PART I THE BASICS

1. Types of Research
   1.1 Introduction
   1.2 Hypothesis generating vs hypothesis testing
   1.3 Description vs explanation
   1.4 Non-experimental vs experimental
   1.5 Process research vs product research
   1.6 Longitudinal vs cross-sectional
   1.7 Case studies vs group studies
   1.8 Qualitative vs quantitative
   1.9 In situ/naturalistic research vs laboratory research
   1.10 The approaches taken in this book

2. Systematicity in Statistics: Variables
   2.1 Introduction
   2.2 Research design
   2.3 Why do we need statistics?
   2.4 Variables and operationalization

3. Descriptive Statistics
   3.1 Introduction
   3.2 Describing datasets: means and dispersion
   3.3 A different view on variability
   3.4 Frequency distributions

4. Statistical Logic
   4.1 Introduction
   4.2 The chance of making the wrong decision
vi Contents

4.3 Statistical decisions 43
4.4 Degrees of freedom 46
4.5 Parametric and non-parametric statistics 46
4.6 Checking assumptions 47
4.7 Alpha and beta 48

5. Doing Statistics: From Theory to Practice 51
5.1 Operationalization 51
5.2 Forming hypotheses 52
5.3 Selecting a sample 52
5.4 Data collection 54
5.5 Setting the level of significance 55
5.6 Statistics 55
5.7 Interpretation of the data 56
5.8 Reliability of outcome 57

6.1 Introduction 59
6.2 Comparing groups 59
6.3 Assessing relations 65
6.4 Analysing frequencies 72
6.5 Doing statistics 74

7. Conclusion 77
7.1 Statistical dangers and limitations 77
7.2 Validity and reliability 77
7.3 Meaningful outcomes 79
7.4 Statistics and the real world 80

Notes 81
References 83

PART II HOW TO SPSS

1. How to Do Descriptive Statistics 87
2. How to Check Assumptions 91
3. How to Do a $t$-test 95
4. How to Do a One-way ANOVA 101
5. How to Do a Two-way ANOVA 107
6. How to Do a Correlation Analysis 113
7. How to Do a Simple Regression Analysis 117
8. How to Do a Chi-square Analysis 121

PART III  SPSS PRACTICALS

1. Exploring SPSS and Entering Variables 131
2. Descriptive Statistics 133
3. Calculations Using SPSS 137
4. Inductive Statistics 143
5. Miscellaneous Assignments 147
6. Miscellaneous Assignments (Revision) 151

Index 155
Chapter 1

Types of Research

1.1 Introduction

The field of applied linguistics is a large one and this means that applied linguists are interested in many issues. Here is a list of a variety of topics:

- Measuring the effectiveness of early bilingual education: how effective is an early start?
- The relation between characteristics of linguistic input and language development in early bilingual children.
- Assessment of problems of elderly migrants learning Swedish.
- The lag in development in language proficiency of migrant children or children in deprived areas where a local dialect is spoken.
- The storage of multiple languages in our mind: language selection and language separation: how do we keep our languages apart?
- The possibility of ‘blocking’ languages we know while we are listening or reading.
- The impact of learning a third language on skills in the first language (L1) and the second language (L2).
- The role of interaction in the language classroom: who’s talking, what is the input?
- The nature of the impact of information and computing technology on language learning.
- The question of how a threatened language can best be protected.
- The question of why prepositions in an L2 are so difficult to learn.
- The question of if and how a forgotten language can be reactivated.

This list could be extended for pages. A quick look at the Current Contents list of journals in the arts and humanities, which shows the tables of contents of over 1,000 journals, will make clear that the scope is huge and that, even for journals focusing on second language development, the range of issues is breath-taking.
There are many topics of research, but the range of types of research in applied linguistics is much more limited. In this chapter we will provide a systematic overview of different types of research: what are their relevant distinctions and how are they related? There will be no single optimal method for all research topics, since each topic can be worked out in different ways. For your understanding of the research literature, it may be useful to become acquainted with the major categories, so you know what to expect and are able to evaluate the use of a particular design. Also, some familiarity with research terminology and categorization will be helpful in finding relevant research on your own topic.

