Give a man a fish and he eats for a day. Teach him how to fish and he eats for a lifetime.

# Applied Linguistics:

## Research Methods and Thesis Writing

## 应用语言学研究方法与论文写作

文秋芳 著

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

# Applied Linguistics:

## Research Methods and Thesis Writing

# 应用语言学

## 研究方法与论文写作

文秋芳 著

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

(京)新登字 155 号

## 图书在版编目(CIP)数据

应用语言学研究方法与论文写作/文秋芳著. －北京:外语教学与研究出版社,2001
ISBN 7－5600－2365－7
Ⅰ. 应… Ⅱ. 文… Ⅲ. ①外语教学－研究方法 ②外语教学－论文－写作
Ⅳ. H09

中国版本图书馆 CIP 数据核字(2001)第 048803 号

**应用语言学研究方法与论文写作**
文秋芳 著

* * *

**责任编辑**: 刘博然
**出版发行**: 外语教学与研究出版社
**社　　址**: 北京市西三环北路 19 号 (100089)
**网　　址**: http://www.fltrp.com.cn
**印　　刷**: 北京师范大学印刷厂
**开　　本**: 850×1168 1/32
**印　　张**: 13.75
**字　　数**: 294 千字
**版　　次**: 2001 年 7 月第 1 版 2001 年 7 月第 1 次印刷
**印　　数**: 1—5000 册
**书　　号**: ISBN 7－5600－2365－7/H·1258
**定　　价**: 17.90 元

* * *

# Preface

" Give a man a fish and he eats for a day. Teach him how to fish and he eats for a lifetime."

Doing research is a mystery to many graduate students and second language teachers in China. This book aims to demystify the process of carrying out research in the field of applied linguistics, particularly in the area of second language acquisition and teaching. This book is written for those who have no previous research experience at all or those who have done some research without formal training. It is particularly useful for graduate students of applied linguistics, second language teachers and teacher-trainers.

This book is concerned with both how to conduct research and how to write a thesis or a dissertation. Graduate students, I have found, face difficulties not only in actual research but also in writing up their theses/dissertations. This book addresses questions that have actually been experienced by me and my students. For example. How can I select a suitable topic? Which research design should I employ? What is the best time for me to collect data and in which way? Should I analyze the data quantitatively or qualitatively? What is the difference between reporting a result and discussing a result? What should be coverad by a literature review?

In response to these challenges, this book provides the readers with solutions, and with suggestions if there is no definite answer. Part I serves as a general introduction in which some basic concepts are explained; Part II focuses on the research process and covers different phases of a single

study. Part III describes the overall structure of a thesis/dissertation and what is expected to be included or not included in different chapters.

This book is written in an accessible style and interactive manner so that the readers can follow the text without too much difficulty. The book has cited various examples from theses and papers written by Chinese scholars and graduate students in order to facilitate the understanding of abstract terms and the research process. Moreover, unlike books on research methodology published in the western context, this book has particularly focused on Chinese graduate students' difficulties and problems in their research and thesis writing. Therefore, this book can be used as a textbook for a course on research tuethods as part of postgraduate program as well as a reference hook for completing a master's thesis or a doctoral dissertation.

A book such as this cannot be produced without the assistance of many people. First of all, I would like to extend my thanks to the students in Nanjing University who have survived the course in which early versions of this text were used. Their responses to this text help me identify which part was well-written and which part needed clarification and elaboration.

I would also like to give my special thanks to Joanna Radwanska-Williams, Professor from the Chinese Utiversity of Hong Kong who carefully edited from Chapter One to Chapter 11 and offered many invaluable comments as well as suggestions; Nancy Pine, Professor from Mount St. Mary's College, Los Angeles, who patiently read the first six chapters of the manuscript and corrected a number of non-native expressions; and Kate Parry, Professor from Hunter College, New York, who skillfully copyedited the first five chapters of the manuscript and made constructive suggestions on improvetuent. Their kind and timely help has made the book better than it would have been otherwise.

I am most indebted to R.K. Johnson, my former supervisor as well as my lifelong friend, who took great pain to proof-read the whole manuscript and provide insightful comments as well as suggestions.

I am very grateful to Ling Wang who helped me check the references and produce some figures; to Qi Chen, Zhihong Qin, Qihong Mao, Yeqiu Zhu and Qin Liu who helped proof-read earlier versions of the text.

Thanks should also be extended to my husband who relieved me from daily housework, and endured late nights and lost weekends. Furthermore, he helped produce some figures and the table of contents.

Finally, I would like to acknowledge the support of the United Board and Harvard-Yenching Institute that provided me with a grant in the academic year of 2000 when I was revising the manuscript for the book as a visiting scholar at Harvard University. I would also like to acknowledge Nanjing Unviersity for granting me a sabbatical leave during this period.

Wen Qiufang
Nanjing University

# Contente

# Part I

# Introduction

Part I is an introduction consisting of two chapters. Chapter One addresses the questions "What is research?" and "How are different types of research classified?" Chapter Two introduces fundamental concepts involved in research. By studying this part, you will be able to:

- Understand the basic criteria for good research。
- Have an overview of the research process。
- Differentiate different types of research。
- Grasp the fundamental concepts involved in research。

# 1. What is Research?

If someone asks you to make a cake, you should know what kind of cake he/she wants. To look for a key, you must know what a key looks like. Without a good understanding of the outcome you intend to obtain, you will experience a lot of frustrations and even failures. The extreme case is that you might have spent your lifelong time and efforts doing a piece of work hut the eventual results you have obtained are not what you desired at all. Similarly, once you make up your mind to make a commitment to research, the first legitimate question you should ask is "What is research?" The importance of such a question is well illustrated in the following parable:

A MAN LOOKING FOR FRUITS

There was once a man who lived in a country that had no fruit trees. This man was a scholar and spent a great deal of time reading. In his readings he often came across references to fruit. The descriptions of fruit were so enticing that he decided to undertake a journey to experience fruit for himself.

He went to the market and asked everyone he met if they knew where he could find fruit. After much searching he located a man who knew the directions to the country and place where he could find fruit. The man drew out elaborate directions for the scholar to follow.

With his map in hand, the scholar carefully followed all of the directions. He was very careful to make all of the right turns and to check out all of the landmarks that he was supposed to observe. Finally, he came to the end of the directions and found himself at the entrance to a large apple orchard. It was springtime and the apple trees were in blossom.

> The scholar entered the orchard and proceeded immediately to take one of the blossoms and taste it. He liked neither the texture of the flower nor the taste. He went to another tree and sampled another blossom, and then another blossom, and another. Each blossom, though quite beautiful, was distasteful to him. He left the orchard and returned to his home country, reporting to his fellow villagers that fruit was a much overrated food.
>
> Being unable to recognize the difference between the spring blossom and the summer fruit, the scholar never realized that he had not experienced what he was looking for.
>
> —From Halcom's Evaluation Parables

The scholar mistook a blossom for the fruit simply because he did not know at the beginning what fruit was. We hope you can bear this parable in mind as you learn about the nature of research.

## DEFINITION OF RESEARCH

In *Collins English Language Dictionary*, "research" is defined as "a detailed study of a subject or an aspect of a subject. If you do research, you collect data and analyze facts and information and try to gain new knowledge or new understanding" (p. 1231). Although this is not a technical explanation, it gives readers a general picture about what research is. By this definition, you may know that research activities include data-collection and data-analysis, and their purpose is to obtain a better understanding of something. Now let us look at a technical definition offered by Hatch and Farhady (1982): Research is "a systematic approach to finding answers to questions" (p. 1). This definition is shorter than the one provided by the dictionary, yet it touches the nature of research. It implicitly tells us three essential elements of research: questions, a systematic approach, and answers. You may use "PPP" to stand for **Purpose** (questions), **Process** (a systematic approach) and **Product** (answers) (See Figure 1.1).

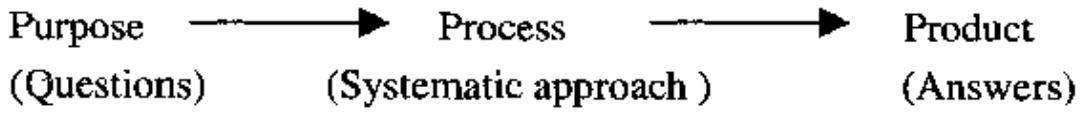



Figure 1.1: A simplified model of research

The above three elements are interrelated to one another. All research starts with questions. Without questions in the first place, there will be no research. Does a good question guarantee a systematic approach used in the research? No way. The selection and construction of an approach needs another set of skills which are different from those required by developing good questions. Without them, the methods chosen are most likely inappropriate, and can never find valid answers. Even if researchers have good research questions and follow the procedures without errors, valid answers may not be natural results since the interpretation of the findings could be illogical or untenable. Therefore, each of the three elements in the definition should be paid equal attention to. If there is any flaw in one element, the whole piece of research will be ruined. The following sections will describe each of the three elements.

## Having Good Questions

Good questions ensure the research goes in the right direction, delimit the research boundary and keep you focus on what you intend to do. To have questions may not be that difficult, but to have good questions is not easy at all. What kind of questions can be qualified as good questions? The qualifying features can be illustrated by three adjectives: **significant**, **original**, and **answerable**.

A **significant** question must be of practical and/or theoretical value. Consider, for example, this question: "Do proficient writers make fewer grammatical mistakes in L2 compositions than less proficient writers?" Surely proficient writers should make fewer grammatical mistakes than less proficient ones. Otherwise they cannot be called proficient writiers.

Therefore, the question is trivial since the answer is self-evident and the findings can neither help improve teaching and learning nor contribute to theory-building. However, if you change the question into "How do proficient writers differ from non-proficient writers in grammatical accuracy?" the answer will be of importance. From the practical point of view, the findings might help teachers understand specific differences between proficient and less proficient writers and thus they can help both less proficient and proficient writers to improve their grammatical accuracy more effectively. For theory-builders, the findings might provide evidence to construct a model for L2 interlanguage development or in support of or against the existing model.

A research question is regarded as **original** when it is different from questions which have been asked by other researchers in one or more aspects, such as differing in learning contexts, or in types of learners, or in the methods used in data collection and in data analysis. In other words, an original question does not need to be totally new. In reality, originality can be a matter of degree.

An **answerable** question is one that can be tackled by the researcher within the time and resources available. This requirement may appear to be unnecessary or the easiest to follow. However, almost all beginning research students fail to meet it because they tend to be over-ambitious and they lack the experience to anticipate difficulties.

To satisfy the above three criteria is invariably the most difficult part of research. Nevertheless, the importance of choosing appropriate questions is often under-estimated and the difficulty in doing so is usually not fully recognized. Research is like taking a long journey. Asking an ill-formulated question is the same as traveling in the wrong direction and can result only in wasted time and effort. In this sense, "it is worth spending as much time as is necessary to get the question right"(Nunan, 1992: 211). How you can develop good research questions will be further discussed in Chapter Three "Developing research questions".

## Employing a systematic approach

By using a **systematic approach**, we mean that research should follow a set of procedures which are clearly described and can be fully justified. The research procedures in some cases are predetermined in the sense that they are decided before the data-collection while in other cases they are developed during the research process. In either case, the procedures used for selecting subjects, data-collection and data-analysis should be recorded and reported to other researchers. Furthermore, the reasons why certain procedures are adopted should be explained in terms of established principles in the discipline. Being transparent and justifiable, the procedures thus can be easily replicated by other researchers.

One thing is worth mentioning here: no approach is perfect, particularly when the research is to study human beings. Thus a systematic approach should not be understood as an impeccable approach. Actually, it is common for researchers to admit there are limitations in their studies.

