a method of drawing together experts who are geographically dispersed to develop curricula (Paterson Davenport et al. 2004). Again it is a structured process that, for example, a video conference meeting would not replicate.

Observation studies can use both quantitative and qualitative methods, and many use both side by side. Qualitative approaches to observation are less widely recognised than interviews, but have much to contribute to the understanding of

process. They may also be followed up with interviews that seek further elucidation of what appear to be key incidents that occurred during the observation. They can be very time consuming, yield much data and pose particular challenges for analysis. Again, where something can be well recorded or filmed, the data may be particularly rich and can be studied many times over and in great depth; however, filming is not always possible for reasons of cost and ethics – and it may not actually capture all the nuances seen by a trained observer with a full view of everything (e.g. in a teaching episode on whom should the camera focus?). Participant observation may yield greater nuance and understanding, and more data. Human observers typically take free flowing notes or focus on particular features that are thought to be most pertinent to the study. All types of observation research run the risk of 'normal' behaviour being altered by the intrusion of the observer's eye, lens or microphone. Attempts to mitigate this with a camera usually take the form of making the instrument as unobtrusive as possible and with participant observation of greater immersion, not less, of the researcher in the context, so that they become almost an accepted part of the situation. A case in point here would be the famous 'boys in white' study of medical students (Becker 1976) or a more recent example from surgery (Bosk 2003). The methodology of observation studies draws heavily on ethnography.

The analysis of written documents and of discourse more generally is another area that might be described as a qualitative method (see Gill and Griffin 2010, for example), in that issues of interpretation cannot be ignored. There is overlap with how interview data (for example) is approached. Textual analysis can be undertaken in a range of ways that are reflective of some of the dilemmas and divisions at the heart of being a qualitative researcher (e.g. how far one imposes an external structure on written data and how far one seeks to have that structure emerge from the data itself). Other considerations that arise are similar to those that historians need to consider, for example, the bias of the writer. Some of these issues are explored further in the section below.

As with quantitative studies, piloting is usually a vital part of qualitative research and will be another feature that critical readers of such research will look for. Piloting, for example, would include trying out interview questions on two or three subjects and seeing if questions are clear, unambiguous, yield the complexity and depth of data required etc. It would also be checked that the proposed approach to analysis will work. A pilot also enables a check that appropriate subjects are being targeted. Adjustments will be made before the study resumes. This process helps to ensure the research is as effective as possible and resources are not committed to larger scale activity before it is warranted.

4.4.3 Analysis

There are many possible approaches to analysing qualitative data, but most fall into one of two camps. Taking a grounded theory approach (Glaser and Strauss 1967)

means that no preconceived concepts, fields or themes are imposed on the data. Instead the data is read and re-read closely, and emerging themes and concepts (ones that are 'grounded' in the data, not, for example in previously published studies) are noted. As more data is read and provisionally analysed, such emerging themes are refined and altered. The text is marked up or coded with these emergent categories in an iterative process. At the extreme of this approach, the signal to end data collection arises when no new categories or concepts are emerging from the data – when saturation is reached. Thus, in grounded theories, categorisations of ideas and actions emerge from the data through careful analysis.

An alternative approach is to come to the data with preconceived ideas of themes or categories and look for these in the data. Both approaches can have strengths and weaknesses and be used better and less well. The grounded approach generally requires more time. Both approaches can use software, or be conducted in more old fashioned ways using highlighter pens of different colours for each 'code' or literal cut and paste of transcribed text. It is partly because of the fundamental difference between a pre- and post-imposition of categorisation that careful description of methods is vital in qualitative research. The more inductive the approach the more vital it is to be able to return to the actual words and actions used to show how these verify one's research. Notes of interviews do not enable the same close analysis as recording and transcription.

However, analysis into major themes should not be the end of research. The hallmark of fully analysed data is the emergence of models and theories about the phenomena being researched. These will be abstractions, but are intended to be the embodiment of the context or situation being studied and will be based on the themes analysed out of the data. They are not models or theories that actors in the situation/process would themselves readily articulate, but should subsequently be recognisable as 'true' to that situation by participants. They usually offer insight. Sections 4.5.1 and 4.5.2 illustrate this feature of qualitative research. Qualitative research may or may not be able to make suggestions about how research findings can be used (and often has not had this aim), but may nonetheless yield pointers for further investigation and consideration.