For clarity’s sake, we will make use of contrasting pairs of types of research, but it should be stressed from the outset that these contrasts are actually the extreme ends on a continuum rather than distinct categories and that the categories are partly overlapping.

1.2 Hypothesis generating vs hypothesis testing

Testing and theory formation is a circular process. Theory must be based on empirical findings, and empirical studies can be used to test the theories formulated. One of the issues related to theory formation is to what extent a theory can be tested, that is to what extent the theory can be used to set up hypotheses that can be empirically tested. Quite often, though, this is not a matter of either–or, but a sequential process. Often theories have not been developed yet to the point that real testable hypotheses have been generated. A current example could be the relation between the use of hand gestures and L2 development. We are now only beginning to see the importance of this type of research, and we are still looking for theories that may help us explain why people gesture the way they do (see for instance de Bot and Gullberg, 2010). However, before we can test what explains cross-linguistic influence from L1 to L2 in gesturing, we first need to find out whether there are actually differences in gesturing between the two languages. Once we have established that, we can proceed to think about specific aspects of gesturing, such as the use of gestures with motion verbs.

Activity 1.1

- If you were to categorize the list of research topics in three or four categories according to the type of research required, how would you do it? On what grounds would you put a specific topic in a category?
- How would you label your categories?
- For four of these topics, briefly work out how you would go about investigating it.
In research reports, we often see phrases like ‘In this study, we test the hypothesis that ...’. However, the formulation of appropriate hypotheses is not always obvious. For instance, if someone claims to ‘test the hypothesis that when second language learning starts after puberty, a native level of proficiency cannot be attained’, then we may wonder what that actually means: is that true for every learner, no matter what? If only one single individual can be found who can achieve this, is the hypothesis then falsified? A hypothesis needs to be narrowed down as far as possible to show what will be tested and what outcomes count as support or counter-evidence.

Activity 1.2: Formulating a research hypothesis

It is not easy to formulate a research hypothesis that is not too broad and not too narrow. The more specific the hypothesis is, the better the chance to test it successfully. The development of a research hypothesis typically goes in stages. A hypothesis like

Elderly people forget their second languages easily.

is not really a hypothesis, but rather a statement. Which elderly people? Compared to what other groups? Do younger people not forget their second languages? What does ‘easily’ mean here? So we need to narrow it down:

Elderly people forget their second languages more quickly than middle-aged people.

Still, this is rather broad and some concepts are not clear and we may have to specify the population. The definition of ‘elderly’ and ‘middle-aged’ may be somewhat vague, and in the description of the population it will have to be made clear what the age range is, though for the hypothesis this will do. But do we also want to include elderly people suffering from dementia or other diseases? And do we want to test every part of the language system? Maybe it is better to limit the study to syntax, morphology, lexicon or fluency. And do we want to look at all second languages? How about the level of education, which is likely to play a role? Narrowing the hypothesis down further could result in something like:

Healthy elderly people forget words in their first second language more quickly than education matched middle-aged people.

‘More quickly’ is still a bit underdefined, but probably clear enough. Sometimes it may help to break the larger hypothesis down into a number of smaller ones in which more details can be provided.

➢ For three of the topics in Activity 1.1, formulate a clear and specific research hypothesis.
The discussion on hypothesis testing inevitably leads to the problem of generalizability: When and how can we generalize our findings to more than the people or phenomena we have looked at? In the chapters on statistics, this will be one of the focal issues, because that is precisely what we need statistics for. Here we want to touch on some of the more general issues.

**Generalizability** refers to the extent to which findings in a study can be generalized to a population that is larger than the samples tested. In most cases it is unimaginable that all individuals of a given group can be included in a study. No study of Chinese learners of English will include all those millions of people. What is typically done is that we draw a sample from that larger population. There are different methods for doing this. The most desirable approach is to have a so-called ‘representative’ sample, which means that all the variation in the larger population is represented in the sample tested. The ideal is hardly ever achieved, because it is very difficult to assess what makes a sample representative; we need to know all the traits that may be relevant and should be included in the sampling. No data will tell us exactly what the relevant traits are for drawing a sample from the large population of Chinese learners of English. The best we can do is to guess and use common sense (and all the relevant research there is, of course) to define the sample. This procedure represents the traditional research approach that we will discuss throughout the remainder of the book.