## Obtaining valid answers

The answer to a question, the last element in the definition but not the least important, must be of high **validity**. Validity is an essential yet difficult concept that cannot be explained in one or two sentences. You will understand it gradually through reeding this book. At this initial stage, I will explain it in a very simple way. When an answer is said to be valid, it means that the claimed answer is the only answer we can obtain. If there is any alternative answer, the validity of the study will be called into question. For example, one study attempted to find out whether there is a relation between L2 learners' vocabulary size and their reading amount under the assumption that the more L2 learners read, the bigger vocabulary size they have. The finding from this study said that the amount of reading did affect the size of vocabulary as expected, which

appears to be reasonable and logical. However, an experienced researcher read the report very carefully and found that the vocabulary test was not scientifically designed. In this case, it is not sure the found relation was caused by the reading amount or due to the poorly-designed test. In this case, there are two competing explanations for the said relation. Therefore, people have the reason to say that the answer is not valid. Another study aimed to find out whether there was any gender difference in L2 learning. The study revealed that female English majors outperformed male English majors in an Englisb proficiency test and thus it was concluded that females are more talented than males in L2 learning. Obviously we can find a dozen alternative answers to account for this fact, such as females spending more time learning than males; males heing less serious about testing than females; the most talented males usually going to the science stream rather than majoring in foreign languages. Since all these alternative explanations are plausible, the validity of the answer is thus doubtful.

## VISUALIZING THE RESEARCH PROCESS

Very often, people visualize the research process in various ways and from different perspectives. In this section, two alternative views are described. The first one is proposed by Rudestam and Newton (1992) who visualize the research process as a wheel. The second one is suggested by the author who depicts a research process as a flow chart in which a series of tasks are presented in a sequence.

### The research wheel

According to Rudestam and Newton (1992), we may use a wheel as a metaphor to describe the stages of the research process. The metaphor indicates that a series of steps are repeated recursively over time. To be simplistic, the research process consists of at least two cycles and each cycle contains four stages (See Figure 1.2). The sequential activities within the first cycle include: empirical observation, developing a

proposition, constructing a theoretical framework and generating specific research questions. Those within the second cycle are data collection, data analysis, referring the results back to the conceptual framework and generating further research questions for additional studies.

Rudestam and Newton maintain that both induction and deduction are essential for the whole research process. The activities on the left-hand side involve a process of inductive logic and the activities on the right-hand side, a process of deductive logic.

### *The first cycle*

The first cycle starts with "empirical observation" through which the researcher chooses a topic. The next stage of the research wheel is to formulate a proposition which describes an established relationship in the form of a statement (e.g. Learning purposes are related to the choice of learning strategies). At the third stage, the researcher should relate the proposition to a conceptual framework. In other words, the researcher at this stage needs to propose a theoretical framework, based on relevant theories and previous studies, in which the proposition can be placed. The novice researcher usually finds this task the most demanding and taxing aspect of the thesis process. Moving forward around the wheel, the researcher is to generate specific research questions. Once the questions are specified and stated, the first cycle is finished.

### *The second cycle*

The second cycle begins with collecting data that directly addresses the research questions. The data-collection process can also be regarded as a form of empirical ohservation. Once the data are gathered, the researcher needs to analyze the data according to the research purposes. Results yielded from the analysis are generalizations made through induction. Then the generalizations are linked to the conceptual framework and further research questions and implications for additional studies are recommended.

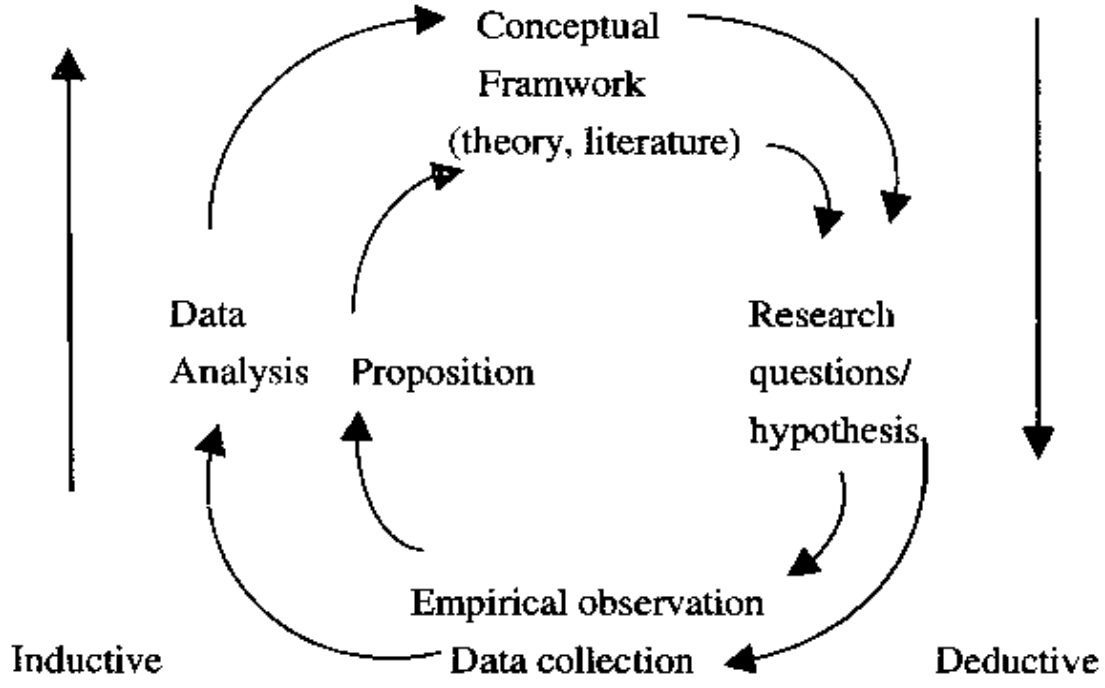



Figure 1.2: The research wheel

The wheel metaphor can successfully capture the dynamic aspect of research. However, it has been built up on only one type of research, that is quantitative in nature aiming at testing hypotheses. If a study is qualitative, the sequence is not the same as the one described above. This issue will be clarified later.

## The flow chart

Compared with the research wheel described above, the flow chart (See Figure 1.3) is much simpler and less technical. It focuses on the activities a researcher must undertake rather than on how these activities interact in the process of research. It is particularly suitable for helping novice researchers understand what they are supposed to do in their research.

According to Figure 1.3, research involves six tasks: developing research questions, reviewing the literature, selecting a research design, collecting data, analyzing the data and writing up a thesis. The arrow that links the last task with the first one means that once you finish your thesis, you may develop new questions for further research.

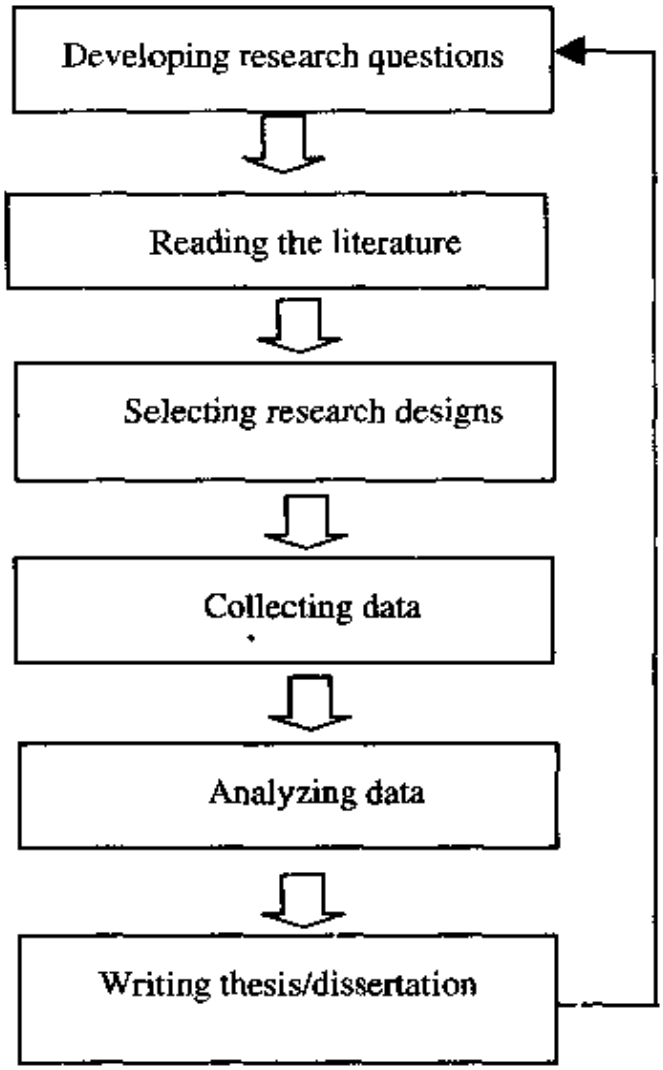



Figure 1.3: The flow chart for research

The six tasks each have their own role to play in the research process and none of them can be overlooked. If one task is not undertaken adequately, the whole study will be put in jeopardy. Furthermore, the tasks are logically interrelated and the sequential order cannot be changed in many cases. For example, data collection cannot be undertaken before

the development of research questions. Similarly, a thesis cannot be written before the analysis of the data. However, the sequence is not always rigidly fixed. In some cases, there may be options. For example, developing research questions, reading the literature and selecting research designs are not necessarily sequential in practice as presented in the flow chart. You may have a rough idea about your research interest and then go to read the relevant literature. Based on the existing literature, you specify your research topic and narrow down the scope. Then you go to read the relevant literature again to develop general as well as specific research questions. The selection of research designs is determined not only by research questions but also by the methods used in previous studies. Therefore, at the stage of constructing your research design, you also need to read the literature. The above description clearly shows that the three activities presented in the flow chart are not happening in a linear fashion and always in a fixed order.

As a would-be researcher, you had better follow the sequence first. Once the cycle is on the track, you are encouraged to be flexible and move back and forth within the sequence. Remember that "real research is inevitably going to be a rather messy process" (Blaxter et al., 1996: 7). In other words, these six tasks are recursive in nature. However, this recursiveness cannot be captured by this two-dimentional flow chart. One obvious advantage of the chart is its simplicity: it enables the reader to remember the tasks without much effort. Furthermore, the tasks described can be applied to both quantitative and qualitative studies.

## CLASSIFICATIONS OF RESEARCH

When you read books on research methods, you may come across various kinds of terms designating research. In this section, I will define some commonly-used terms along two classifying features: aims of research and source of data (See Table 1.1).

| Classifying features | Types of research |
|---|---|
| Aims of research | Theoretical /practical |
| Source of data | Primary/secondary |

Table 1.1: Classifications of research

## Theoretical and practical

In terms of different aims, two types of research can be identified: theoretical and practical. Theoretical research is primarily concerned with constructing theories or testing existing theories. Practical research is usually conducted by people who are directly involved in L2 teaching, such as L2 language teachers, L2 textbook compilers or L2 test designers. The major aim of practical research is to seek a practical solution to a problem in our daily life. In the following section, theoretical and practical research will be discussed one by one.

### *Theoretical research*

Theoretical research aims at developing or testing theories rather than at resolving practical issues. For example, the work undertaken by Krashen (1985) for constructing the "Monitor" theory can be regarded as theoretical research. His model consists of five hypotheses as follows:

(1) The acquisition-learning hypothesis
(2) The natural order hypothesis
(3) The monitor hypothesis
(4) The input hypothesis
(5) The affective filter hypothesis

The above theory was not invented by Krashen himself. For example, the Input Hypothesis was first proposed by Macnamara in 1972 (Krashen, 1985) and the Natural Order Hypothesis, by Corder (1967). Krashen

constructed the theory by studying, analyzing, and synthesising the relevant literature coupled with his own empirical studies.

Examples of such research can be easily found in international journals such as *Language Learning*, *Applied Linguistics* and domestic journals such as *Linguistics and Applied Linguistics* and *Foreign Language Teaching and Research*. Let's look at some examples.

