

1 An introduction to research methods and traditions

Scientists should not be ashamed to admit . . . that hypotheses appear in their minds along uncharted byways of thought; that they are imaginative and inspirational in character; that they are indeed adventures of the mind.

(Peter Medawar, 1963, “Is the Scientific Paper a Fraud?” BBC Presentation)

This book is essentially practical in nature. It is intended as an introduction to research methods in applied linguistics, and does not assume specialist knowledge of the field. It is written in order to help you to develop a range of skills, but more particularly to discuss and critique a wide range of research methods, including formal experiments and quasi-experiments; elicitation instruments; interviews and questionnaires; observation instruments and schedules; introspective methods, including diaries, logs, journals, protocol analysis, and stimulated recall; interaction and transcript analysis; ethnography and case studies. Having read the book, you should have a detailed appreciation of the basic principles of research design, and you should be able to read and critique published studies in applied linguistics. In relation to your own teaching, you should be better able to develop strategies for formulating questions, and for collecting and analysing data relating to those questions.

The purpose of this initial chapter is to introduce you to research methods and traditions in applied linguistics. The chapter sets the scene for the rest of the book, and highlights the central themes underpinning the book. This chapter deals with the following questions:

- What is the difference between *quantitative* and *qualitative research*?
- What do we mean by ‘the status of knowledge’, and why is this of particular significance to an understanding of research traditions?
- What is meant by the terms *reliability* and *validity*, and why are they considered important in research?
- What is *action research*?

Research traditions in applied linguistics

The very term *research* is a pejorative one to many practitioners, conjuring up images of white-coated scientists plying their arcane trade in laboratories filled with mysterious equipment. While research, and the conduct of

Research methods in language learning

research, involves rigour and the application of specialist knowledge and skills, this rather forbidding image is certainly not one I wish to present here.

I recently asked a group of graduate students who were just beginning a research methods course to complete the following statements: 'Research is ...' and 'Research is carried out in order to ...' Here are some of their responses.

Research is:

- about inquiry. It has two components: process and product. The process is about an area of inquiry and how it is pursued. The product is the knowledge generated from the process as well as the initial area to be presented.
- a process which involves (a) defining a problem, (b) stating an objective, and (c) formulating an hypothesis. It involves gathering information, classification, analysis, and interpretation to see to what extent the initial objective has been achieved.
- undertaking structured investigation which hopefully results in greater understanding of the chosen interest area. Ultimately, this investigation becomes accessible to the 'public'.
- an activity which analyses and critically evaluates some problem.
- to collect and analyse the data in a specific field with the purpose of proving your theory.
- evaluation, asking questions, investigations, analysis, confirming hypotheses, overview, gathering and analysing data in a specific field according to certain predetermined methods.

Research is carried out in order to:

- get a result with scientific methods objectively, not subjectively.
- solve problems, verify the application of theories, and lead on to new insights.
- enlighten both researcher and any interested readers.
- prove/disprove new or existing ideas, to characterise phenomena (i.e., the language characteristics of a particular population), and to achieve personal and community aims. That is, to satisfy the individual's quest but also to improve community welfare.
- prove or disprove, demystify, carry out what is planned, to support the point of view, to uncover what is not known, satisfy inquiry. To discover the cause of a problem, to find the solution to a problem, etc.

Certain key terms commonly associated with research appear in these characterisations. These include: inquiry, knowledge, hypothesis, information, classification, analysis, interpretation, structured investigation, understanding, problem, prove, theory, evaluation, asking questions, analysing data, scientific method, insight, prove/disprove, characterise phenomena, demystify, uncover, satisfy inquiry, solution. The terms, taken together, suggest that research is a process of formulating questions, problems, or hypotheses; col-

An introduction to research methods and traditions

lecting data or evidence relevant to these questions/problems/hypotheses; and analysing or interpreting these data. The minimal definition to which I shall adhere in these pages is that *research* is a systematic process of inquiry consisting of three elements or components: (1) a question, problem, or hypothesis, (2) data, (3) analysis and interpretation of data. Any activity which lacks one of these elements (for example, data) I shall classify as something other than research. (A short definition of key terms printed in *italic* can be found in the glossary at the end of the book.)

Traditionally, writers on research traditions have made a binary distinction between qualitative and quantitative research, although more recently it has been argued that the distinction is simplistic and naive. Reichardt and Cook (cited in Chaudron 1988), for example, argue that in practical terms, qualitative and quantitative research are in many respects indistinguishable, and that 'researchers in no way follow the principles of a supposed paradigm without simultaneously assuming methods and values of the alternative paradigms' (Reichardt and Cook 1979: 232). Those who draw a distinction suggest that quantitative research is obtrusive and controlled, objective, generalisable, outcome oriented, and assumes the existence of 'facts' which are somehow external to and independent of the observer or researcher. Qualitative research, on the other hand, assumes that all knowledge is relative, that there is a subjective element to all knowledge and research, and that holistic, ungeneralisable studies are justifiable (an ungeneralisable study is one in which the insights and outcomes generated by the research cannot be applied to contexts or situations beyond those in which the data were collected). In metaphorical terms, quantitative research is 'hard' while qualitative research is 'soft'. Terms (sometimes used in approbation, sometimes as abuse) commonly associated with the two paradigms are set out in Figure 1.1.