However, although the majority of studies in applied linguistics have taken a traditional research approach in which the findings of representative samples are generalized to populations, this is not the only way of investigating second language development. A research perspective that is rapidly gaining ground in developmental psychology and applied linguistics is *dynamic systems theory* (DST) or *complexity theory* (CT). This perspective emphasizes the change of development over time and emphasizes the dynamic interaction of factors affecting the (language-) system over time. DST/CT is thus interested in the process of development rather than in the eventual learning outcomes. The focus on the process has important consequences for the research choices that are made. For instance, a DST/CT approach takes into account that developmental processes are complex processes in which characteristics of the individual learner interact with the environment. Every learner is unique, and her developmental path will be affected by the internal structure of her system and her interaction with the environment. Inevitably, there will be variation between individual learners.

Individual patterns can tell us in detail how an individual developed and what factors may have played a role, but at the same time we need general tendencies, not only because of educational policy reasons, but also because we need information on the likelihood that a given factor has an impact on development, so that we can include it in the study of individual developmental patterns. This is basically a cycle in which we move from factors that seem to have an effect at the individual level to testing that
effect on a larger sample to obtain an estimate of its strength and then back to the individual level again to study the impact in more detail. An example could be the role of motivation: its effect may be suggested when we look at the learning process of an individual learner who indicates why he or she was motivated to invest time and energy in learning a language at some moments in time and not at all at other moments. To know more about the potential strength of such a factor, we may then do a study on a larger sample of similar learners. With that information, we can go back to individuals and see how the factors that appeared to be important in the larger sample affect the individual learner. The general pattern will be less detailed and typically will not give us information about change of motivation over time in the same way that an individual case study can.

A comprehensive perspective on language development means balancing two ways of looking at research: on the one hand one may give more attention to individual developmental patterns, but on the other hand one may accept the relevance of general tendencies for various purposes. It may be true that for some learners there is no impact from the first language when learning the second, and the variation in cross-linguistic influences between individuals may be considerable; but what remains is that in general the first language does play a role in learning a second language. It is important to realize that these are complementary perspectives, each of which is equally relevant. The main focus in this book is on hypothesis testing and generalization, but we should not forget that this is only one side of the coin.

1.3 Description vs explanation

The discussion in the previous section is closely related to the distinction between description and explanation. Before we can explain anything, we first need a good description. For instance, before we can explain why interaction in the classroom is beneficial for second language development, we need to describe what goes on in classrooms: who is actually saying what to whom, how complex the language used is, whether what is said is also understood, and whether the language used is correct or full of errors. Most of the research we do in applied linguistics is descriptive. We describe processes of learning and teaching, naming different factors that play a role in these processes; we describe language policy programmes and their effectiveness; we describe the impact of learning environments on learning; and
Many studies investigate the effect of X on Y. A randomly chosen issue of one of the leading journals in our field, *Studies in Second Language Acquisition*, reveals the following contents. There is a study on the effect of ordering input and interaction on language proficiency (Gass and Torres, 2005), one on the effect of reading and writing on word knowledge (Webb, 2005), and one on the impact of negative feedback and learners’ responses to ESL question development (McDonough, 2005). All of these studies describe the effect of the manipulation of one variable on another one. But even if the relation between two variables is established (and it normally is, because journals do not usually publish null-results), then this is not the same as an explanation of that relation. The only conclusion that can be drawn is that the development of Y is influenced by X. To really account for the influence observed, there must always be a rigorous and detailed theory behind the research, which provides explanations for the phenomena.

There is another sense of the concept of explanation, which is more statistical in nature. As we will see in Chapter 2, variation within individuals and groups vs variation between individuals and groups is the essence of statistical procedures. We try to explain variation in variable A by looking at the impact of the systematic and controlled variation of variable B. For instance, we look at the variation in the acquisition of new L2 words (variable A) by manipulating the methods of teaching (variable B). If the experiment works, the variation in variable A is reduced, because we have taken out the variation that is caused by variable B. Suppose we give two different methods to two different groups, a ‘strong’ group and a ‘weak’ group. If we look at the learning results of the two groups together, we will find some learners who improved a lot, some less, some not at all. If we look at the groups for methods 1 and 2 separately, we may find that for one method most learners will improve a lot, while for the other group learners improve only slightly or not at all. There is a great deal of variation in the two groups taken together because there will be good and bad learners within each group, while there is less variation in the groups separately because the good learners are in one group and the weaker learners are in the other group. Within the groups the learners are more similar and therefore there is less variation among the individuals in the group. In statistical terms this is referred to as ‘variance explained’, variance being a specific type of variation. The goal in experiments is to explain as much variation as possible, because that will tell us to what extent we can explain a given effect. However, this is not an explanation in the theoretical sense, but a description of the relation between variables.