Example One

The study intended to measure the effects of cultural background knowledge on L2 reading comprehension. The researcher chose two passages which were based in different cultures but with similar linguistic difficulty. One passage is about the Chinese Mid-Autumn Festival and the other is about the Thanksgiving Day in the United States. Sixty second-year non-English majors read these two passages each at a time. Once the reading was over, the students took a comprehension test. Eventually, the scores on the two comprehension tests were compared. The results showed that the students displayed a better comprehension in reading a passage about the Chinese Mid-Autumn Festivul (Ye, 1993).

Example Two

The study aimed at finding out to what extent students' pragmatic knowledge was related to their language proficiency. 90 second-year English majors were required to take a test of pragmatic competence. Then the students were divided into three groups according to their overall English proficiency: top, middle and bottom. A statistical procedure was taken to compare these three groups' pragmatic competence. The results indicated that students' L2 proficiency was closely related to their pragmatic knowledge (Wu, 1998).

The above two examples both attempted to test the existing hypotheses. It is generally agreed by researchers that cultural background knowledge is an important factor affecting the quality of reading comprehension. In Example One, Ye tested this assumption in her study in which non-English majors were involved and the result confirmed the assumption. In Example Two, Wu examined the hypothesis that pragmatic knowledge and L2 proficiency are closely related to each other and her empirical data were in support of the hypothesis. Although the findings from the above two studies have implications for L2 teaching and learning, they are not direct solutions to any practical problems.

For an M.A's or Ph.D. student, it is rare to write a thesis or dissertation exclusively on theoretical researcb since this kind of research requires a profound understanding of the topic you are investigating.

### *Practical research*

Practical research attempts to solve concrete problems in classroom teaching or learning or some other situations. The findings from such research usually can be directly tried out by practitioners. The following are the examples:

> Example One
>
> The study tried to see to what extent a language laboratory could be used to teach spoken English (Wen & Wu, 1998) in order to find a practical answer to the question: "How can we make 30 to 40 students active in a speaking class." 29 second-year students who participated in this experiment were from the same class. They had their speaking class in a language laboratory for four months. At the end of the semester, the students were asked to answer a questionnaire anonymously to make an evaluation of the speaking class. The questionnaire items concerned four aspects: (1) the students' attitude towards the speaking class in a language laboratory; (2) the amount of their participation in the language

laboratory in comparison with the speaking class in an ordinary classroom; (3) their degree of nervousness in the language laboratory in comparison with the speaking class before; (4) their rate of progress in spoken English this semester in comparison with their previous learning. Based on their responses to the questionnaire, we concluded that although the students noticed the limitations of the language laboratory setting, they made a positive evaluation of the speaking class in a laboratory.

Example Two

The study intended to see whether a spoken English test in a laboratory is feasible as a large-scale test format (Wen, 1999). The experiments were held for five years consecutively, in which several aspects were examined: (1) the content and the difficulty level of the test; (2) the administration of the test; (3) the evaluation of the tapes. The total number of subjects involved in the experiment was 3,300 second-year English majors from 60 different universities within five years. After each experiment, the researchers made modifications based on the students' and the teachers' responses to the test. Finally, such a test format has been adopted as the English spoken Test-Band 4 for English majors since May 1999.

The two exemplary studies mentioned above are primarily concerned with practical problems. In Example 1, the researchers wanted to find a method to teach spoken English more effectively in a large class and in Example 2, the researcher aimed at developing a large-scale spoken English test for English majors who spread out in different parts of China. Although the results from the two studies all had a direct impact on actual teaching and learning or syllabus design, or testing, there is no theoretical basis for us to claim that teaching spoken English in a language laboratory, or testing spoken English in a language laboratory is better

than other forms. Actually, the solutions are selected only due to practical constraints.

### *Aims of research: Discrete categories or a continuum?*

In the above discussion, I tried to give you various examples to illustrate the differences between theoretical and practical research according to their aims. You might get an impression that sucb a distinction is clear-cut and easy to identify. Actually it is not always true. In reality, such a theoretical-practical divide is defined only in a relative sense. They do not form discrete categories but a continuum. Moreover, in practice, one piece of research more often than not fulfils a double or triple purpose. It is not uncommon for theoretical research to have practical implications while practical research has theoretical value. The theoretical-practical distinction is sometimes blurred and they differ in degree rather than in kind.

## Primary and secondary

The distinction between primary research and secondary research depends upon the source of data. By saying primary research, I mean the data are collected directly from our lived experiences. These data have not existed in any documents before. They are first-hand and original information. Secondary research is a kind of study which makes use of data in documents, books and journals. These data have been collected by other people for their own purposes.

### *Secondary research*

Secondary research is often called documentary or library research. Let us first consider the following example of a term paper.

Suppose as a requirement of the course, MA students are asked to write a term paper on the topic "Select two learner factors which you think are most important in accounting for individual differences in L2 learning outcomes." To accomplish the task, the students use libraries to search for the written wisdom of other scholars. Suppose the students

select "motivation" and "learning strategies". First of all, they may search the books which include these two variables and then find out the papers which have reported empirical studies on them. They then synthesize diverse views from these secondary sources and various findings related to the issue. The resulting work, if not too bad, can provide some useful ideas about the topic at hand and the best papers can develop creative and productive insights into a given topic.

A review paper is a typical example of secondary research. The researcher reviews the recent work in a defined area and then summarizes, analyzes, evaluates, or synthesizes information that has already been published. Although the materials the researcher reviews are not new, the best review papers are insightful in the sense that they offer new syntheses, new ideas and theories. For example, a paper entitled "Research on language learning strategies: Methods, findings and instructional issues" written by Oxford and Crookall (1989) is a review paper which intends to "survey research on language learning strategies (LLSs)". The authors describe and evaluate various primary research studies on LLSs according to the research methods used. Finally they put forward some valuable suggestions for future research on this topic.

### *Primary research*

Primary research is also called empirical research because its data are derived from the primary source (e.g. students who are learning a second language or teachers who are teaching a second language), in contrast to secondary research, which depends on secondary sources (e.g. books about L2 learning and teaching). Primary research can be theory-oriented or practice-oriented. The following hypothetical studies are all primary research.

> Example One
>
> In order to find out how good learners and poor learners differ in reading strategies, the researcher asked 60 second-year non-

English majors to answer a questionnaire which contained 25 statements with a five-point scale ranging from "This statement is never or almost never true of me" to "This statement is always or almost always true of me." The responses were compared by statistical analysis.

Example Two

In order to find out how good learners and poor learners differ in reading strategies, 12 students, divided evenly between good and poor learners, were asked to read one passage while thinking aloud. The whole process was recorded individually and then transcribed. Their reading strategies that were thus idenditifed were categorized based on the verbal protocols and the categories were compared.

Example Three

The researcher aimed at examining to what extent second-year English majors used L1 in the process of L2 writing. 50 students were asked to answer a questionnaire that contained 15 items concerning whether L1 was used at different stages of writing. Each item was responded to on a three-point scale, i.e. never-sometimes-usually. The responses to the questionnaire were analyzed first by statistical procedures. Then the 8 students who were reported to use L1 most frequently and the least frequently were selected respectively out of the 50 for interviewing. In the interview, they each were asked to describe their L2 writing process in detail, and to explain how and why L1 was used in the process. The interview data were categorized and reported as supplements to the responses to the questionnaire.

Recently, due to the advancement of computer technology, quite a few linguistic corpora have been developed at home and abroad. These

corpuses can provide the researcher with recorded authentic speech or written texts. It is very convenient for a beginning researcher to work with an existing set of data since it is extremely difficult for any individual researcher to carry out large-scale data collection given that time and funding are limited. Do we call this kind of research primary or secondary? The data in these corpuses were collected by other people. However, they contain raw data in the sense that they have not been analyzed. In my opinion, this kind of research is still primary by nature.

In the above discussion, I tried to explain to you how primary research is different from second research. Actually, they each can hardly be conducted in isolation. Secondary research can only exist based on primary research, while primary research must start with secondary research.

## Requirements for graduate students

As graduate students in applied linguistics, you might ask, "What kind of research are we expected to do?" You are expected to learn to do all of the above: theoretical/practical research and primary/secondary research. For your theses, you have to carry out primary research that is supported by second research. In the case of writing a term paper, you usually do secondary research by searching the library, which may focus on findings from theoretical or practical research or both. One thing that has to be mentioned here is that a doctoral dissertation should report primary research which must have theoretical value.

## SUMMARY

Research is defined as a systematic approach to finding answers to questions. For a piece of work to be qualified as research, it must meet three requirements: (1) questions that are significant, original and answerable; (2) explicit and transparent research procedures that can be justified in terms of established principles in the discipline concerned; (3) answers that address the questions being asked. The process of research

may be metaphorically described as a wheel or a flow chart. The metaphor of a wheel emphasizes the dynamic aspect of research while the image of a flow chart stresses six distinct activities that have to be carried out in a sequence. A novice researcher had better first follow the sequence of these activities and later show flexibility. Research can be classified in terms of aims as theoretical or practical or in terms of its sources ,as primary or secondary. Postgraduate students are expected to learn to do various types of research. However, in Master's or Doctoral theses for applied linguistic programs, primary research is usually the major component.

## DISCUSSION QUESTIONS

1. How did you define "research" before you read this chapter?
2. How does the definition of research in this book differ from your own definition, and from the definitions you have read in other books?
3. Find out one research paper in a recently-published international journal and discuss the following questions with your classmates:
   1) What are the research questions listed in the paper?
   2) What are the procedures described in the paper?
   3) What are the answers to the questions?
4. What is the relation among the three components: questions, a systematic approach and answers?
5. Compare the two graphic representations of the research process and identify their similarities and differences.
6. What tasks does a researcher have to accomplish if a piece of research is conducted effectively?
7. How does the author classify different types of research?
8. Use examples to show what is theoretical/practical research..
9. In what ways can primary research be supported by secondary research?
10. If you are asked to carry out secondary research, what are you expected to do?

# 2. Fundamental concepts

This chapter will introduce to you some fundamental concepts which will repeatedly occur in the book. A good understanding of these concepts will facilitate your comprehension of the remaining chapters. The specific concepts to be discussed in this chapter are the following: hypothesis, theory, model, population and sample, variable, levels of measurement, and operationalization.

## HYPOTHESES, THEORY AND MODEL

No matter what kind of research you are engaged in, it is almost impossible for you to avoid the use of the concepts of hypothesis, theory or model. They are so essential that a total absence of these terms may lead to doubts on the scientific nature of the work. However, simply using the terms does not make work scientific.

### Different types of hypotheses

As was mentioned before, all research aims at finding the answer to a question. Very often the investigator has already formed a tentative answer to the question before the research starts, and the function of the research is to check whether the conjectured answer can be confirmed or refuted by the evidence. Such a tentative answer is called a hypothesis. To be precise, a hypothesis in this case should be defined as a declarative statement that describes the hypothetical relationship between two or more variables. It is not a wild guess. Instead, it is made based on previous research findings or established theories.

Suppose the research question is "What is the relationship between the use of L1 hy learners in learning English and their learning outcomes?" Logically speaking, there are four possible hypotheses about the relationship between them:

(1) $H_0$ = null hypothesis: There is no relationship between the use of L1 and English learning outcomes.

*(2)* $H_1$ = positive, directional hypothesis: There is a positive relationship between the use of L1 and English learning outcomes.

*(3)* $H_2$ = negative, directional hypothesis: There is a negative relationship between the use of L1 and English learning outcomes.

*(4)* $H_3$ = non-directional hypothesis: there is a relationship but the direction of the relationship is not specified.

The hypotheses ($H_1$, $H_2$, $H_3$ ) that are in contrast with the null hypothesis are termed alternative hypotheses. In other words, both directional and non-directional hypotheses are alternatives to the null one (See Figure 2.1).