In an attempt to go beyond the binary distinction between qualitative and quantitative research, Chaudron (1988) argues that there are four research traditions in applied linguistics. These are the psychometric tradition, interaction analysis, discourse analysis, and ethnography. Typically, *psychometric* investigations seek to determine language gains from different methods and materials through the use of the 'experimental method' (to be dealt with in detail in Chapter 2). *Interaction analysis* in classroom settings investigates such relationships as the extent to which learner behaviour is a function of teacher-determined interaction, and utilises various observation systems and schedules for coding classroom interactions. *Discourse analysis* analyses classroom discourse in linguistic terms through the study of classroom transcripts which typically assign utterances to predetermined categories. Finally, *ethnography* seeks to obtain insights into the classroom as a cultural system through naturalistic, 'uncontrolled' observation and description (we shall deal with ethnography in Chapter 3). While Chaudron's aim of attempting to transcend the traditional binary distinction is a worthy one, it could be argued that discourse analysis and interaction analysis are methods of data

Research methods in language learning

Qualitative research

Advocates use of qualitative methods
 Concerned with understanding human behaviour from the actor's own frame of reference
 Naturalistic and uncontrolled observation
 Subjective
 Close to the data: the 'insider' perspective
 Grounded, discovery-oriented, exploratory, expansionist, descriptive, and inductive
 Process-oriented
 Valid: 'real', 'rich', and 'deep' data
 Ungeneralisable: single case studies
 Assumes a dynamic reality

Quantitative research

Advocates use of quantitative methods
 Seeks facts or causes of social phenomena without regard to the subjective states of the individuals
 Obtrusive and controlled measurement
 Objective
 Removed from the data: the 'outsider' perspective
 Ungrounded, verification-oriented, confirmatory, reductionist, inferential, and hypothetical-deductive
 Outcome-oriented
 Reliable: 'hard' and replicable data
 Generalisable: multiple case studies
 Assumes a stable reality

Figure 1.1 Terms commonly associated with quantitative and qualitative approaches to research (adapted from Reichardt and Cook 1979)

collection rather than distinct research traditions in their own right. In fact these methods can be (and have been) utilised by researchers working in both the psychometric and ethnographic traditions. For example, ethnographers can use interaction analysis checklists to supplement their naturalistic observations, while psychometric research can use similar schemes to identify and measure distinctions between different classrooms, teaching methods, approaches, and teachers (the studies reported by Spada 1990 are excellent examples of such research).

Grotjahn (1987) provides an insightful analysis of research traditions in applied linguistics. He argues that the qualitative-quantitative distinction is an oversimplification and that, in analysing actual research studies, it is necessary to take into consideration the method of data collection (whether the data have been collected experimentally or non-experimentally); the type of data yielded by the investigation (qualitative or quantitative); and the type of analysis conducted on the data (whether statistical or interpretive). Mixing and matching these variables provides us with two 'pure' research paradigms. Paradigm 1 is the 'exploratory-interpretive' one which utilises a non-experimental method, yields qualitative data, and provides an interpretive analysis of that data. The second, or 'analytical-nomological' paradigm, is one in which the data are collected through an experiment, and yields quantitative data which are subjected to statistical analysis. In addition to these 'pure' forms, there are six 'mixed' paradigms which mix and match the three variables in different ways. For example, there is an 'experimental-qualitative-interpretive' paradigm which utilises an experiment but yields qualitative

An introduction to research methods and traditions

data, which are analysed interpretively. The different research paradigms resulting from mixing and matching these variables are set out in Figure 1.2. (It should be pointed out that, while all of these various ‘hybrid’ forms are theoretically possible, some are of extremely unlikely occurrence. For example, it would be unusual for a researcher to go to the trouble of setting up a formal experiment yielding quantitative data which are analysed interpretively.)

While I accept Grotjahn’s assertion that in the execution of research the qualitative-quantitative distinction is relatively crude, I still believe that the distinction is a real, not an ostensible one, and that the two ‘pure’ paradigms are underpinned by quite different conceptions of the nature and status of knowledge. Before turning to a discussion of this issue, however, I should like to outline a model developed by van Lier (1988; 1990) for characterising applied linguistic research.

Van Lier argues that applied linguistic research can be analysed in terms of two parameters: an interventionist parameter and a selectivity parameter. Research is placed on the interventionist parameter according to the extent to which the researcher intervenes in the environment. A formal experiment which takes place under laboratory conditions would be placed at one end of the interventionist continuum/parameter, while a naturalistic study of a classroom in action would be placed at the other end of the continuum. The other parameter places research according to the degree to which the researcher prespecifies the phenomena to be investigated. Once again, a formal experiment, in which the researcher prespecifies the variables being focused on, would be placed at one end of the continuum, while an ethnographic ‘portrait’ of a classroom in action would occur at the other end of the continuum. Figure 1.3 illustrates the relationship between these two parameters.

The intersection of these two parameters creates four ‘semantic spaces’: a ‘controlling’ space, a ‘measuring’ space, an ‘asking/doing’ space, and a ‘watching’ space. The controlling space, which is characterised by a high degree of intervention and a high degree of control, contains studies in which the experimenters focus their attention on a limited number of variables and attempt to control these in some way. For example, in an investigation into the effect of cultural knowledge on reading comprehension, the investigator may set up an experiment in which subjects from different cultural backgrounds read texts in which the content is derived from their own and other cultures. In such an experiment, the focus is on a single variable (cultural background) which is controlled through the reading texts administered to the subjects.