### 1.4 Non-experimental vs experimental

Applied linguistics has positioned itself as part of the social sciences and distanced itself from the humanities by adopting research techniques and
Types of Research

Activity 1.4

One of the subdivisions we can make with regard to types of research is the one between experimental and non-experimental.

➢ Which of the topics in Activity 1.1 can best be investigated using experimental research and which with non-experimental research?
➢ Mention one disadvantage and one advantage of experimental research.

paradigms based on the science model. In this model, quantitative empirical research and controlled experiments are often considered the only way to make progress. The aim is to decompose complex processes in parts that can be studied and manipulated experimentally. Experiments and statistical manipulations provide ‘hard’ evidence as compared to the ‘soft’ evidence evolving from the more interpretative research that dominates in the humanities. If one looks at the bulk of research as reported in the most prestigious journals and books in applied linguistics, the experimental approach is dominant, though at the same time it is obvious that it is no longer seen as the only way to practise research. The choice of an experimental or a non-experimental approach largely depends on what a researcher wants to know. For a study of the organization of the bilingual lexicon or the perception of foreign accents by native speakers, experimental approaches may be the logical choice; the way human language processing takes place is not open to introspection and we can study this effectively through controlled experiments. Other aspects, such as non-instructed L2 development, we can study better through non-experimental techniques, such as observations and analyses of spontaneous speech.

In studying Second Language Development (SLD), a wide range of experimental techniques have been used, ranging from grammaticality judgements to lexical decision tasks, and more recently different neuro-imaging techniques have been used that provide insight into brain activity while processing language. A detailed discussion of various techniques is beyond the scope of this book, but a good overview of different experimental techniques that have been used for the study of SLD can be found in Mackey and Gass (2005) and thorough discussions of different brain-imaging techniques have been provided by Brown and Hagoort (1999).

Sometimes there is a choice between using an experimental technique and a non-experimental technique. The choice for one or the other is determined by the aim of the study. An interesting example is the study of L2 pragmatic competence. Hendriks (2002) studied the acquisition of requests by Dutch learners of English. Her aim was to study the impact of power relations, social distance and conversational setting on the use of politeness strategies in requests. She could have made recordings of requests in
spontaneous conversations, but she wanted to study the systematic effects of each of these variables and their interactions. She would have needed a very large corpus of utterances to find sufficient examples of requests that differed in terms of power, social distance and setting. Therefore, she decided to use an experimental technique that tries to mimic real life interaction while allowing for systematic variation of variables. The technique used is the Discourse Completion Task. In this task a short description of a conversational setting is presented and the participant has to construct a sentence as a reaction. Here are two examples:

*The living room*

You were in your room upstairs doing your maths homework, but you were not able to do the sums. You need some help. You go down to the living room where your dad is watching a documentary on television. What do you say to your dad?

*The supermarket*

You are standing in line at the checkout with a shopping trolley full of groceries. You are late for an important meeting. There is one man in front of you. What do you say to the man in front of you?

The use of such tasks allows for systematic variation of the variables, but it is of course not a natural setting or real conversation. This shows that the choice for an experimental or non-experimental approach is always a trade-off between controllability and ecological validity. Ideally, data from controlled experiments should be validated through a comparison with ‘real’ conversational data.