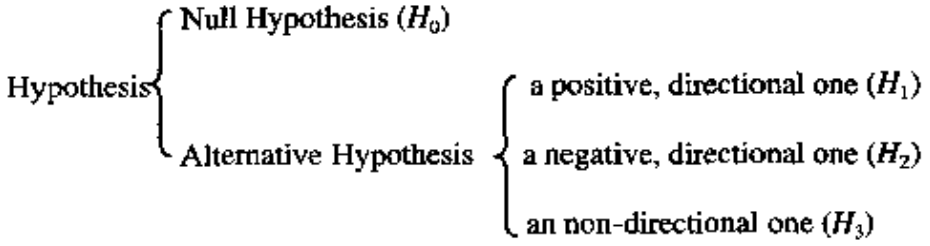



Figure 2.1: Different types of hypotheses

Given strong evidence from previous research, you may feel that it is possible to predict whether the relationship is positive or negative. Such hypodieses are directional. However, it is often the case that there are inconsistencies in previous research or even contradictory findings among various studies. Thus, you may not feel sure in which direction your hypothesis should follw. In this case, you may put forward a non-directional hypothesis, i.e. there is a relationship between the use of L1

and English learning outcomes. Or you even do not know whether there is any relationship at all. Consequently, you formulate a null hypothesis which predicts no relationship between the two variables. Nevertheless, no statistical procedures can directly confirm or disconfirm a hypothesis, i.e., an alternative hypothesis. What they can do is to reject or accept the null hypothesis (See detailed explanations in Chapter 8). When the researcher does not explicitly state a null hypothesis, the logical reasoning behind the research is inevitably related to the null hypothesis.

Instead of forming hypotheses in advance and testing them in the designed research, some studies that are exploratory in nature do not have any hypothesis in the first place. The researcher keeps open to all possibilities. The answer to the question gradually emerges in the process and becomes clear in the end. These eventual answers more often than not serve as hypotheses for confirmation in future research. For example, in the study carried out by Wen and Guo (1998), the researchers did not know how L1 was used in the process of English picture composition when they started collecting data. The findings yielded from the study are hypotheses. Such hypotheses need to be tested by other studies. Furthermore, the hypotheses in this case may not be concerned with the relationship between two or more variables but with variables themselves (See the hypotheses in a theory about L2 sentence-generating in the next section).

To sum up, hypotheses-formulation is essential for all research although the hypotheses may be constructed at different stages of a study. The research may start with a hypothesis or end up with a hypothesis. In the former case, it is a process of confirming a hypothesis and in the latter, it is a process of generating one.

## Theory

A theory is broader than a hypothesis, because it is a complete account or worldview about the phenomena studied. A theory is usually capable of generating new hypotheses within the general framework of

understanding that it provides. Like hypotheses, a theory is testable and falsifiable. A typical example is the theory of second language acquisition proposed by Krashen[1] (1985), which consists of the following five hypotheses (pp. 1-3):

(1) The acquisition-learning hypothesis

Acquisition and learning are two independent ways of developing L2 ability. Acquisition is a subconscious process which is similar to the process of learning one's native language while learning is a conscious process in which the learner only obtains L2 linguistic knowledge.

(2) The natural order hypothesis

The order in which learners acquire the rules of language is predictable but it does neither depend on the perceived formal complexities nor the sequence of teaching.

(3) The monitor hypothesis

Conscious knowledge obtained through learning can oaly be used to serve as an editor under two conditions: The first condition is that the learner must be consciously concerned with accuracy and the second condition is that the learner must know the rule concerned.

(4) The input hypothesis

L2 language proficiency develops when the learner is exposed to sufficient comprehensible input which is a bit beyond the learner's current level of competence.

(5) The affective filter hypothesis

The affective filter is a hypothesized mental barrier which may be up or down. When the learner is poorly motivated

---

[1] Many people have pointed out that his "theory" cannot be falsified and is therefore not a theory.

and has a high level of anxiety, the filter is up, which will prevent the input from reaching the long-term memory.

Theories can provide you with direction to or the focus on research by pointing to variables relevant to the study. Krashen's theory emphasizes the degree of consciousness of learning, acquisition order, monitoring, comprehensible input and affective states. Your study may select any one of the above variables as the focus. Theories also specify the relationships between the variables. In the above theory, Krashen proposes that the less the confidence and the higher the degree of anxiety level the learner displays, the less input the learner can process. Such hyputhetical relationships can serve as a basis on which you formulate hypotheses for your own study.

Instead of being motivated by others' theories, your own study may develop a theory which may stimulate others to conduct research. For example, in the study of Wen and Guo (1998) in which the process of L2 picture composition was examined with a focus on the use of L1, they proposed a theory based on the students' verbal protocols. The theory describes the active use of L1 in L2 sentence-generating when the learner is given a picture as a stimulus. The theory consists of the following hypotheses:

(1) Once L2 writers receive a pictorial stimulus, they start sentence-generating by choosing one of three paths: (a) Pictorial stimulus ⇒ L2; (b) Pictorial stimulus ⇒ L1, mediator for retrieving L2; (c) Pictorial stimulus ⇒ L1, mediator for generating ideas.

(2) L1 serves as a mediator for generating ideas, retrieving L2, monitoring L2, coafirming content and controlling L2 writing procedures.

(3) No matter which path L2 writers choose, none of them can totally avoid the use of L1 in producing L2 sentences.

Furthermore, they rely on the use of L1 for different purposes in different situations.

The above theory was further examined recently by Wang (2000) who found that the functions of L1 use listed above are far from complete.

## Model

Very often the terms "model" and "theory" are used interchangeably but they do show a difference sometimes. A model refers to the graphic representation of a theory while a theory refers to the verbal description alone. For example, Krashen's theory can be diagrammatically described as a model illustrated in Figure 2.2 (Krashen, 1989).

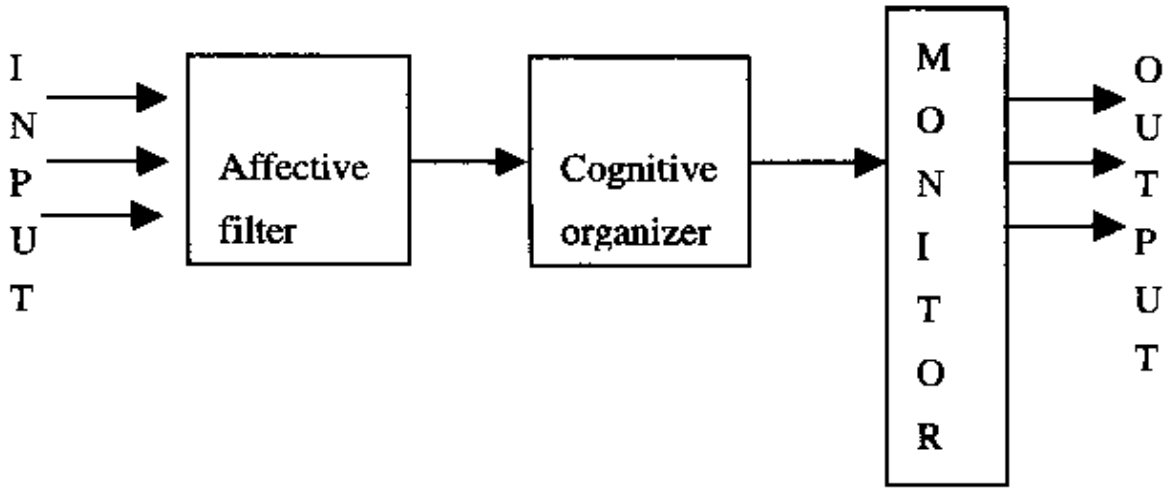



Figure 2.2: A model of second language acquisition

Obviously, a graphic form is more vivid and straightforward than a verbal account, but the graphic form alone is not sufficient for understanding. A verbal account describing the interrelations among the variables is more explicit and less ambiguous than a graphic form. Therefore, it is recommended that in thesis/dissertation writing a graphic form and a verbal account be presented side by side so that readers can easily comprehend it.

# POPULATION AND SAMPLE

## Definitions

The researcher defines the population as whatever the researcher is studying. It can be a collection of persons, textbooks, or test papers. However, in most cases, it refers to people. For example, if a researcher intends to measure the secondary school students' English proficiency level in Nanjing, the population consists of students. A researcher might be interested in evaluating the difficulty level of Matriculation English Tests in the past 10 years. The population is then made up of Matriculation English Test papers.

A sample is a representative portion of the whole target population, which is typically obtained through random sampling. In random sampling every unit of the population has an equal opportunity to be selected into the sample. The detailed description of random sampling techniques will be presented in Chapter 6.

## Relation between population and sample

Since a sample is randomly selected, it is assumed to be a miniature of the population. By studying the sample, the researcher hopes in the end to make inferences about the whole target population from which the sample is drawn.

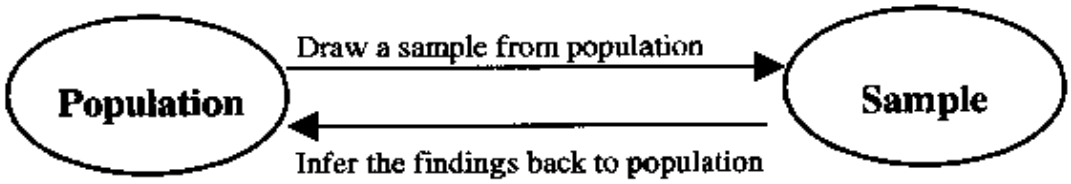



Figure 2.3: populations and samples (Punch, 1998:106)

As shown in Figure 2.3, you first draw a sample from the population and then work at the sample. Once you have finished the work with the sample, you will get the results concerning the sample. However, you do

not stop there. The next step is an attempt to generalize the results from the sample back to the population. How much generalizing power the results have depends on to what extent the sample can represent the population.

## VARIABLES

Brown (1990) emphasizes the importance of understanding the concept of variable by saying "The variables are, after all, the focus and center of any study"(p. 9). In this section, I will first introduce to you the definition of the term "variable" and then discuss different types of variables in terms of their functions and their levels of measurement in research.

### Definition of a variable

A variable, as its name suggests, is a feature that must be varied from person to person or from object to object. Some variables have a very limited number of variations. For example, gender only has two, either male or female. Marital status has four variations: unmarried, married, divorced, or widower/widowed. Education level may have four or more variations depending on how the education level is measured. It may include: primary education, secondary education, tertiary education and postgraduate education. However, some variables have a lot more variations. For example, the scores students got from an exam can be varied from 0 to 100. People's height and weight are even more varied than test scores.

With reference to the functions of variables in a study, we have independent variables, dependent variables, moderator variables, control variables and intervening variables. The following section will describe each type of variables.

## *Independent and dependent variables*

Independent and dependent variables are interrelated. They each cannot exist without the other. In other words, independent variables cannot exist without dependent variables, or the other way around. Therefore, these two kinds of variables always go together.

Independent/dependent variables can be used in two ways. In the first case, an independent variable is a stimulus variable or input which is manipulated by the researcher while the dependent variable is a response variable or output. Suppose the researcher studies the causal relationship between the two variables of X and Y. If their relationship can be explained as "What will happen to Y, if X is made greater or smaller?" then the researcher is thinking of X as an independent variable and Y as a dependent variable.

In the second case, an independent variable is the variable that is used to predict another variable while the dependent variable is the variable predicted by an independent variable. These two variables are correlated to each other but the relationship cannot be defined as a causal one on the basis of the evidence available. For example, one researcher carried out a correlation study to find out to what extent gender can predict one's L2 achievement. Then Gender in this study is an independent variable (a predictor) and L2 achievement is a dependent variable (a predicted factor).

To sum up, in the first case, it is clear that the independent variable is a cause which produces changes in the dependent variable; whereas in the second case, the independent variable is a predictor which is used to predict the dependent variable.