The measuring space encloses those research methods involving a high degree of selection but a low degree of control. ‘One selects certain features, operationally defines them, and quantifies their occurrence, in order to establish a relationship between features, or between features and other things,

Research methods in language learning

PURE FORMS

Paradigm 1: exploratory-interpretive

- 1 non-experimental design
- 2 qualitative data
- 3 interpretive analysis

Paradigm 2: analytical-nomological

- 1 experimental or quasi-experimental design
- 2 quantitative data
- 3 statistical analysis

MIXED FORMS

Paradigm 3: experimental-qualitative-interpretive

- 1 experimental or quasi-experimental design
- 2 qualitative data
- 3 interpretive analysis

Paradigm 4: experimental-qualitative-statistical

- 1 experimental or quasi-experimental design
- 2 qualitative data
- 3 statistical analysis

Paradigm 5: exploratory-qualitative-statistical

- 1 non-experimental design
- 2 qualitative data
- 3 statistical analysis

Paradigm 6: exploratory-quantitative-statistical

- 1 non-experimental design
- 2 quantitative data
- 3 statistical analysis

Paradigm 7: exploratory-quantitative-interpretive

- 1 non-experimental design
- 2 quantitative data
- 3 interpretive analysis

Paradigm 8: experimental-quantitative-interpretive

- 1 experimental or quasi-experimental design
- 2 quantitative data
- 3 interpretive analysis

Figure 1.2 Types of research design (from Grotjahn 1987: 59–60)

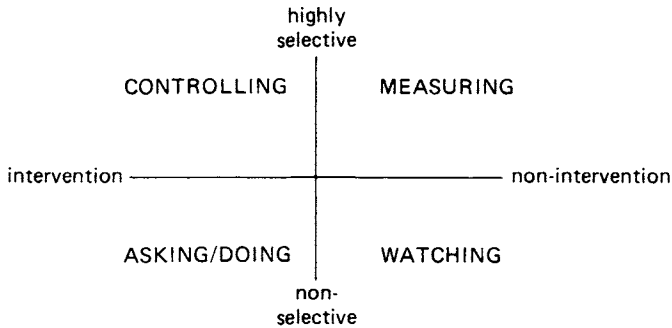
An introduction to research methods and traditions

Figure 1.3 Parameters in research design (after van Lier 1988)

such as educational outcomes' (van Lier 1990: 34). For example, the researcher may be interested in the effect of teacher questions on student responses. Armed with a taxonomy of teacher questions, the researcher observes a series of classes, documenting the types of questions asked and the length and complexity of the responses. Here the researcher is highly selective in what he or she chooses to look at or for, but does not attempt to control the behaviour of either the teacher or the students.

The asking/doing space contains studies in which there is a high degree of intervention, but a low degree of control. 'One investigates certain problem areas by probing, trying out minor changes, asking for participants' views and concerns, and so on. After a while it may be possible to pinpoint the problem so precisely that a controlled environment can be created in order to conduct an experiment, thus moving from [asking/doing] through watching to controlling. On the other hand, increased understanding through interpretation can also make experimentation unnecessary' (van Lier 1990: 34–35).

The final semantic space, watching, is characterised by a lack of selectivity and a lack of intervention. The researcher observes and records what happens without attempting to interfere with the environment. Additionally, the researcher does not decide which variables are of interest or of potential significance before engaging in the research. While some form of quantification or measurement may be used, it is seen as no more than one tool among many, and not inherently superior to any other way of analysing data. An example of a study fitting into this final semantic space would be one in which the researcher wishes to provide a descriptive and interpretive portrait of a school community as its members go about their business of living and learning together.

I find van Lier's model of types of research a useful one, although, as van Lier himself points out, it is a simplification of what really happens when research is carried out. In reality, a particular piece of research may well tran-

Research methods in language learning

scend its initial 'semantic space'. An investigation may well begin in the 'watching' space, and then, as issues emerge, the focus may become narrower. The researcher may then decide to establish a formal experiment to test an hypothesised relationship between two or more variables. In this instance, the research will have moved from the 'watching' space to the 'controlling' space. Regardless of the fact that it is a simplification, it does serve to highlight two of the most important questions researchers must confront at the beginning of their research, namely:

- To what extent should I attempt to prespecify the phenomena under investigation?
- To what extent should I attempt to isolate and control the phenomena under investigation?

Brown (1988) provides a very different introduction to research from van Lier, being principally concerned with quantitative research. In his framework for analysing types of research, he draws a distinction between primary and secondary research. Secondary research consists of reviewing the literature in a given area, and synthesising the research carried out by others. Normally, this is a necessary prerequisite to primary research, which 'differs from secondary research in that it is derived from the primary sources of information (e.g., a group of students who are learning a language), rather than from secondary sources (e.g., books about students who are learning a language)' (1988: 1). Hence, it has the advantage of being closer to the primary source of information. Primary research is subdivided into case studies and statistical studies. Case studies centre on a single individual or limited number of individuals, documenting some aspect of their language development, usually over an extended period of time. Statistical studies, on the other hand, are basically cross-sectional in nature, considering 'a group of people as a cross section of possible behaviors at a particular point or at several distinct points in time. In addition, statistical analyses are used in this approach to estimate the probability, or likelihood, that the results did not occur by chance alone' (p. 3). In Brown's model, statistical studies are further subdivided into survey studies and experimental studies. Survey studies investigate a group's attitudes, opinions, or characteristics, often through some form of questionnaire. Experimental studies, on the other hand, control the conditions under which the behaviour under investigation is observed.

For instance, a researcher might wish to study the effects of being male or female on students' performance on a language placement test. Such research might involve administering the test to the students, then separating their scores into two groups according to gender, and finally studying the similarities and differences in behavior between the two groups. Another type of experimental study might examine the relationship between students' scores on a language aptitude test and their actual

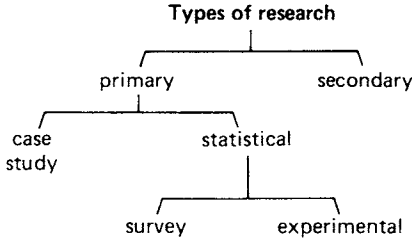
An introduction to research methods and traditions

Figure 1.4 *Types of research (after Brown 1988)*

performance in language classes, as measured by course grades. Experimental studies, then, can be varied in the types of questions being asked . . . (p. 3)

Brown's characterisation of types of research is set out in Figure 1.4.