### 1.5 Process research vs product research

The distinction between process research and product research has to do with how change takes place and what the result of the change is. An example may help to explain what we mean by this. There is an abundance of research on different types of immersion education. Not only in a bilingual country like Canada but also in Europe, so-called CLIL programmes have shown that these approaches are very effective. CLIL means Content Based Instruction in North American terms or Content and Language Integrated Learning in European terms. A study by Huibregtse (2001; see also Admiraal et al., 2006) showed that Dutch secondary education students in a CLIL programme outperformed a matched control group with respect to the development of different aspects of English. This is typically a product study: there is an effect, but no indication of what might have caused the effect. To evaluate the causes of the effect, it would be necessary to have a detailed look at what goes on in CLIL classes, what the impact of such an approach is on
students’ attitudes to English and how the proficiency changes over time as a function of contact. That would then be the process part of that study. Of course, there are layers in both ‘products’ and ‘processes’. Going back to the CLIL study: there are products in terms of gains in proficiency on different levels (syntax, lexicon, reading, writing), but there can also be more long-term products, such as better jobs or more study abroad with the CLIL students as compared to the control students. There are layers in processes too: we referred to classroom processes, but ultimately we are interested in the impact of external resources on the individual student and how that leads to learning: what are the characteristics of the setting and how does the individual system react to that. Most studies in applied linguistics have been effect studies that focus on the products of language development, but recently the importance of process-oriented research has been emphasized. Investigating processes can be done with test-retest designs (‘repeated measurements’), to see how a certain intervention (like a teaching method) may have affected learning outcomes over time. In such designs there are usually two groups, a control group (that does not get the ‘treatment’) and an experimental group, and there are usually two or three test moments. However, to really make sense of the process of development, more measurements will have to be taken over a relatively long period of time, in what is referred to as ‘longitudinal’ research.

1.6 Longitudinal vs cross-sectional

Longitudinal research is research in which individuals’ development over time is studied. Most studies of children growing up bilingually are examples of longitudinal research: the child is typically video/audio-taped at regular intervals over a longer period of time, sometimes for more than three years, and transcripts of the recordings are analysed with respect to relevant aspects, such as mean length of utterance and lexical richness. But other types of development can also be studied longitudinally: Hansen et al. (2010) looked at the attrition of Korean and Japanese in returned missionaries, who typically acquired the foreign language up to a very high level, used it in their work as missionaries and after they returned hardly ever used that language. This study is unusual, because it is longitudinal with only two moments of measurement in ten years. In most longitudinal studies there are more moments of measurement with smaller time intervals. Such studies necessarily take a long time; even three-year data collection periods may be too short to cover a significant part of the developmental process. And funding agencies are not very keen on financing projects that take more than four or five years. Therefore, the number of longitudinal studies is small, but those projects (like the European Science Foundation study on untutored L2 development in different countries (see Klein and Perdue, 1992; Becker and Carroll, 1997)) have had a major impact on the field.
Because of the time/money problem of longitudinal studies, many researchers use cross-sectional designs. In cross-sectional research, individuals in different phases of development are compared at one moment in time. For the study of the development of morphology in French as an L2, a researcher may compare first, third and fifth graders in secondary schools in Denmark. Rather than follow the same group of learners for four years as they progress from first to fifth grade, different groups in the three grades are compared at one moment in time.

Both longitudinal designs and cross-sectional designs have their problems. In longitudinal studies the number of participants is generally very small because a large corpus of data on that one individual is gathered. Large numbers of participants would make both the data collection procedure and the processing and analysis of the data extremely demanding and time consuming. The small numbers mean that the findings may be highly idiosyncratic and difficult to generalize. As we have seen, this may not be a problem in studies that use the uniqueness of the individual’s development as the central issue, as is normally the case in DST/CT approaches to language development. Another problem of longitudinal studies is subject mortality, the dropping out of subjects in a study. With each new measurement, there is a risk of this happening, and the longer and more demanding the study, the higher the risk of drop out. An additional problem is that in such studies drop out is typically not random, but selective or biased: in a study on acquisition or attrition, subjects that do not perform well will be more likely to lose their motivation and drop out than more successful ones, leaving a biased sample that is even less generalizable.