A study may simply have a single independent variable and a single dependent variable. Or it may include a series of independent variables but only one dependent variable, or the other way around. Now I will give you a number of hypotheses where independent and dependent variables have been identified.

Example one:

Hypothesis: L2 learners who have done six revisions of the same composition progressed faster than those who simply wrote six different compositions without revisions

Independent variable: Methods of teaching writing

Method 1: multiple revisions;

Method 2: multiple compositions without revision

Dependent variable: Improvement in L2 writing

In Example One, the researcher assumes that the method of revising the same composition several times is more effective than writing several compositions without revision in developing L2 writing abilities. In other words, the method of multiple revision can produce more positive effects on the development of L2 writing than the method of multiple compositions. Therefore, methods of teaching writing is an indepenent variable while L2 writing improvement is a dependent variable.

Example Two

Hypothesis: L2 learners who use more varieties of learning strategies and use them more frequently learn English better than those who use fewer varieties of learning strategies and use them less frequently.

Independent variable: Varieties and frequency of the strategy use

Dependent variable: English learning outcomes

In Example Two, the researcher hypothesizes that the use of learning strategies can determine L2 learners' English learning outcomes. The more varieties of the learning strategies they use and the more frequently the learning strategies they use, the bettter the English outcome they can achieve. That is to say, there is a causal relation between the strategy use and the learner's English

learning outcomes. The strategy use is a cause, or an independent variable while the English learning outcomes is a result or a dependent variable.

Example Three
Hypothesis: Students' perceptions of a 'good' teacher are in part predicted by their attitude toward education.
Independent variable: Students' attitude toward education
Dependent variable: Their perceptions of a "good" teacher

In Example Three, the researcher hypothesizes that students' attitude towards education can predict their perceptions of a "good" teacher. Therefore, the former is an independent variable or a predictor while the latter is a predicted variable or a dependent variable.

Example Four
Hypothesis: Students' way of learning English is determined by their beliefs about what can lead to success in learning.
Independent variable: Learners' beliefs about English learning
Dependent variable: Learners' way of learning

The hypothesis in Example Four means that students' beliefs about the way of English learning is closely related to their actual behaviors in learning English. In other words, students' beliefs can be used as an independent variable to predict the dependent variable, i.e. their actual English learning behaviors.

### *The moderator variable*

A moderator variable is measured to see whether it modifies the relationship between the independent variable and the dependent variable. It is sometimes called a secondary independent variable. The word

"moderator" indicates the reason why this secondary independent variable has been singled out for study. If the researcher is interested in studying the effect of independent variable X on dependent variable Y but suspects that the nature of the relationship between X and Y is altered hy a third factor Z, then Z is called a moderator variable.

Consider two examples. First, the researcher wanted to compare the effectiveness of a visual approach (using pictures) to an auditory approach (using audio tapes) for teaching an English lesson. First of all, she randomly assigned the students into two classes: one for the visual approach and the other one for the auditory approach. After three months' experiment, all the students were tested together for achievement, and the results of the two approaches appeared to he the same. The researcher suspected that students' preferred learning styles might affect the results since some students had propensity to receive the input through eyes or ears. Thus she designed a questionnaire to measure students' preferred learning styles. Then she separated visual learners from auditory ones and the blurred picture became clear. The visual and auditory approaches were more effective if they matched with learners' preferred learning styles In this case, learners' preferred learning styles would be seen to moderate the relationship between instructional approacb (the independent variable) and effectiveness of the approach (the dependent variable).

A second example is a study of the relationship between the conditions under which a test is taken (the independent variahle) and test performance (the dependent variable). Assume that the researcher varies test conditions between ego orientation ("Write your name on the paper, we are measuring you") and task orientation ("Don't write your name on the paper, we are measuring the difficulty level of the task"). The test-taker's test anxiety level is analyzed as a moderator variable. The results show that the persons with a high level of test anxiety did better under task orientation and the persons with a low level of text anxiety

performed better under ego orientation. Then there is interaction between the independent variable, moderator variable, and the dependent variable.

The situations in second language teaching and learning are usually quite complex and therefore, the inclusion of at least one moderator variable in a study is highly recommended. Often the nature of the relationship between X and Y remains poorly understood because of the researcher's failure to single out and measure vital moderator variables — Z, W, and so on. Let us look at more examples.

Example one

Hypothesis: Male researchers get more effective performances from the whole group of subjects than do female researchers, but the female researchers are more effective with male subjects.

Independent variable: The sex of the researcher

Moderator variable: The sex of the subject

Dependent variable: Effectiveness of performance of subjects

Example two

Hypothesis: Using L1 in the process of learning English has a negative effect on students' learning outcomes but it has a stronger negative effect on adults than children.

Independent variable: The use of L1

Moderator variable: Age

Dependent variable: Learning outcomes

Example three

Hypothesis: Management strategies have stronger effects at the advanced level of learning than at the elementary level of learning.

Independent variable: Management strategies

Moderator variable: Proficiency level

Dependent variable: Learning outcomes.

### *Control variables*

It is impossible to study all of the variables in an environment at the same time; some must be neutralized to guarantee that they will not have a moderating effect on the relationship between the independent variable and the dependent variable. The variables whose effects are neutralized or kept constant are called control variables. The purpose of such control is to cancel out or neutralize any effect the variables might otherwise have on the observed phenomenon. For example, in a study on the effects of outside reading on the development of reading comprehension abilities, the researcher purposely selected all the students who had scores above 80 on College English Test-Band 4. In this way, the subjects' overall English proficiency was selected as a control variable. In another study, the researcher intended to see to what extent testing strategy training might affect the test scores on the English test. The subjects chosen were all girls and thus gender was selected as a control variable in the study.

Certain variables appear repeatedly as control variables, although they occasionally serve as moderator variables. For example, sex, intelligence, and learners' previous performance are three variables that are commonly controlled. In constructing a study, the researcher always needs to decide which variables will be studied and which will be controlled.

### *Intervening variables*

The term intervenings or extraneouss is an umbrella term that can refer to any variables that are not measured in a study but produce effects together with the independent and moderator variables on the dependent variables. There are two types of intervening variables. The first type includes the variables that are difficult, if not totally impossible, to measure even if the researcher wants to. The second type refers to the variables that can be measured but the researcher does not want to study these variables due to limited time and resources or the researcher ignores them due to insufficient research experience. Very often these two types of intervening variables are present simultaneously. For example, you

plan to conduct a study aiming to find out the relationship between the amount of reading (Independent variable) and the size of L2 vocabulary(dependent variable). Between these two variables, there might be a list of other variables functioning at the same time, such as the previous vocabulary size, motivation, the strategies used in memorizing words, the ability of memorizing and the process of taking in, storing and retrieving words. The first three variables can be measured if you decide to, while the last one, i.e. the process of taking in, storing and retrieving is not easily measured even if you are curious about it.

Suppose in another study, you intended to contrast presenting an English lesson on closed-circuit TV with presenting it via live lecture. The independent variable in the study is the mode of presentation; the dependent variable is some measure of learning. Through three months' experiment, the final result was that students learned English better through live lecture than through closed-circuit TV. Apparently, it is the mode of instruction that had the effect on learning outcomes. If you asked why one way of instruction was more effective than the other, your answer might be live lectures could provide students with a lot of opportunities to participate in class activities while closed-circuit TV was one-way communication. In this case, the amount of students' participation was an intervening variable, which produced effects on students' learning outcomes. Actually, you can identify a list of intervening variables in this study such as the opportunity to replay the TV presentation, the students' unfamiliarity with the use of closed-circuit TV, etc.

You, as a good researcher, must be able to identify the intervening variables in your study. Knowledge of intervening variables can help you explain why the independent variable causes changes in or predicts the dependent variable. Furthermore, a clear understanding of the intervening variables existing in your study leads to a cautious interpretation of your findings which, otherwise, tend to be overclaimed, or even wrong.

### *The relationship among variables*

How are the above five types of variables interrelated? The following diagram (Figure 2.4) adapted from Brown's (1990, p.13) illustrates such relationships. The relationship between independent and dependent variables is either causal or correlated. Moderator variables are secondary independent variables, which can modify the relationship between independent and dependent variables. Control variables are the variables that are kept constant during the study in order to examine the robust effects of the independent variable on the dependent variable. Intervening variables are not observable or not measured and thus placed in a dotted arrow. The more intervening variables there are, the harder it is to claim a causal relation.

The function of a variable is not inherent in the variable itself. Thus, the conceptual status for any variable can change from study to study, or from part to part within the same study. A variable such as L2 motivation may be an independent variable in one study, a dependent variable in another, a moderator variable in a third, a control variable in a fourth, and an intervening variable in a fifth. The researcher, however, must clearly specify the conceptual status of each variable at each stage of the study.

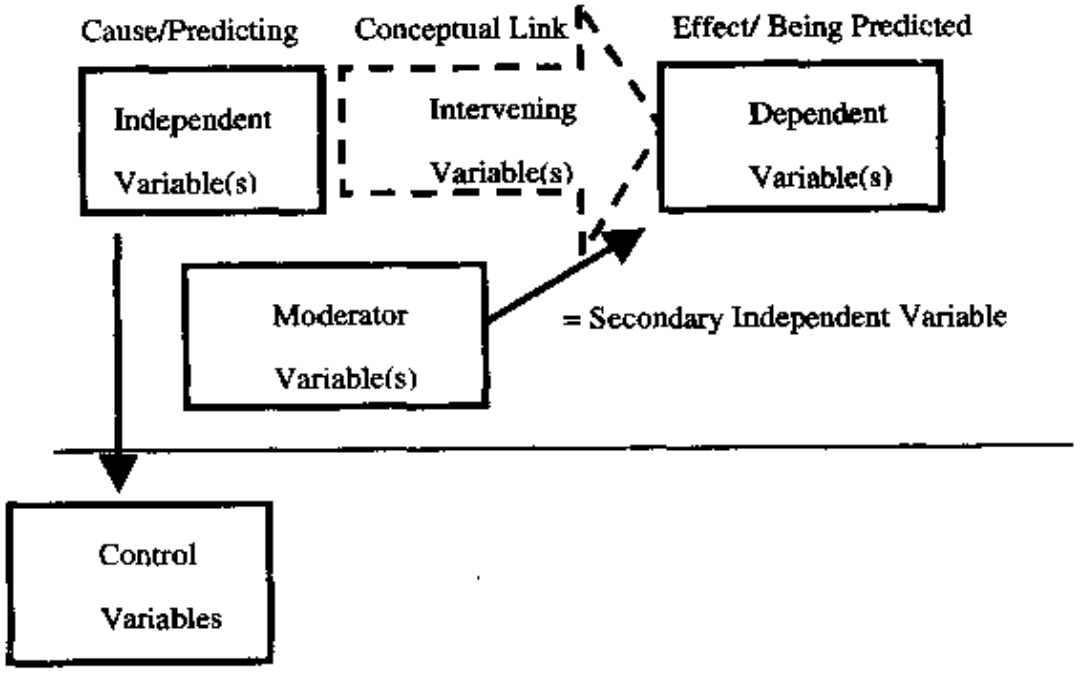



Figure 2.4: Relations between different types of variables

## LEVELS OF MEASUREMENT

In addition to the functional differences, the variables also show differences in their levels of measurement, i.e. the ways in which they are measured. Generally speaking, there are three distinct levels of measurement[2] in the field of applied linguistics: nominal scale, ordinal scale and interval scale. Accordingly, there are three kinds of variables: nominal variables, ordinal variables and interval variables. The following section will discuss these three scales one by one.

### Nominal scale

A nominal scale, as its name indicates, is used to name objects or classify objects. It contains a list of categories, which are mutually exclusive. In other words, each category is unique. For example, you may classify individuals based on their first language, sex, or nationality. This type of measurement may seem rather simple, but many of the variables of greatest interest are nominal in nature. Among the more commonly studied nominal scales are sex, occupation, and mother tongue.