According to Brown, experimental research should exhibit several key characteristics. It should be systematic, logical, tangible, replicable, and reductive, and one should be cautious of any study not exhibiting these characteristics. A study is *systematic* if it follows clear procedural rules for the design of the study, for guarding against the various threats to the internal and external validity of the study, and for the selection and application of statistical procedures. A study should also exhibit *logic* in the step-by-step progression of the study. *Tangible* research is based on the collection of data from the real world. 'The types of data are numerous, but they are all similar in that they must be *quantifiable*, that is, each datum must be a number that represents some well-defined quantity, rank, or category' (p. 4). *Replicability* refers to the ability of an independent researcher to reproduce the study under similar conditions and obtain the same results. In order for a reader to evaluate the replicability of a study, it should be presented clearly and explicitly. *Reductivity* is explained in the following way: '... statistical research can reduce the confusion of facts that language and language teaching frequently present, sometimes on a daily basis. Through doing or reading such studies, you may discover new patterns in the facts. Or through these investigations and the eventual agreement among many researchers, general patterns and relationships may emerge that clarify the field as a whole' (p. 5). Most of these characteristics can ultimately be related to issues of validity and reliability, and we shall look in detail at these critical concepts later in the chapter. Table 1.1 summarises the key characteristics of good experimental research according to Brown.

In this section I have reviewed the recent literature on research traditions in applied linguistics. My main point here is that, while most commentators reject the traditional distinction between qualitative and quantitative research as being simplistic and naive, particularly when it comes to the anal-

Research methods in language learning

TABLE 1.1 CHARACTERISTICS OF GOOD EXPERIMENTAL RESEARCH

<i>Characteristic</i>	<i>Key question</i>
Systematic	Does the study follow clear procedural rules?
Logical	Does the study proceed in a clear step-by-step fashion, from question formation to data collection and analysis?
Tangible	Are data collected from the real world?
Replicable	Could an independent researcher reproduce the study?
Reductive	Does the research establish patterns and relationships among individual variables, facts, and observable phenomena?

Source: Based on Brown (1988).

ysis of published research, the distinction between the research traditions persists. Ultimately, most researchers will admit to subscribing to one tradition rather than another. How, then, are we to account for the persistence of a distinction which has been so widely criticised?

The status of knowledge

One reason for the persistence of the distinction between quantitative and qualitative research is that the two approaches represent different ways of thinking about and understanding the world around us. Underlying the development of different research traditions and methods is a debate on the nature of knowledge and the status of assertions about the world, and the debate itself is ultimately a philosophical one. It is commonly assumed that the function of research is to add to our knowledge of the world and to demonstrate the 'truth' of the commonsense notions we have about the world. (You might recall the statements made by students of research methods, some of which are reproduced at the beginning of this chapter.) In developing one's own philosophy on research, it is important to determine how the notion of 'truth' relates to research. What is truth? (Even more basically, do we accept that there is such a thing as 'truth?') What is evidence? Can we ever 'prove' anything? What evidence would compel us to accept the truth of an assertion or proposition? These are questions which need to be borne in mind constantly as one reads and evaluates research.

In a recent television advertising campaign, the following claim was made about a popular brand of toothpaste: 'University tests prove that Brand X toothpaste removes 40% more plaque'. (The question of 40% more than what is not addressed.) By invoking the authority of 'university tests' the manufacturers are trying to invest their claim with a status it might otherwise lack. There is the implication that claims based on research carried out in

An introduction to research methods and traditions

universities are somehow more 'scientific' and therefore believable than claims made on the basis of anecdotes, the experience of the layperson, or the in-house research of the manufacturers themselves. According to Winograd and Flores (1986), the status of research based on 'scientific' experiments and, indeed, the rationalist orientation which underlies it, is based on the success of modern science.

The rationalist orientation . . . is also regarded, perhaps because of the prestige and success that modern science enjoys, as the very paradigm of what it means to think and be intelligent. . . . It is scarcely surprising, then, that the rationalistic orientation pervades not only artificial intelligence and the rest of computer science, but also much of linguistics, management theory, and cognitive science . . . rationalistic styles of discourse and thinking have determined the questions that have been asked and the theories, methodologies, and assumptions that have been adopted. (p. 16)

The following assertions have all been made publicly. You might like to consider these, and the evidence on which they are based, and reflect on which deserve to be taken seriously on the balance of the evidence provided.

ASSERTION 1

Second language learners who identify with the target culture will master the language more quickly than those who do not. (Evidence: A case study of an unsuccessful language learner.)

ASSERTION 2

Schoolchildren are taught by their teachers they they need not obey their parents. (Evidence: A statement by a parent on a radio talk-back program.)

ASSERTION 3

Immigrants are more law abiding than native-born citizens. (Evidence: An analysis of district court records.)

ASSERTION 4

Deaf children are more successful in school if their parents do not succumb to a sense of powerlessness when they experience difficulty communicating with their children. (Evidence: A study based on data from 40 deaf and 20 hearing children.)

ASSERTION 5

Affective relationships between teacher and students influence proficiency gains. (Evidence: A longitudinal ethnographic study of an inner city high school class.)

Research methods in language learning

ASSERTION 6

Students who are taught formal grammar develop greater proficiency than students who are taught through 'immersion' programs. (Evidence: A formal experiment in which one group of students was taught through immersion and another group was taught formal grammar.)