Cross-sectional designs can be problematic, because the assumption that the groups that are compared behave like one group tested longitudinally may not be true. Referring to the example above, there may be specific characteristics of different age groups, such as changes in the curriculum, natural disasters, changes in demographic trends and changes in school population that can make the three groups very different. One solution for this so-called cohort effect is to take more than one cohort; so rather than only testing groups 1, 3 and 5 in year x, one also tests the groups 1, 3 and 5 of the next year or cohort. If the findings for the two cohorts are similar, it is assumed that the groups do not behave atypically. Some studies try to get the best of two worlds by combining longitudinal and cross-sectional designs: in research on ageing such cross-sequential designs have been used frequently, and it was also used by Weltens (1989) and Grendel (1993) in their studies of the attrition of French as a foreign language in the Netherlands.

1.7 Case studies vs group studies

Most longitudinal studies will be case studies, while cross-sectional studies tend to be based on group data. Some studies use multiple cases, and
although in terms of numbers they may seem to be the same as a group, the approach is fundamentally different. In case studies, we typically find a holistic approach, which aims at trying to integrate as many aspects that are relevant for the individual case as possible. In group studies, it is similarity rather than difference that counts: a group is selected with specific characteristics – for example, Turkish undergraduate students doing a course on Academic Writing in English – while other differences are either ignored or controlled for by using background questionnaires and specific statistical techniques to cancel out such differences.

1.8 Qualitative vs quantitative

The discussion of qualitative versus quantitative studies would take another book to discuss in sufficient detail. It has been one of the main rifts in the social sciences including applied linguistics over recent decades. It looks as if the fiercest controversy is over now, but the different communities still view each other with distrust. For a long time a researcher had to be in one or the other community, but now it seems acceptable to take an elective stance and use more qualitative or more quantitative methods depending on the type of research question one wants to answer. Following Mackey and Gass (2005, p. 363), the two approaches can be defined as follows:

**Qualitative**: Research in which the focus is on naturally occurring phenomena and data are primarily recorded in non-numerical form.

**Quantitative**: Research in which variables are manipulated to test hypotheses and in which there is quantification of data and numerical analyses.

From these definitions it follows that the two approaches differ fundamentally in epistemological terms and with respect to the research methods used. Qualitative research is holistic, trying to integrate as many aspects that are relevant into a study. It is also by definition interpretative and therefore in the eyes of its opponents ‘soft’. In qualitative research a number of techniques are used, such as diaries of learners, interviews, observations and introspective methods such as think-aloud protocols (see ibid., ch. 6; Brown and Rodgers, 2002, chs 2–4 for discussions of various methods). One of the main problems is the lack of objectivity in those methods: in all methods, the researchers interpret what goes on, and some form of credibility can only be achieved through combinations of data (‘triangulation’) and the use of intersubjective methods, in which the interpretations of several ‘judges’ are tested for consistency. All of this may not satisfy the objections raised by hard-core quantitativists. For them only objective quantitative data are real ‘hard’ data; for example, there is little a researcher can change or interpret in the latencies in reaction-time experiments. The starting point
of quantitative research is that the entire world is one big mechanism and by taking it apart and studying its constituent parts we will in the end understand the whole machine. Exactly this position is criticized by qualitative researchers.

A problem with the experimental and quantitative approach is that it is not always clear what participants in such experiments actually do. There is a substantial set of studies on the recognition of pseudo-homophones (like English ‘coin’ and French ‘coin’, which means ‘corner’ in English). The list of words to be recognized typically consists of many regular words with some of these pseudo-homophones interspersed. The researcher’s hope is that the participants will not notice these words and become aware of the fact that they are special, because that could have an effect on their strategies in processing. To what extent participants actually do notice the trick is often unclear. Participants in such experiments are typically psychology students who have to take part in many different types of experiment and who have accordingly become quite clever in detecting the trick.

1.9 In situ/naturalistic research vs laboratory research

In situ or naturalistic research studies a phenomenon in its normal, natural setting and in normal, everyday tasks, while laboratory research both isolates a phenomenon from its normal setting and uses data that are an artefact of the procedures used. Laboratory research aims at finding ‘pure’ effects that are not tainted by the messiness of everyday life. In such studies either the grammaticality of sentences in isolation is tested through grammaticality judgements or the process of lexical access is studied using reaction-time experiments. Experimental laboratory research has reached extremely high standards, mainly through a very successful experimental psychology tradition in North America and Western Europe. Therefore, that type of research is held in high regard and is also used in applied linguistics. The counter-movement that advocates a more qualitative and naturalistic approach has long been marginalized and has created its own subculture and its own journals and societies. Their main argument is that