Nominal measurement is qualitative in nature. When you assign numbers to categories on a nominal scale, you must remember that numbers are only used as labels of categories and nothing else and you do it simply for the convenience of operating some statistical analyses on qualitative data. For example, you assign the number 1 to males and the number 2 to females or the other way around. Then you may easily find out the percentage of males and females in your study. Such practice does not make the variable quantitative because the number 2 here is not twice as big as the number 1 and it cannot be added, subtracted,

---

[2] In many books under the topic of levels of measurement, the ratio level is regarded as a distinct level separating it from the interval level. In this book, I refer to ratio scales as interval ones because "it is common practice in the social sciences" (Bernard, 1994:34).

multiplied or divided. The numbers assigned are arbitrary since it makes no difference what number you assign to what category, so long as each category has a unique number. Nominal scales thus possess the property of distinctiveness. The following examples are nominal scales.

| Example 1: Sex | Example 2: Marital status |
|---|---|
| 1=male | 1= single |
| 2= female | 2= married |
| | 3= divorced |
| | 4= widowed/widower |

| Example 3: Mother tongue | Example 4: Nationality |
|---|---|
| 1= Chinese | 1= American |
| 2= Russian | 2= French |
| 3= English | 3= German |
| 4= Japanese | 4= British |
| 5= Korean | 5= Australian |
| 6= Swedish | 6= Canadian |

## Ordinal scale

An ordinal scale provides information about the relative amount of some trait possessed by objects, in addition to naming it. For example, in an English speech contest, you might select the six best speakers and award them prizes. These selected speakers might be further ranked as the first prize winner, the second prize winner, the third prize winner, and so on. Such a rank order is an ordinal scale which can indicate who is better than who. However, the difference between the first prize winner and the second one or the difference between the second prize and the third prize winners cannot be measured precisely. In another case where students' compositions are evaluated, instead of giving them precise scores, you give them five categories, namely A, B, C, D and E. It is clear that the students who got A did better than the students who got B.

But what is the exact difference between A and B or B and C? This kind of information is absent in an ordinal scale.

In the process of data-coding, you often use Arabic numbers to represent the ordinal information. For example, you might use 5 to represent A and 4 to represent B. However, assigning numbers to the ordinal data still cannot quantify the exact distance between every two ranks. If you want to compare two ordinal variables, the direct way is 1< 2 <3 or 3> 2>1. Or you may say Zhang did better than Li in the English speaking test but less well than Zhu.

## Interval scale

In contrast to nominal and ordinal scales, interval scales can provide information about the distance between two attributes. In other words, interval measurement has the property of equal distances between every two consecutive points on the scale. For example, the difference between one minute and two minutes is the same as the difference between four minutes and five minutes. Similarly, the interval distance between two metres and three metres is the same as the distance between four metres and five metres. Therefore, the measurement of time and length is an interval scale.

In education, a 100-point marking system for a test usually involves an interval scale because the distance between every two successive points on the scale is the same, namely one point. If two subjects obtain scores 95 and 100 respectively, you say that one is better than the other to the extent of the value of the five points. In other words, you assume that each of the five points has equal value. Assigning number to scores from 1 to 100 is based on the assumption that the intervals between 1 and 2 or 5 and 6 or 99 and 100 are of equal value.

Nevertheless, it can be well argued that the above view is being overly simplistic here. Items on the test may vary considerably in difficulty. Therefore, the points earned by doing different items successfully are not of the same value. As a general practice, however, this difficulty is

ignored in education, and such measurements are considered to be interval.

The difference between an ordinal scale and an interval scale is shown in Figure 2.5 which is adapted from Bachman (1997: 28). For instance, the student who ranked the first got 95 points and the one who ranked the second got 94.5 points. The difference between them is only 0.5 points. But the person who ranked the third got 90 which is 4.5 points lower than the second rank. Obviously, the distance between the first rank and the second one, and the distance between the second and the third are not equal at all. By contrast, the distance between any two consecutive points on an interval scale is one point.

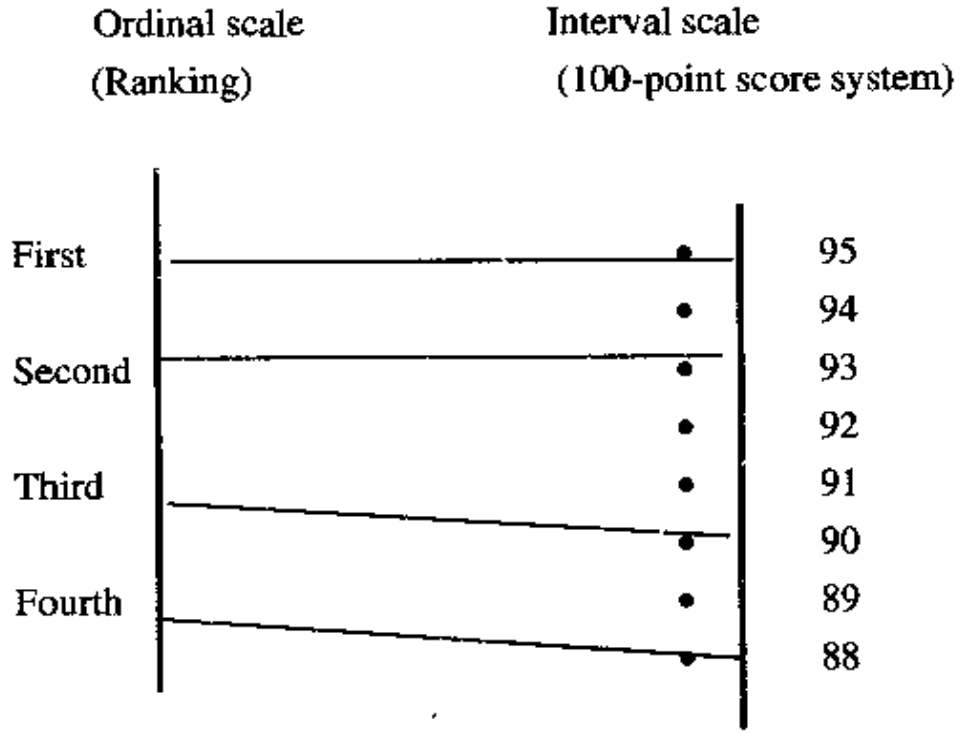



Figure 2.5: Comparison between ordinal and interval scales

Ideally, when you measure how much of a variable is present on an interval scale, you expect the intervals tu be of equal value. In some cases there can be little argument about the value of intervals, such as time and weight; in other cases you may not be so sure. For example, if age is the variable you are researching, you can assume that a year is a

year. But the value of a year may differ along the scale for second language learning. Obviously the differences between each year in 10- to 20- year range may be much smaller in value than the year intervals between 60 and 70.

Recent studies tend to treat some traditionally-called ordinal scales as interval scales for the convenience of statistical analysis (Boyle, 1970; Labovitz, 1971). For example, there is no instrument that can give us precise and accurate measurement of people's interest in learning a foreign langauget. What you might do is to operationally define the variable in terms of three continuous categories: not interested, interested and very interested. Then you rank the subjects along this continuum. In the process of coding such data, you assign a number to each category. In this case, you code the answer "not interested" as 1, "interested" as 2 and "very interested" as 3. Then 1-2-3 are regarded as interval scales and the means are calculated to compare. Theoretically speaking, such an analysis is problematic since the distance between "not interested" and "interested" , and between "interested" and "very interested" cannot be measured in the first place, nor can we speak of equal distance between every two consecutive points. However, in practice, many studies in social sciences do take such responses to questionnaire items as interval scales. It seems to me such practice has already been widely-accepted (Rudestam & Newton, 1992).

## Relations among the scales

Nominal, ordinal and interval scales form three levels of measurement in terms of precision. The nominal scale is at the lowest level, the least precise; the interval scale, at the highest level, the most precise; and the ordinal one is in between. Nominal scales only categorize the data. Ordinal scales not only categorize hut also indicate the ordering of those categories. Interval scales, in addition to the information about categories and their ranking, indicate the distance between the points in the ordinal

ranking. The three levels of scales along with their characteristics are summarized in Table 2.1 which is adapted from Brown (1990: 23).

| Levels of measurement | | Category | Ordering | Distance |
|---|---|---|---|---|
| highest level ↑ | Interval scale | + | + | + |
| | Ordinal scale | + | + | — |
| lowest level | Nominal scale | + | — | — |

Table 2.1: Levels of measurement and their features

The scales at the higher level can be converted into a lower level of scale but not the other way around. Let's take students' proficiency level for example. You may easily change exact scores on a 100-point scale into five bands: A, B, C, D and E. However, it is not possible for you to convert the rank order of the five bands into the exact scores.

Whether a variable is placed on a nominal, ordinal, or interval scale is usually decided by the researcher. The rule of thumb in research is "always measure things at the highest level of measurement possible. Don't measure things at the ordinal level if you can measure them intervally" (Bernard, 1994: 35).

## OPERATIONALIZATION

Whatever variables are to be involved in your study, they have to be defined so explicitly that you can measure them without difficulty. This process is called operationalization. In this section, I will discuss first the differences between conceptual and operational definitions and then how a variable is defined operationally.

### Conceptual and operational definitions

A variable needs to be defined conceptually and operationally. A conceptual definition is a theoretical explanation that helps clarify people's understanding of the variable investigated. It is also called a

construct, which will be used as an equivalent to the conceptual definition of a variable in this book. The operational definition contains specified instructions which explicitly show how a variable can be measured. The following examples illustrate how a conceptual definition is different from an operational one.

Let us first look at one example from our daily life. You might have heard someone say, "Zhang's apartment is spacious." What does "spacious" mean? *Longman Dictionary of English Longuage and Culture* defines spacious as "baving a lot of room"(1993, p. 1271). By this definition, it is still difficult to decide whether someone's apartment is spacious or not because people have different conceptions of "spaciousness". In a study of people's housing problem, the researcher defines spaciousness as the average living area of each person being more than 15 square metres. By the researcher's definition, it is easy for us to make a comparison between the areas of two apartments. In this case, the dictionary definition is conceptual while the definition given hy the researcher is operational.

The second example is from the field of second language research. Gardner (1985) sees motivation as "the combination of effort plus desire to achieve the goal of learning the language plus favourable attitudes toward learning the language"(p.10). That is to say, motivation includes three components: attitude, desire and effort. Operationally, in his study, he regarded motivation as the responses to the questionnaire items concerning these three components.

Although a conceptual definition facilitates our understanding of the concept, it is of limited use because it does not allow us to measure anything. Without measurement we can not make direct comparisons. An operational definition permits the researcher not only to measure the variable investigated but also to discuss it with another researcher and evaluate whether the measurement specified by the operational definition is appropriate or not.

In the field of applied linguistics and other social sciences, theoretical definitions often occur in literature reviews and operational definitions in descriptions of methodology. In defining a variable conceptually, you have to notice that the conceptual definition of the same term is varied in the literature due to the researchers' different views or due to the development of theories. In this case, you have to take a position and select/adapt from one of them with justifications. Take L2 language proficiency for example. L2 proficiency was once defined as knowledge about grammar, vocabulary and phonology. In the 50s and 60s, it was defined as four skills: listening, speaking, reading and writing (Lado, 1961). In the 70s, it is regarded as communicative competence (Hymes, 1971). In the 80s and 90s, the concept of L2 proficiency is further expanded to include strategic competence (Ellis, 1994). If you want to investigate your students' L2 proficiency, you face several choices. Which conceptual definition you choose will directly affect your operational definition. Suppose you view L2 proficiency in terms of the four skills. Your operationally defined L2 proficiency must accordingly be related to tests of the four skills. That is to say, the two kinds of definition of the same term within a study must be compatible and logically related. Similarly, suppose you believe L2 proficiency consists of linguistic competence, pragmatic competeuce and strategic competence. If the test you select to measure your students' L2 proficiency is TOEFL, which does not aim at examining pragmatic and strategic competence, your operational and conceptual definition are incompatible.