In actual fact, all of these assertions can be challenged on the basis of the evidence advanced to support them. Some critics would reject assertions 1, 2, and 5 on the grounds that they are based on a single instance (in the case of 1 and 2 on the instance of a single individual, and in the case of 5 on the instance of a single classroom). Such critics would argue that the selection of a different individual or classroom might have yielded a very different, even contradictory, response. (We shall return to the issues of 'representativeness' and 'typicality' of data again in later chapters, particularly Chapter 3 on ethnography, and Chapter 4 on case study.) Assertion 3 could be challenged on the grounds that the causal relationship between fewer court convictions and demographic data has not been demonstrated. (It might simply be, for example, that criminals from immigrant communities are smarter, and therefore less likely to be caught than native-born criminals.) The problem with this study is that we can account for the outcomes through explanations other than the one offered by the researchers. Someone versed in research methods would say that the study has poor internal validity. (We shall look at the question of validity in the next section.) Assertion 4 might be criticised on the grounds that 'power' and 'powerlessness' have not been adequately defined. Such a criticism is aimed at the construct validity of the study. (We shall also look at issues related to constructs and construct validity in the next section.) The final assertion can be challenged on the grounds that the two groups might not have been equal to begin with.

In the final analysis, the extent to which one is prepared to accept or reject particular methods of inquiry and the studies utilising these methods will depend on one's view of the world, and the nature of knowledge. For some people the notion that there are external truths 'out there' which are independent of the observer is self-evident. For others, this notion, which underlies the quantitative approach to research, is questionable (see, for example, Winograd and Flores 1986).

Some key concepts in research

In this section, we shall look in greater detail at some key concepts which have to this point only been touched on in passing. We shall look in particular at the concepts of reliability and validity. First, however, I should like briefly to discuss two other terms. These are *deductivism* and *inductivism*.

An introduction to research methods and traditions

Two procedures open to researchers are inductivism and deductivism. *Deductive research* begins with an hypothesis or theory and then searches for evidence either to support or refute that hypothesis or theory. *Inductivism* seeks to derive general principles, theories, or 'truths' from an investigation and documentation of single instances. Numerous commentators have criticised what is called naive inductivism (see Chalmers 1982), which is the belief that we can arrive at the 'truth' by documenting instances of the phenomenon under investigation. Popper (1968, 1972) illustrated the naivety of inductivism with his celebrated swan example. He pointed out that we are never entitled to make the claim that 'All swans are white', regardless of the number of sightings of white swans. Though we may have sighted one thousand white swans, there is nothing to say that the one thousand and first sighting will not be a black swan. This led Popper to advance his falsificationist principle. This principle states that while we can never conclusively demonstrate truth through induction, we can in fact falsify an assertion through the documentation of a single disconfirming instance (as in the case of the black swan). According to Popper, all hypotheses should therefore be formulated in a way which enables them to be falsified through a single disconfirming instance. Taken to its logical conclusion, this view would have it that all knowledge is tentative and that, in fact, 'absolute truth' is an ideal which can never be attained.

Chalmers (1982) introduces the falsificationist's position in the following manner:

According to falsificationism, some theories can be shown to be false by an appeal to the results of observation and experiment. I have already indicated in Chapter 2 that, even if we assume that true observational statements are available to us in some way, it is never possible to arrive at universal laws and theories by logical deductions on that basis alone. On the other hand, it is possible to perform logical deductions starting from singular observation statements as premises, to arrive at the falsity of universal laws and theories by logical deduction. . . . The falsificationist sees science as a set of hypotheses that are tentatively proposed with the aim of accurately describing or accounting for the behaviour of some aspect of the world or universe. However, not any hypothesis will do. There is one fundamental condition that any hypothesis or system of hypotheses must satisfy if it is to be granted the status of a scientific law or theory. If it is to form part of science, an hypothesis must be falsifiable. (pp. 38–39)

The argument that progress in applied linguistics should be through the formulation and testing of hypotheses which are falsifiable has been advanced by numerous researchers. Pienemann and Johnston (1987) mount a vigorous attack on a major and influential research program in applied linguistics on the basis that it is not falsifiable. McLaughlin (1987) also argues that falsifiability or disconfirmation is the most important means to achieving scientific progress in applied linguistics.

Research methods in language learning

In any scientific endeavour the number of potentially positive hypotheses very greatly exceeds the number of hypotheses that in the long run will prove to be compatible with observations. As hypotheses are rejected, the theory is either disconfirmed or escapes from being disconfirmed. The results of observation ‘probe’ but do not ‘prove’ a theory. An adequate hypothesis is one that has repeatedly survived such probing – but it may always be displaced by a new probe. (McLaughlin 1987: 17)

In reality, comparatively few hypotheses in applied linguistics can be demolished by a single disconfirming instance. In most cases we are interested in general trends and statistical tendencies rather than universal statements. Even researchers who claim their research is falsifiable have ways of protecting their theories from attack. For example, some second language acquisition researchers (see, for example, Pienemann and Johnston 1987) claim that the morphosyntax of all learners of English as a second language passes through certain developmental stages. These stages are defined in terms of the morphosyntactic items that learners are able to control at a particular stage, which in turn are governed by speech-processing constraints. According to the researchers, it is impossible for learners to ‘skip’ a stage, and if a single learner were to be found who had mastered, say, a stage 4 grammatical item while still at stage 2, then the developmental hypothesis would have been falsified. In fact, when such instances occur, it may be claimed that the learners in question have not really internalised the item but are using it as a formulaic utterance. Given the difficulty in determining with certainty whether or not an item is or is not a formulaic utterance, it is highly unlikely that the theory will ever be falsified.

Two terms of central importance to research are *reliability* and *validity*, and I shall return to these repeatedly in the course of this book. *Reliability* refers to the consistency of the results obtained from a piece of research. *Validity*, on the other hand, has to do with the extent to which a piece of research actually investigates what the researcher purports to investigate. It is customary to distinguish between internal and external reliability and validity, and I shall deal with each of these briefly in this section. The description and analysis provided here is developed and extended in subsequent chapters.