Activity 1.5

- Which of the topics in Activity 1.1 can best be investigated using longitudinal research and which can best be investigated using cross-sectional research?
- Mention one advantage and one disadvantage of cross-sectional research.
- How can the cohort effect be avoided?
reductionist research has no ‘ecological validity’ in that it does not actually tell us what reality looks like. Researchers in the laboratory tradition have problems countering this argument, because their research does not always lead to the kind of deeper insight it is supposed to bring. The gap between the methods used and the reality it claims to inform us about has become so wide that even researchers themselves may have problems showing the relevance of what they do. To give an example, there is a large body of research on word recognition, mostly using the lexical decision paradigm in which participants are presented with letter strings on a computer screen and are asked to indicate as quickly as possible whether the letter string is a word in a given language or not. At the beginning of this field of research, it was assumed that word recognition data would inform us about the process of normal reading, but over time researchers working on word recognition have developed their own sets of questions that may be only marginally linked to the process of normal reading.

Researchers using naturalistic data claim that their research is more ecologically valid because it focuses on the tasks in their normal setting. Up to a point this is probably true, but with the use of different introspective methods, they may also have crossed the line and adapted methods that may create their own type of data that are as far removed from reality as the reaction times and error rates of word recognition researchers. The validity of introspection has been questioned and the core of the problem is sublimely expressed in the title of Klein’s (1989) review of Kasper and Faerch on introspective methods in L2 research: ‘Introspection into What?’.

Activity 1.6

Here is the summary of an article by Webb and Kagimoto (2011) from the journal *Applied Linguistics*:

This study investigated the effects of three factors (the number of collocates per node word, the position of the node word, synonymy) on learning collocations. Japanese students studying English as a foreign language learned five sets of 12 target collocations. Each collocation was presented in a single glossed sentence. The number of collocates (6, 3, 1) varied per node word between three of the sets, the position of the node word (+1, −1) varied between two of the sets, and the semantic relationship between collocations (synonyms, non-synonyms) varied between two sets. Productive knowledge of collocation was measured in pre- and post-tests. The results showed that more collocations were learned as the number of collocates per node word increased, the position of the node word did not affect learning, and synonymy had a negative effect on learning. The implications for teaching and learning collocations are discussed in detail.

What do you think are the main characteristics in terms of research types of the findings reported?
Index

A
Alpha error (α) 41–3, 48, 55
Alternative hypothesis 42, 43, 52
ANOVA 63
Assumptions 39, 46–8, 91–3

B
Bar graph 33
Beta error (β) 41–2, 48–9
Bivariate correlation see Correlation
Bonferroni correction 62
Boxplot 27, 134

C
Case studies 12–13
Causal modelling 67
Chi-square (χ²) 72–4, 121–7
Cohort effect 12
Complexity theory 6
Construct 19–21
Construct validity 77–8
Contingency table see Cross-tabulation
Correlation 65–9, 71, 113–116, 145
Cronbach’s alpha 58, 78, 146
Cross-sectional research see Research
Cross-tabulation 72–3, 121

D
Data collection 54
Data view 131
Degrees of freedom (Df) 46, 98
Description 7–8
Descriptive research see Research
Descriptive statistics 25, 55, 87–90, 133–5
Discourse completion task 10
Dispersion 26–8
Distribution see Frequency
DST or DST/CT see Dynamic Systems Theory
Dynamic Systems Theory 6, 12, 30, 47, 80

E
Effect 8
size 48, 53, 64–5
Equality of variance see Variance, homogeneity
Error 70
Expected value 74, 121, 122–4
Experimental research see Research
Explore 89

F
F 60–2, 103
Factor see Variables, dependent
Falsification (principle of) 42–3
Frequency 31–4, 87
distribution 31–2
analysis 72
polygon 35
tally 31

G
Gabriel’s procedure 104
Games-Howel procedure 104
Generalization 6, 12, 39, 45
G-power 47
Grouping variable see Variables, independent