However, compatibility does not mean that the operational definition must measure virtually everything described in the conceptual definition. Often, due to the weakness of instruments and limitations of time and resources, what is actually measured is less than what is conceptually defined. In other words, operational definitions are necessarily incomplete representations of conceptual definitions. For example, no

single English test score could represent all aspects of the construct of L2 proficiency.

Furthermore, the operational definitions used in various studies may be different even if the conceptual definitions are essentially the same. For example, some researchers may define language aptitude operationally as the scores on the Modern Language Aptitude Test (MLAP) developed by Carroll and Sapon in the 50s while the others, as the scores on the Language Aptitude Battery (LAB) constructed by Pimsleur in the 60s. Although both MLA and LAB claim to measure language aptitude, they actually do not measure exactly the same aspects. Skehan (1989) points out "Compared to the MLAT, the LAB emphasizes inductive language learning capacities and auditory ability....What the LAB does not have, in comparison with the MLAT, is a test of grammatical sensitivity, on the one hand, and any effective coverage of memory."(p. 29)

Logically speaking, you should start with a conceptual definition of a variable and then move to its operational definition. In practice, the process more often than not is non-linear. For example, initially, you might like to follow the most comprehensive view of L2 proficiency and define it as a combination of linguistic competence, pragmatic competence and strategic competence. Later you might realize that the authoritative test that can measure the three components simultaneously is not available at all. If you stick to the conceptual definition, you are not able to operationalize it. Therefore, you have to modify your conceptual definition.

## Operational definitions

Operational definitions specify exactly what you have to do in order to measure something that has been defined conceptually. Such definitions are characterized by explicitness which makes possible not only measurement but also evaluation and replication. The following

examples are all operational definitions from my own study on the relationship of modifiable learner variables to learning outcomes:

(1) L2 language proficiency:

The scores on the Test for English majors-Band 4.

(2) Effort:

The amount of time spent outside class studying L2 within a week as reported by the subjects.

(3) Learning purpose:

The responses to the questionnaire items presented in the appendix, concerning learning reasons such as interest in the language and culture or longing for a better career in the future.

(4) Management strategies:

The responses to the questionnaire items presented in the appendix, concerning strategies used in planning, goal-setting and evaluation of the learning process and the learner's personal characteristics.

If you stick to those definitions in constructing measurements, then you can compare the findings when the subjects are from different universities. For example, you can tell if the students in University A scored higher than the students from University B on the Test for English Majors-Band 4; you can tell if the students in University A spent more hours outside class studying L2 within a week than the students in University B.

How is a variable operationalized? By operationalization, you should first study the corresponding conceptual definition and make it clear what aspects you plan to measure in your empirical study. For example, in my study, the variable "Effort" is conceptually defined as the quantity and quality of time spent on learning L2 in and outside class. In other words, the conceptual definition contains two components: the quantity and

quality of time. It is obvious that both the quantity and quality of time are important for English learning outcomes. However, it is difficult to measure the quality aspect. Furthermore, the subjects involved in my study were second-year Englisb majors who had a more or less similar number of English lessons every week. Therefore, the operational definition of the variable "Effort" only focuses ou the quantity of time spent outside class learning English. Once I had chosen which aspect to be investigated, I had to think how to measure this selected aspect. Eventually, I decided to ask the subjects to report how many hours they spent outside class learning English within a week. Finally, "Effort" is operationally defined as the amount of time spent on L2 learning outside class reported by the subjects. To sum up, you operationalize a variable by first selecting which aspect(s) in its conceptual definition is to be examined and then by making the selected aspect(s) measurable.

The quality of operational definitions is essential for research. If an operational definition is bad, so are all the conclusions drawn from using it to measure something, e.g. if you defme L2 proficiency as scores on a test designed by a single teacher, which is obviously questionable in terms of validity and reliability. By defining a variable operationally, the flaws in your measurement can be easily detected by your supervisor or other researchers and will be eradicated at the initial stage of the study.

## SUMMARY

Fundamental research concepts discussed in this chapter include hypothesis, theory, model, population and sample, variable, levels of measurement, conceptualization and operationalization. A hypothesis is a conjectured answer to the research question that is developed out of research findings or theories and is expressed in the form of a declarative statement. It may be tested in a new study, or generated from a study to be proved in future research. A null hypothesis explicitly or implicitly stated in a study is a statement in which no relation is assumed between two or more variables. Alternative hypotheses are either directional or

non-directional. A set of hypotheses with specified relations among the variables form a theory, which can be described verbally or graphically. In order to differentiate these two modes of representation, the verbal account is usually called a theory while the graphic description is called a model. A population is the whole of whatever a researcher is investigating and a sample is a representative portion of the target population. The results from a sample are often used to infer information about the population. Variables are attributes of a person or an object that can be varied. With reference to their functions, variables are differentiated into independent and dependent variables, moderator variables, control variables and intervening variables. In terms of the levels of measurement, variables are grouped into nominal variables, ordinal variables and interval variables. All the variables involved in a study have to be defined both conceptually and operationally. A conceptual definition is a theoretical explanation that helps clarify people's understanding of the target variable while an operational definition contains a set of specified features of the target variable that make explicit measurement possible.

## DISCUSSION QUESTIONS

1. What are different types of hypotheses? Find these different types of hypotheses in research papers published in an international journal.
2. What is the difference between a theory and a model? Can you give examples to illustrate such a difference?
3. What is a population and what is a sample? What are the possible relationships between a population and its sample? Please illustrate your explanations by examples.
4. What are independent, dependent, control, moderator and intervening variables? Find a real study from an international journal to illustrate these different types of variables.
5. What are the differences between conceptual and operational definitions? How is a variable operationalized?

6. Identify the conceptual and operational definitions of five variables from a published paper. Share your findings with your classmate.
7. How can a variable be measured differently in terms of levels? Use examples from a study reported in an international journal to illustrate these different levels of measurement.
8. Select one suitable research paper from an international journal and read it carefully from the beginning to the end. Then list all fundamental concepts mentioned in the paper and explain their meanings to your classmate.

# Part II

# Researching

Part II consists of nine chapters which deal with all the important issues occurring at the different stages of research. It starts from the development of research questions and ends up with the analysis of the data gathered. For each stage, the procedures are described and common problems are mentioned together with suggested solutions. By studying this part, you will be able to

- Understand the basic procedures of research.
- Undertake the specified task at each stage
- Use the strategies in coping with the frequently-occurring problems or difficulties.

# 3. Developing research questions

Developing research questions is not an instantaneous process, but takes place over time. During this period, you are likely to do a lot of background reading, discuss with other researchers, and think about the significance, originality and feasibility of the questions. The following section will describe general procedures for the development of research questions and discuss some common problems in formulating research questions.

## PROCEDURES

In developing research questions, you usually undertake the following tasks as shown in Figure 3.1:

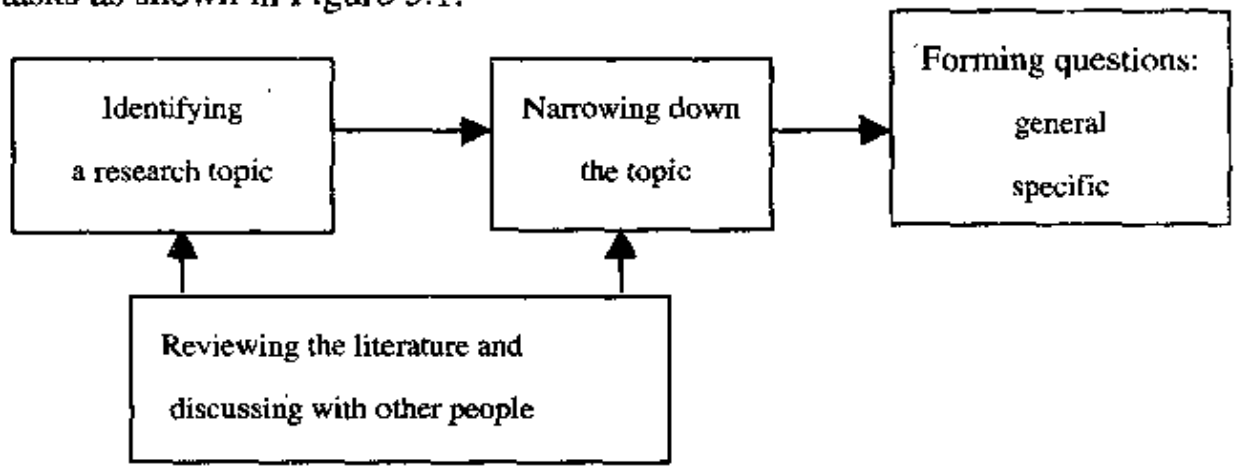



Figure 3.1: Developing research questions

According to Figure 3.1, you start with identifying a general topic followed by narrowing down the topic to get a focus. While you are researching a general topic to find the focus, you need to read relevant literature extensively and discuss with your supervisors or classmates frequently. Finally, you are expected to construct general questions as

well as specific questions. In the following section, I will explain how to accomplish the above four tasks one by one.

## Identifying a research topic

As an initial step, you engage in a literature search for a general research topic. You might decide that you are interested in the area of L2 teaching, or of L2 learning. These two areas are related but each has its own focus and different perspectives. If L2 learning is your starting point, then the topic is extremely general. Suppose your initial chosen area is L2 writing or more specifically, L2 expository writing. Then your topic is obviously less general than the previous one, "L2 learning". Evidently, the journey of developing research questions is varied from person to person when the starting points are different. The more specific the topic you start with, the less heavy the workload for you on the second step, i.e. narrowing down the research topic.

What needs to be considered in topic-selection? If you asked me for suggestions, I would say personal interest should be placed on the top of the list. The reasons are obvious. In the process of postgraduate study, there are more hardships than happiness. You might have to key in thousands of responses to questionnaire items; you might have to transcribe dozens of hours of tapes; you might have searched in the library for several days without finding the materials you wanted. Your personal interest in the topic is always the best driving force for you to reach the end of your arduous journey.

The second thing you need to consider is that you should not select topics which may evoke strong emotional reactions that can lead you astray. For example, some students show a high degree of anxiety and are very much afraid of speaking. Suppose one of them decides to study the relation between the degree of anxiety and oral English proficiency. S/he is likely to have a "position" on the subject that will interfere with her/his completing the research satisfactorily on a number of levels. The most important of these is the self-fulfilling prophecy phenomenon

(Ambady & Rosenthal, 1992). If s/he is convinced that a high degree of anxiety will lead to failure in acquiring good speaking skills, the data collected can be biased. Furthermore, suggestions from supervisors or other people might be hard to accept when they are inconsistent with their personal understanding of the problem. However, according to the constructivist's view, all studies are somewhat biased. What is essential for the researcher is to reduce the degree of bias.

The third suggestion is that you should "avoid a topic which is overly ambitious and overly challenging" (Rudestam & Newton, 1992: 10). Many postgraduate students, before embarking on. their journey of research, have made up their minds to do something unique so thet they can make a remarkable contribution to the field they are investigating. These students, I should say, should be admired but their perception of the thesis is a bit unrealistic. Those apparently spectacular theses may end up with two fates: either remaining unfinished or being completed with lower quality. Therefore, you have to temper your enthusiasm with practical concerns. Remember that even the best thesis is very often a result of compromises among our own ambition, the requirements of the supervisor and practical constraints. What is being said here, however, does not mean that you are encouraged to choose a simple and easy topic, and write a mediocre thesis.