Reliability refers to the consistency and replicability of research. *Internal reliability* refers to the consistency of data collection, analysis, and interpretation. *External reliability* refers to the extent to which independent researchers can reproduce a study and obtain results similar to those obtained in the original study. In a recent investigation into classroom interaction, one of my graduate students coded the interactions of three teachers and their students using an observation schedule developed for that purpose. I also coded a sample of the interactions independently. When the student and I compared the categories to which we had assigned interactions, we found that we were in agreement in 95% of the cases. We took this high level of agreement as an

An introduction to research methods and traditions

indication that this aspect of the study had high internal reliability. If a second graduate student were to conduct the study a second time and obtain the same results, we could claim that the study was externally reliable. (This 'inter-rater reliability' procedure is but one way of guarding against threats to the internal reliability of a study. We shall consider alternative procedures in Chapter 3.)

There are two types of validity: internal validity and external validity. *Internal validity* refers to the interpretability of research. In experimental research, it is concerned with the question: Can any differences which are found actually be ascribed to the treatments under scrutiny? *External validity* refers to the extent to which the results can be generalised from samples to populations. Researchers must constantly be alive to the potential and actual threats to the validity and reliability of their work. Table 1.2 provides two sample studies which illustrate the threats to validity posed by poor research design.

One of the problems confronting the researcher who wishes to guard against threats to external and internal validity is that measures to strengthen internal validity may weaken external validity and vice versa, as Beretta has shown.

Internal validity has to do with factors which may directly affect outcomes, while external validity is concerned with generalisability. If all variables, such as treatments and sampling of subjects, are controlled, then we might say that laboratory conditions pertain and that the experiment is more likely to be internally valid. However, what occurs under such conditions may not occur in typical circumstances, and the question arises as to how far we may generalise from the results. (Beretta 1986a: 297)

However, if the researcher carried out the study in context, this may increase the external validity but weaken the internal validity.

In addition to internal and external validity, researchers need to pay close attention to *construct validity*. A construct is a psychological quality, such as intelligence, proficiency, motivation, or aptitude, that we cannot directly observe but that we assume to exist in order to explain behaviour we can observe (such as speaking ability, or the ability to solve problems). It is extremely important for researchers to define the constructs they are investigating in a way which makes them accessible to the outside observer. In other words, they need to describe the characteristics of the constructs in a way which would enable an outsider to identify these characteristics if they came across them. If researchers fail to provide specific definitions, then we need to read between the lines. For example, if a study investigates 'listening comprehension', and the dependent variable is a written cloze test, then the default definition of 'listening comprehension' is 'the ability to complete a written cloze passage'. If we were to find such a definition unacceptable, we would be questioning the *construct validity* of the study. Construct validity

Research methods in language learning

TABLE 1.2 THREATS TO INTERNAL AND EXTERNAL VALIDITY POSED BY POOR RESEARCH DESIGN: SAMPLE STUDIES

<i>Example</i>	<i>Critique</i>
<p><i>Internal validity under threat</i></p> <p>In an investigation of three different methods of teaching grammatical structure, three teachers in three different schools are each trained in one of the methods and apply it to their classes. One teacher has three mixed ability classes, another has four mixed ability classes, and the third has two homogeneous groups of fast track learners. At the end of the term, each group is administered a test devised by their teacher. Group means for each group are computed and compared.</p>	<p>In this investigation, the results are uninterpretable. It is impossible to say whether the results are due to the method, the proficiency of the students, the skill of the teacher, or the ease of the test.</p>
<p><i>External validity under threat</i></p> <p>(Adapted from Wiersma 1986) A study investigates the effect of length of visual exposure on the ability to memorise and recall nonsense words. Subjects are ten postgraduate students who are undertaking a master of arts program in psychology. There are five different lengths of exposure, so five groups of two volunteers each receive different lengths of exposure. A volunteer participates in the study by being exposed to 20 nonsense words individually. After each exposure, the volunteer is to reproduce the nonsense word.</p>	<p>Assuming that the performance scores generally increase with increased length of exposure, the question remains: To which populations and conditions can the results be generalised? Can they be generalised to primary and secondary students learning meaningful material? Can they be generalised to young adults working on meaningful tasks in a highly structured situation? The answer to both questions is no. The results may not even be generalisable to the graduate student population, since the participants were volunteers.</p>

has to do with the question: Is the study actually investigating what it is supposed to be investigating? Brown characterises the notion of a psychological construct in the following way:

A psychological construct is a theoretical label that is given to some human attribute or ability that cannot be seen or touched because it goes on in the brain. . . .

It is through tests that . . . constructs are measured indirectly. But researchers cannot take the constructs out and show that the tests are measuring them.

An introduction to research methods and traditions

TABLE 1.3 QUESTIONS FOR ESTABLISHING THE RELIABILITY AND VALIDITY OF A STUDY

<i>Type</i>	<i>Key question</i>
Internal reliability	Would an independent researcher, on reanalysing the data, come to the same conclusion?
External reliability	Would an independent researcher, on replicating the study, come to the same conclusion?
Internal validity	Is the research design such that we can confidently claim that the outcomes are a result of the experimental treatment?
External validity	Is the research design such that we can generalise beyond the subjects under investigation to a wider population?

Therefore, they do the next best thing: They try to demonstrate experimentally that a given test is measuring a certain construct. . . . The experiment may take numerous forms but, most commonly, it is in the form of a differential-group or intervention experiment. A *differential-group experiment* might compare the performance of two groups on a test: one group that obviously has the particular construct and another group that clearly does not. . . .

There are numerous ways to go about establishing the construct validity of a test, but the basic strategy is always the same. The test developer sets up an experiment to demonstrate that a given test is indeed testing the construct that it claims to be testing. (Brown 1988: 103–104)

The central concepts of validity and reliability are extremely important in language research (as indeed they are in all other types of research), as we shall see in the succeeding chapters of this book. I have summarised the discussion in this section by setting out, in Table 1.3, the key questions one needs to ask in relation to reliability and validity.