H
Histogram 33, 138
Hochberg’s GT2 104
Homogeneity of variance see Variance
| Hypothesis 5 | formulation 5, 52 |
| In situ research see Research |
| Inductive statistics 25, 39, 45, 143 |
| Interaction 63, 110–11 |
| Intercept 70, 117 |
| Interquartile range 27 |
| Intervention 11 |
| Introspection 15 |
| Kendall’s Tau 71 |
| Kolmogorov-Smirnov test 47, 92, 140 |
| Kolmogorov-Smirnov Z-test 100 |
| Kruskal-Wallis H-test 63, 106 |
| Kurtosis 36–8, 47, 92, 140 |
| Lambda 124, 126 |
| Level of significance see Significance |
| Levels 21, 59 |
| Levene’s test 48, 93, 98–9, 139, 141 |
| Line graph 34 |
| Longitudinal research see Research |
| Lower quartile 27 |
| Main effect 110 |
| Mean 25 |
| Meaningfulness 49, 79 |
| Means analyses 60 |
| Median 26, 27 |
| Min-Max graph 30–1 |
| Mode 26 |
| Mortality (subject mortality) 12 |
| Multiple regression analysis 71 |
| Multivariate ANOVA 63 |
| Non-parametric statistics 39, 46, 59, 63–4, 71, 100, 106 |
| Normal distribution 34–5, 41, 47, 92, 138, 140 |
| Null-hypothesis (H₀) 42–3, 52, 55, 66 |
| Observer’s paradox 54 |
| One-tailed 44, 52, 55, 98 |
| One-way ANOVA 60–3, 101–106 |
| Operationalization 19–24, 51, 78 |
| Output file 133 |
| Parametric statistics 46–8, 59 |
| Partial correlation 68–9 |
| Pearson r 66, 113 |
| Percentile 35 |
| Population 6, 25, 39, 45, 52 |
| Posthoc analysis 62, 103–105 |
| Power 48, 53 |
| Predictor variable see under Variables |
| p-value 45, 103 |
| Qualitative research see Research |
| Quantitative research see Research |
| Range 26–7 |
| Regression 69–71, 117–20 |
| Regression line 70 |
| RegWQ 104 |
| Reliability 57–8, 78 |
| Repeated measurements 11 |
| Repeated measures 63 |
| Representative sample see Sample |
| Research |
| cross-sectional 11–2 |
| descriptive 7 |
| experimental 8–10, 14, 53 |
| in situ 14–5 |
| longitudinal 11–2, 31 |
| process 10–1 |
| product 10–1 |
| qualitative 13–4 |
| quantitative 13–4 |
| quasi-experimental 53 |
| Researcher expectancy 54 |
| Sample 6, 25, 39, 45, 52–3 |
| random 52–3 |
| representative 6, 7, 52–3 |
| stratified 52 |
Sample size (N) 46, 48, 53
Scale 21–2
Scatterplot 66, 114–116
SD see Standard deviation 30
Second-language development 9
Self-selection 54
SEM see Standard error of the mean
Shapiro-Wilk test 47, 92, 140
Significance 43–4, 51, 55, 57, 80
level of 54
Skewness 36–8, 47, 92, 140
negatively skewed 37
positively skewed 37
SLD see Second-language development
Slope 70, 117
Spearman’s Rho 71, 113
Split-half correlation 57–8, 78
Standard deviation (SD) 29, 45
Standard error of the mean (SE or SEM) 45, 98
Statistic 55
Stratum see Sample, stratified
Subject expectancy 54
Subject mortality 12

T
Test variable see Variables, dependent
Test-retest reliability 57
t-test 48, 55–7, 60, 95–100, 139
one-sample 96
independent samples 96–9, 143
paired samples 96
Tukey HSD 104–5
Two-tailed 44, 52, 55
Two-way ANOVA 60, 107–12
Type I error see Alpha error
Type II error see Beta error

U
Upper quartile 27

V
Validity 10, 15, 19–20, 77–8
Variable view 131
Variables 19–24, 51
control 23
dependent 22, 97
independent 22, 97
interval 21
nominal 21
ordinal 21–2
predictor 69, 117–8
ratio 21
Variance (s²) 8, 29, 64
homogeneity of 46–8, 93, 141
Variation 8
Volunteer effect 54

W
Weight cases 121–2
Welch’s F 103

Z
z-score 30