4.4.4 Writing Up

It will be apparent that when using qualitative methods the influence of the researcher cannot be completely nullified (although neither can it in quantitative research, in that the researcher in both instances has made choices about methods and analysis that influence the output). The researcher sets the questions, does the coding, selects the extracts, may be present during data collection and affect the context etc. Qualitative research compensates for this usually through using different forms of triangulation, that is, through the use of multi-methods, multiple researchers comparing their analysis of the same data, checking findings and conclusion with study participants for 'recognition', etc.

Transparency is vital, in the sense of absolute clarity as to methods and modes of analysis. Being able to present depth and complexity is another necessary feature. Educational qualitative research is typically written up with reference to and use of existing research and theory that is pertinent; results and conclusions are not necessarily separated – nor is any of this necessarily separated from theory and already published research. This is quite unlike the approach taken in science papers. Qualitative research aims to generate insight and understanding. As has been mentioned, this is typically presented as themes, or ideally developed into models to present a representation of reality that participants will recognise, but would not have been conscious of themselves before the research (e.g. Stewart 2008). As has been noted, a critique of the research is also usually expected from the authors, pointing to strengths and limitations of the work.

4.5 Case Studies of Educational Research

The purpose of this section is to briefly describe a few pieces of educational research so as to 'put flesh' on the bones of methods and approaches described above.

4.5.1 Case Study 1: The Qualitative Paradigm

Meyer and Land are the authors of Chap. 6 of this book. They have developed a theory of learning that, to simplify, suggests that in each (sub) discipline area, there is a key concept that is crucial to understanding that subject but which often poses challenges for learners such that some learners can never get beyond it – a threshold concept. Meyer, Land and others have conducted further studies to investigate what these threshold concepts might be for a number of disciplines. In this book, Meyer and Land present their first attempt to explore threshold concepts in surgical education. They developed their ideas through a series of interviews with consultant and trainees surgeons. Through asking pertinent questions, analysing respondents' replies and 'applying' their theory they emerged with a suggestion for a threshold concept in surgery. They quote from their interviews in showing how this concept emerged; it is akin to the development of a model or theory as referred to above.

4.5.2 Case Study 2: The Qualitative Paradigm

Jane Stewart adopted a grounded theory approach to the data she collected about how junior doctors make decisions about calling out a senior (Stewart 2008). She used purposive sampling, semi-structured interviews and group presentation and 32 hours of direct observation. From these she constructed a complex model that

attempted to capture the criteria that governed their decisions ‘to call or not to call’. The importance of the research is that it elucidated not only the actual practice in these situations but also that the researcher was then able to draw out suggestions about how junior doctors hone their skills and decision-making capacity and the implications this has for education and training.

4.5.3 Case Study 3: The Quantitative Paradigm

Simulation-based training for health-care teams (especially in interventional specialties such as surgery, obstetrics and anaesthesia) has been highlighted in the surgical education literature as a key learning environment for trainee as well as expert clinicians to hone their skills in crisis management, as well as communication, leadership and other ‘nontechnical’ factors (Issenberg et al. 2005; Lynagh et al. 2007; Sutherland et al. 2006; Sturm et al. 2008). Working independently, a number of research teams across the world piloted simulation-based training modules for individual clinicians and entire teams (e.g. operating theatre teams or maternity care teams) and assessed their efficacy using participants’ attitudes to safety, and also validated, quantitative metrics of performance (Ellis et al. 2008; Moorthy et al. 2006; Paige et al. 2009; Robertson et al. 2009; Undre et al. 2007b). Following initial studies, prospective cross-sectional studies were subsequently carried out, with randomised controlled studies being increasingly reported in the international literature across a variety of specialities (Thomas et al. 2010). Learning outcomes are increasingly paired with clinical outcomes in such studies, aiming to demonstrate direct, quantitative translation of education and training into improved care processes, and ultimately, patient outcomes (Haller et al. 2008; Nielsen et al. 2007).

4.5.4 Case Study 4: Using Mixed Methods

In a recent series of interrelated studies, Arora and her colleagues have explored the impact of stress on the performance of surgeons, and how best to educate and train them to enable them to better cope with it in their clinical practice. This work started with reviewing the relevant evidence base, to determine what is known about the topic (Arora et al. 2010c). It then proceeded with a qualitative assessment, having conducted semi-structured interviews, of what surgeons feel about stressors in their environment, and what they would like a training and educational intervention to consist of (Arora et al. 2009). This work delivered a very rich context for an intervention to be delivered, with both senior and junior surgeons highlighting the role of systematic feedback on their performance as a key driver to learning and better coping with stress. The role of high-fidelity simulation environments also emerged as safe environments for learning and rehearsal of cognitive strategies to