Action research

A form of research which is becoming increasingly significant in language education is *action research*. This research has been defined in a number of different ways. Kemmis and McTaggart (1988), for example, argue that the three defining characteristics of action research are that it is carried out by practitioners (for our purposes, classroom teachers) rather than outside researchers; secondly, that it is collaborative; and thirdly, that it is aimed at changing things. 'A distinctive feature of action research is that those affected by planned changes have the primary responsibility for deciding on courses

Research methods in language learning

of critically informed action which seem likely to lead to improvement, and for evaluating the results of strategies tried out in practice. *Action research is a group activity*' (Kemmis and McTaggart 1988: 6). A piece of descriptive research carried out by a teacher in his or her own classroom, without the involvement of others, which is aimed at increasing our understanding rather than changing the phenomenon under investigation, would not be considered by these commentators to be 'action research'. For Kemmis and McTaggart, the essential impetus for carrying out action research is to change the system.

Cohen and Manion (1985) offer a similar set of characteristics. They argue that action research is first and foremost situational, being concerned with the identification and solution of problems in a specific context. They also identify collaboration as an important feature of this type of research, and state that the aim of action research is to improve the current state of affairs within the educational context in which the research is being carried out.

While collaboration is highly desirable, I do not believe that it should be seen as a defining characteristic of action research. Many teachers who are interested in exploring processes of teaching and learning in their own context are either unable, for practical reasons, or unwilling, for personal reasons, to do collaborative research. The work that such people carry out should not necessarily be excluded as action research. I would also dispute the claim that action research must necessarily be concerned with change. A descriptive case study of a particular classroom, group of learners, or even a single learner counts as action research if it is initiated by a question, is supported by data and interpretation, and is carried out by a practitioner investigating aspects of his or her own context and situation. That said, I know of few such studies which have not resulted in change of some sort.

Figure 1.5 illustrates the scope of action research and the various stages involved. Several points are worth noting from this example. In the first place, the research is initiated by the practitioner and is derived from a real problem in the classroom which needs to be confronted. Secondly, the research is collaborative – not, in this instance, between colleagues, but between a teacher and a university-based researcher. Thirdly, the teacher collects objective data in the form of classroom interactions and learner language. Fourthly, the results are disseminated. Finally, the project takes the form of an ongoing cycle (Kemmis and McTaggart speak of the 'action research spiral') in which the teacher reflects on, returns to, and extends the initial inquiry.

Is this activity research? I would argue that it is, in that it fits my minimalist definition, containing a question/issue, data, and interpretative analysis. Others may argue that such activity can only lay claims to being research if the teacher has taken steps to guard against threats to the reliability and validity of the research. I believe that care needs to be taken over the reliability of all forms of inquiry, but that for action research there is not the same imperative to deal with external validity. In many cases practitioners are less con-

An introduction to research methods and traditions

- | | |
|--|---|
| Step 1: Initiation | → A teacher comes to me with a problem: His current group of students do not seem interested or motivated. What should be done? |
| Step 2: Preliminary investigation | → We spend some time collecting baseline data through observation and recording classroom interaction. |
| Step 3: Hypothesis | → After reviewing the initial data, we form the hypothesis that the students are unmotivated because the content of the classroom is not addressing the needs and interests of the students. |
| Step 4: Intervention | → The teacher devises a number of strategies for encouraging the students to relate the content of the lessons to their own backgrounds and interests. These include increasing the number of referential over display questions. |
| Step 5: Evaluation | → After several weeks, the class is recorded again. There is much greater involvement of the students, and the complexity of their language and student-led interactions is enhanced. |
| Step 6: Dissemination | → The teacher runs a workshop for colleagues and presents a paper at a language conference. |
| Step 7: Follow-up | → The teacher investigates alternative methods of motivating students. |

Figure 1.5 Steps in the action research cycle

cerned with generating generalisable knowledge than with solving pressing problems associated with their own particular workplace. (Allwright 1991 prefers the term ‘puzzle’ to ‘problem’, in that it avoids the possible negative connotations of ‘problem’.) While such activities therefore fulfil a professional development function, I still believe that if they address questions of interest to other practitioners, if they generate data, and if they contain analysis and interpretation, then they qualify as research. In the sample study summarised in Figure 1.5, extreme caution needs to be exercised in making strong claims about the research outcomes. While the reliability of the research was strengthened by the involvement of an outside researcher, the internal validity of the research is particularly problematic, and it would be extremely unwise for the teacher (or anyone else) to claim that improvements in the students’ language were a result of interventions such as the increased use of referential questions. Numerous competing explanations suggest themselves.

Research methods in language learning

For example, because the class was in progress for several weeks between steps 4 and 5, it could be argued that improvements were simply due to maturation, and that progress would have been recorded regardless of the types of questions asked by the teacher. Despite these problems, I still believe that the investigation was worth carrying out and reporting, particularly as it is the sort of investigation which can be replicated rather easily by other teachers.