manage stress. These findings led to the subsequent development and validation of a mental practice protocol for surgeons to mentally rehearse their responses to stressors (Arora et al. 2010a; Arora 2010d). This was done by developing the protocol, administering it to expert and novice surgeons (two independent groups design), and comparing statistically the quality of their self-reported mental imagery pre- and post-exposure to the protocol. Once the protocol was shown to improve both novice and expert surgeons' imagery, it was deployed within a context of full immersion simulation (two independent groups design again), in which surgeons had to deal with a range of stressors in their theatre environment. One group of surgeons was exposed to the protocol (intervention group) whereas a second group of surgeons was not (control group). Findings of this study showed that surgeons' imagery did improve as a result of the protocol, and so did their response to stress and their performance in the simulation – all measured using standardised, validated scales (Arora 2010d).

4.6 Conclusion

This chapter has set out some of the key features of the two main paradigms used in surgical education research. It has used examples and case studies to exemplify the use of these varied approaches. It has also attempted, very superficially, to indicate the differing philosophical underpinnings of these paradigms, their differing starting points and intents. The purpose of the chapter has been to provide a sufficient account of the techniques of educational research for readers unfamiliar with this area to appreciate the research and approaches referred to in this book.

References

Albert, M., & Reeves, S. (2010). Setting some new standards in medical education research. *Medical Education, 44*(7), 638–639.

Arora, S. (2010d). Stress, safety and surgical performance. Unpublished Doctoral Dissertation, Imperial College, London.

Arora, S., Sevdalis, N., Nestel, D., Tierney, T., Woloshynowych, M., & Kneebone, R. (2009). Managing intra-operative stress: What do surgeons want from a crisis training programme? *American Journal of Surgery, 197*, 537–543.

Arora, S., Sevdalis, N., Aggarwal, R., Sirimanna, P., Darzi, A., & Kneebone, R. (2010a). Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surgical Endoscopy, 24*, 2588–2593.

Arora, S., Tierney, T., Sevdalis, N., Aggarwal, R., Nestel, D., Woloshynowych, M., Darzi, A., & Kneebone, R. L. (2010b). The Imperial Stress Assessment Tool (ISAT): A feasible, reliable and valid approach to measuring stress in the operating room. *World Journal of Surgery, 34*, 1756–1763.

Arora, S., Sevdalis, N., Nestel, D., Woloshynowych, M., Darzi, A., & Kneebone, R. L. (2010c). The impact of stress on surgical performance: A systematic review of the literature. *Surgery, 147*, 318–330.

Becker, H. S. (1976). *Boys in white*. New Brunswick: Transaction Publishers.
Blalock, HM, Jr. (1979). *Social statistics* (Revised 2nd edn). New York: McGraw-Hill.
Bosk, C. L. (2003). *Forgive and remember. Managing medical failure*. Chicago: University of Chicago Press.
Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London: Routledge Falmer.
Ellis, D., Crofts, J. F., Hunt, L. P., Read, M., Fox, R., & James, M. (2008). Hospital, simulation center, and teamwork training for eclampsia management: A randomized controlled trial. *Obstetrics and Gynecology, 111*, 723–731.
Gill, D., & Griffin, A. (2009). Reframing medical education research: Let's make the publishable meaningful and the meaningful publishable. *Medical Education, 43*(10), 929–1024.
Gill, D., & Griffin, A. (2010). Good medical practice: What are we trying to say? Textual analysis using tag clouds. *Medical Education, 44*(3), 316–322.
Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. Chicago: Aldine.
Haller, G., Garnerin, P., Morales, M. A., Pfister, R., Berner, M., Irion, O., Clergue, F., & Kern, C. (2008). Effect of crew resource management training in a multidisciplinary obstetrical setting. *International Journal for Quality in Health Care, 20*, 254–263.
Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher, 27*, 10–28.
Kneebone, R., & Fry, H. (2009). Principles and methods in qualitative research. In T. Athanasiou, H. Debas, & A. Darzi (Eds.), *Key topics in surgical research and methodology* (pp. 243–254). Heidelberg: Springer. chapter 20.
Lynagh, M., Burton, R., & Sanson-Fisher, R. (2007). A systematic review of medical skills laboratory training: Where to from here? *Medical Education, 41*, 879–887.
Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks: Sage.
Moore, C. M. (1987). *Group techniques for idea building*. Newbury Park: Sage.
Moorthy, K., Munz, Y., Forrest, D., Pandey, V., Undre, S., Vincent, C., & Darzi, A. (2006). Surgical crisis management skills training and assessment: A simulation-based approach to enhancing operating room performance. *Annals of Surgery, 244*, 139–147.
Nielsen, P. E., Goldman, M. B., Mann, S., Shapiro, D. E., Marcus, R. G., Pratt, S. D., Greenberg, P., McNamee, P., Salisbury, M., Birnbach, D. J., Gluck, P. A., Pearlman, M. D., King, H., Tornberg, D. N., & Sachs, B. P. (2007). Effects of teamwork training on adverse outcomes and process of care in labor and delivery: A randomized controlled trial. *Obstetrics and Gynecology, 109*, 48–55.
Paige, J. T., Kozmenko, V., Yang, T., Paragi Gururaja, R., Hilton, C. W., Cohn, I., Jr., & Chauvin, S. W. (2009). High-fidelity, simulation-based, interdisciplinary operating room team training at the point of care. *Surgery, 145*, 138–146.
Paterson Davenport, L. A., Hesketh, E. A., Macpherson, S. G., & Harden, R. M. (2004). Exit outcomes for the PRHO year: An evidence base for informed discussion. *Medical Education, 38*(1), 67–80.
Pope, C., & Mays, N. (2006). *Qualitative research in health care* (3rd ed.). London: BMJ Books/Blackwell Publishing.
Robertson, B., Schumacher, L., Gosman, G., Kanfer, R., Kelley, M., & DeVita, M. (2009). Simulation-based crisis team training for multidisciplinary obstetric providers. *Simulation in Healthcare, 4*, 77–83.
Robson, C. (2002). *Real world research* (2nd ed.). Oxford: Blackwell.
Seale, C., Gobo, G., Gubrium, J., & Silverman, D. (Eds.). (2006). *Qualitative research practice*. Thousand Oaks: Sage.
Sevdalis, N., Davis, R. E., Koutantji, M., Undre, S., Darzi, A., & Vincent, C. A. (2008). Reliability of a revised NOTECHS scale for use in surgical teams. *American Journal of Surgery, 196*, 184–190.