Conclusion

In this chapter, I have dealt with some of the central themes and issues associated with research into language learning and use. I have argued that, while the distinction between qualitative and quantitative research is simplistic in many ways, it does represent a real, not an ostensible, distinction. However, the distinction is a philosophical one which is not always reflected in the actual conduct of empirical investigation. Underpinning quantitative research is the positivistic notion that the basic function of research is to uncover facts and truths which are independent of the researcher. Qualitative researchers question the notion of an objective reality. As Rist asserts:

Ultimately, the issue is not research strategies, per se. Rather, the adherence to one paradigm as opposed to another predisposes one to view the world and the events within it in profoundly different ways. (1977: 43)

In the chapters which follow, we shall take up and explore these issues in greater detail. Chapter 2 provides an introduction to the use of the experimental method. We shall also look at the use of statistics and the logic of inferential statistics, which enables us to make generalizations beyond the subjects we have studied to a wider population. Issues associated with descriptive and interpretive research are taken up in Chapter 3, which looks at ethnography, and Chapter 4, which deals with case study methodology, including single case research. Chapter 5 looks at aims, issues, and methods in classroom observation. In Chapter 6, the focus of concern is introspection and the use of introspective methods in research, including think-aloud techniques, diaries, and retrospection. The focus of Chapter 7 is the collection and analysis of speech data collected in naturally occurring interactions. The theme of Chapter 8 is elicitation, and the chapter deals with a number of different methods, such as the interview and questionnaire, which are designed to elicit data from language learners and users. Chapter 9 looks at some of the theoretical and practical issues involved in program evaluation, and raises the question of whether or not program evaluation is a form of research. In the final chapter, practical questions associated with the formulation of a research question or hypothesis, the selection of an appropriate research design, and the analysis and presentation of data are taken up.

*An introduction to research methods and traditions***Questions and tasks****1. Complete the following statements**

Research is . . .

Research is carried out in order to . . .

2. Here is a list of questions which have been addressed in the research literature. (a) What are the key constructs associated with each question? (b) Which do you think might best be investigated through some form of experiment, and which might best be investigated through naturalistic investigation? (c) Can you find any studies which might be investigated *either* through an experiment or a naturalistic study?

- Are authentic materials more effective in bringing about learning than materials written specifically for the classroom?
- Does learning a second language involve the same psycholinguistic processes as learning a first language?
- Are there significant differences in the ways in which people interact with members of the same/opposite sex?
- Do learners from the same ethnic background share learning strategy preferences?
- In classrooms containing both first (L1) and second (L2) language learners, should teachers use different language and interactional patterns with L1 and L2 speakers?
- Do learners who have grammatical rules explained to them learn more effectively than those who learn inductively?
- What happens when teachers share decision-making with their learners?
- Is there a positive correlation between the language addressed to a child in its preschool years by the primary caregiver and ultimate academic attainment?
- Is the difficulty of a listening text influenced by the listener's background knowledge of the subject in question?
- How do people keep casual conversations going?
- Is there a 'critical period' for language acquisition, after which it is much more difficult to acquire a second language?
- Do children consciously try and work out rules as they acquire their first language, or is it a subconscious process?
- How are power relationships in the multilingual workplace linguistically marked?
- Are doctors who are trained in the language and culture of patients from different ethnic backgrounds able to diagnose more effectively?
- Do first language learners learn to do discourse before they learn grammar?
- How do parents help their children acquire language?

Research methods in language learning

3. Which of the following statements from Chalmers (1982) are falsifiable and which are not?

- It never rains on Wednesdays.
- All substances expand when heated.
- All points on a Euclidean circle are equidistant from the centre.
- Heavy objects, such as a brick, when released near the surface of the Earth fall straight downwards if not impeded.
- When a ray of light is reflected from a plane mirror, the angle of incidence is equal to the angle of reflection.
- Luck is possible in sporting speculation.

There are some types of applied linguistic research in which a single disconfirming instance is sufficient to invalidate the claim, hypothesis, or theory under investigation. Which of the following statements would you accept as being invalidated by the existence of a single disconfirming instance?

- Learners will acquire the ability to form questions through inversion before they acquire Wh- questions formed through 'do' insertion.
- Authentic listening materials are more effective than materials specially written for the classroom.
- Parents of hearing and hearing-impaired children will code-switch to accommodate the hearing status of the child.
- The degree to which a learner acculturates to the target language group will control the degree to which he or she acquires a second language.

(In the first part of the preceding task, assertions c and f are not falsifiable. Assertion c is a definition and therefore a necessary truth. Assertion f is quoted from a newspaper horoscope, and, as Chalmers (1982: 40) says: 'It typifies the fortune-teller's devious strategy. The assertion is unfalsifiable. It amounts to telling the reader that if he has a bet today he might win, which remains true whether he bets or not, and if he does, whether he wins or not.')

4. Review one or more studies concerned with some aspects of language learning and use which has been published in a language journal such as *Language Learning*, *Modern Language Journal*, *Applied Linguistics*, *TESOL Quarterly*, *Canadian Modern Language Review*, *JALT Journal*, or *Studies in Second Language Acquisition*. Make a note of the functions of the following components of the report – in other words, what is the author trying to do in each of these sections? (Not all reports will necessarily contain all of these elements, which is why you may need to look at several.)

Abstract

Introduction

Rationale

Literature review

Hypothesis or research questions

An introduction to research methods and traditions

Data collection instruments or methods
Research procedure
Subjects
Data analysis
Results
Discussion
Conclusions

5. What do you see as the potential threats to the validity and reliability of the action research project described in Figure 1.5? What steps might be taken to guard against these threats?
6. What are some of the questions, issues, or problems from your own professional context which might be investigated through action research?

Further reading

Chalmers (1982) provides a detailed introduction to the nature and philosophy of scientific research. He deals at some length with the problems of deduction, induction, and falsifiability.

Nunan (1989) is intended as a practical introduction to action research for those classroom practitioners interested in carrying out such research in their own classrooms. A useful collection of papers on action research can be found in Lomax (1989).

Chaudron (1988) provides an extremely detailed and comprehensive introduction to issues in second language classroom research. Although the emphasis in the book tends to be towards quantitative rather than qualitative research, both are dealt with.

For an introduction to the ethnography of classroom research, see van Lier (1988). Key studies in classroom research are published with a critical commentary in Allwright (1988).