Sevdalis, N., Lyons, M., Healey, A. N., Undre, S., Darzi, A., & Vincent, C. A. (2009). Observational teamwork assessment for surgery: Construct validation with expert vs. novice raters. *Annals of Surgery, 249*, 1047–1051.

Shadish, W. R., Cook, T. D., & Cambell, D. T. (2002). *Experimental and quasi experimental design for generalized causal inference*. Boston: Houghton-Mifflin.

Somekh, B., & Lewin, C. (Eds.). (2005). *Research methods in the social sciences*. Thousand Oaks: Sage.

Stewart, J. (2008). To call or not to call: A judgement of risk by pre-registration house officers. *Medical Education, 42*(9), 938–944.

Sturm, L. P., Windsor, J. A., Cosman, P. H., Cregan, P., Hewett, P. J., & Maddern, G. J. (2008). A systematic review of skills transfer after surgical simulation training. *Annals of Surgery, 248*, 166–179.

Sutherland, L. M., Middleton, P. F., Anthony, A., Hamdorf, J., Cregan, P., Scott, D., & Maddern, G. J. (2006). Surgical simulation: A systematic review. *Annals of Surgery, 243*, 291–300.

Swanwick, T. (Ed.). (2010). *Understanding medical education: Evidence, theory and practice*. Oxford: Wiley-Blackwell.

Teunissen, P. W. (2010). On the transfer of theory to the practice of research and education. *Medical Education, 44*(6), 534–535.

Thomas, E. J., Williams, A. L., Reichman, E. F., Lasky, R. E., Crandell, S., & Taggart, W. R. (2010). Team training in the neonatal resuscitation program for interns: Teamwork and quality of resuscitations. *Pediatrics, 125*, 539–546.

Undre, S., Sevdalis, N., Healey, A. N., & Vincent, C. A. (2007a). The Observational Teamwork Assessment for Surgery (OTAS): Refinement and application in urological surgery. *World Journal of Surgery, 3*, 1373–1381.

Undre, S., Koutantji, M., Sevdalis, N., Selvapatt, N., Williams, S., Gautama, S., McCulloch, P., Darzi, A., & Vincent, C. A. (2007b). Multi-disciplinary crisis simulations: The way forward for training surgical teams. *World Journal of Surgery, 3*, 1843–1853.

Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G., & Paterson-Brown, S. (2008). Surgeons' non-technical skills in the operating room: Reliability testing of the NOTSS behaviour rating system. *World Journal of Surgery, 32*, 548–556.

Zeller, R. A., & Carmines, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge: Cambridge University Press.