In a word, I would suggest that you should selecte a small piece of an important topic or a small piece of an area that intrigues you.

## Narrowing down the research topic

Narrowing down the topic usually causes researchers, particularly novices, the most trouble. There is no clear-cut rule specifying what is the right size of the topic that is sufficient for an MA thesis or a Ph.D. dissertation. Experienced supervisors can give their students help in this regard.

You can start with the questions beginning with "who", "what", or "how". Suppose you have selected the research topic "the acquisition of

English vocabulary". Obviously, the topic is rather broad. Now you may first ask yourselves "Who are the learners?" Are they middle school students or university students? If they are university students, are they English majors or non-English majors? If they are English majors, are they freshmen, or sophomores, or juniors or seniors? Let's decide to select sophomores. The second question is what kind of vocabulary you are interested in. Are you interested in receptive or productive vocabulary? If your interest is in productive vocabulary, then you have to decide whether you are interested in vocabulary in speaking or in writing. Let's say you are interested in speaking. The last question concerns how you will go about your research. Do you interview the students or ask them to answer a questionnaire, or observe their performance in class? Do you plan to measure the size of their productive vocabulary? Finally, how do you do it? Do you give them each a personal interview or an oral test in a laboratory? By keeping on asking yourselves such questions, the topic will become narrower and narrower and better defined. When will this end? It depends on your research purpose.

## Reviewing the literature and discussing with reseerchers

In the process of research, the literature is reviewed from time to time, but for different purposes. As stated in Figure 3.1, you do a literature review both in researching a general topic and deciding the research focus. If you have not found any topics interesting, reading others' research may give you some inspirations. You may pick up any recent issue of a well-established journal in the relevant field such as *Applied Linguistics*, *TESOL Quarterly*, and *Language Learning*, and read the papers which appeal to your interest. Usually, published papers include recommendations for further research in the last section. You may read the last section more carefully and see whether you find any of the suggested topics attractive. It is also possible for you to be motivated by the findings of a reported study. For example, in Huang's study (1984),

she concluded that functional practice (i.e. undertaking communicative activities) is a powerful predictor of the Chinese learners' proficiency in oral English. You may be motivated to find out whether functional practice is also a good predictor of the Chinese learners' proficiency in reading and writing.

If the research topic is initially motivated by your personal experience or observation, you have to read through the literature to examine whether the selected topic is worth investigating. You will go through major influential journals published in the past decade searching for the papers relevant to your topic focus and reading the abstracts of each paper. If the reviewing of the literature reveals that your chosen focus has not been studied before or that yours differs from the pravious ones in one or two aspects, you are assured that you can move to undertake the next task, i.e. developing general questions into specific ones.

While reviewing the literature, you need to discuss your research topic with other people, espeoially your supervisor. Such a discussion is very often stimulating and enlightening. It can facilitate your research topic identification and focus specification. One problem I have found is that some students do not know how to make full use of human resources around them. They try to work on their own, seldom discussing their research with other people. Therefore, they will miss opportunities to learn from others and gain help from others.

## Formulating general and specific questions

General questions serve as a blueprint that provides the direction for your research but they are not specific enough to be answered. Specific questions that are derived from the general questions are directly related to the details of research procedures such as subject-selection, data-collection and deta-analysis.

The questions slıould not be finalized in a burry since hasty decisions might overlook possibilities. You had better budget enough time to look for all the possibilities available before you reach closure on the specific

questions. Often you may experience a stage where a small set of questions are expanded into a large set which you have to delimit again afterwards. Such experiences are very common and you do not need to worry about them. In Punch's opinion (1998), the absence of such experiences may indicate insufficient time spent generating possibilities in the first place.

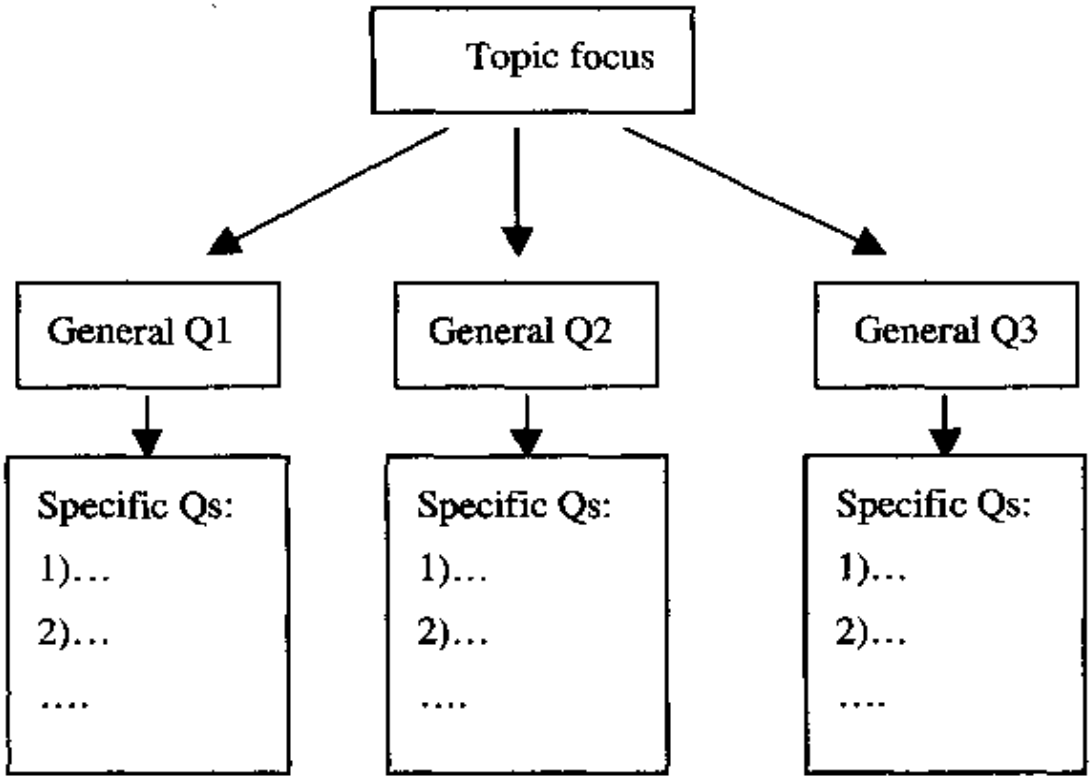



Figure 3.2: Relations between general and specific questions

Once the expansion of the initial set of questions is over, you must examine all the questions to weed out those unimportant ones. The general principle here is that it is better to answer fewer questions thoroughly than many questions superficially. As a study for MA or doctoral programs, you are unlikely to tackle five or more major questions. If you have more than five, you should probably be thinking of cutting them down in number and focusing on fewer.

Research questions should finally be grouped and ordered in a logical sequence. In other words, the hierarchical relationship among the

questions should be self-evident. Such a relationship can be described diagrammatically in Figure 3.2.

Questions for quantitative research are often, if not always, constructed ahead of empirical work. They are sometimes stated as hypotheses. But for most people, straightforward questions are fine. Questions for qualitative research are usually general before data-collection and specific questions only emerge as the research unfolds.

The following are research questions selected from some of our MA students' theses in 1998, quoted here in slightly modified form:

Example One (Wu, 1998:3)

Research topic: L2 argumentative writing

Topic focus: A study of university students' argumentative writing in English: rhetorical knowledge and discourse pattern

General and specific questions:

1. What is the relationship between the students' rhetorical knowledge and discourse patterns?

   (1) Does the subjects' perception of the rhetorical organization of English argumentative discourse vary from the expected English norms?

   (2) Do these English argumentative essays written by the subjects demonstrate any deviations in rhetorical organization from the expected English norms?

   (3) Are there any correlations between the subjects' rhetorical knowledge and the discourse pattern of their English argumentative texts?

2. Does L2 proficiency moderate the relationship between rhetorical knowledge and the discourse pattern?

   (1) Are the subjects of different linguistic proficiency levels different from one another in their perception of the rhetorical organization of English argumentative text?

(2) Do they show any differences in their performance in creating an English argumentative text?

(3) What is the relationship between the students' rhetorical knowledge and their performance in writing if the subjects are divided into groups according to their L2 proficiency?

Example Two

Research topic: EFL writing (Xu, 1998: 14)

Topic focus: Cohesion in English majors' writing

General and specific questions:

1. To what extent does L2 proficiency affect cohesive pattern in EFL writing?
   (1) Does cohesive patterning in EFL writing significantly vary according to students' proficiency level?
   (2) If so, what are the differences in cohesive patterning between learners at a high proficiency and those at a low level?
2. To what extent does L2 proficiency affect cohesive-error patterning?
   (1) Does patterning of cohesive errors in EFL writing vary according to students' proficiency level?
   (2) If so, what are the differences in cohesive-error patterning between learners with a high proficiency and those at a low level?
3. To what extent does cohesive-error patterning affect the quality of writing?
   (1) Do differences in cohesive and cohesive-error patterning between grade levels lead to variation in writing quality?
   (2) If so, what are the differences in cohesive and cohesive-error patterning between good and poor writing?
   (3) Which cohesive devices are the better predictors of the quality of writing?

Example Three (Zhu, 1998: 11-12)

Research Topic: L2 oral performance

Topic focus: Influence of different tasks and lengths of planning time on second language oral performance

General research questions and specific hypotheses:

1. What is the relationship between the familiarity of tasks and L2 oral performance?

   Hypothesis 1:

   The task with which students are most familiar would be performed with the highest degree of accuracy while with the most unfamiliar and difficult task, students might produce the lowest accuracy.

   Hypothesis 2:

   The task which was most unfamiliar and difficult for subjects would generate language with the highest degree of complexity. In contrast, the task which was most familiar to subjects might be performed with the lowest language complexity.

   Hypothesis 3:

   The task which was most familiar to subjects might be performed with the highest degree of fluency while the task that was least familiar to subjects might be performed with the lowest degree of fluency.

2. What is the relationship between lengths of planning time and L2 oral performance?

   Hypothesis 1: With more planning time, greater degrees of accuracy will be observed.

   Hypothesis 2: With more planning time, more complexities in language will be displayed.

   Hypothesis 3: With more planning time, more fluency will be achieved.

## PROBLEMS IN QUESTION FORMATION

As mentioned before, developing good questions is often the first difficulty beginning research students encounter in preparing for their theses. According to my experiences as a superviosr, the following are the common problems they come across.

### Problem 1: No reseerch idea

Some postgraduate students by the time they have finished all their courses have not yet developed any specific research interest. When it is the time for them to write a proposal, they do not know what topic they are going to investigate. They expect their supervisors to choose a research topic for them.

This is obviously a wrong and unrealistic expectation because a supervisor, as the name suggests, is responsible for supervising your work rather than doing the work for you. Moreover, to learn to identify a research topic independently is part of postgraduate study. Once you have developed such an ability, you are capable of pursuing your research after graduation. Otherwise, the accomplishment of your theses would be the end of your research career.

Some supervisors who are lacking experience in supervision might provide the students with suggested topics out of their good will. However, it has more disedvantages than advantages. Apart from not giving the students proper training in developing research questions, sucb practice suffers from at least two additional drawbacks which may not be discerned by either the teacher or the student at the beginning. First of all, a thesis/dissertation normally takes years to finisb. Often there is a period of time when the students feel bored or/and begin to lost interest in the topic. If they did not select the topic themselves, such feelings will be much stronger than otherwise. Some of them even feel regretful that they did not make their own choice of topic. Secondly, such students usually expect the supervisor to fmd a solution to the problem since they

think this is not their topic. Therefore, if the cycle of postgraduate training is to be complete, I strongly suggest that nobody should skip over this process.

How can you find an interesting topic? In addition to the existing literature, the following are two important sources from which you may find one that appeals to you..

### *Personal experience*

One's personal experiences of learning an L2 and being taught an L2 are usually primary and rich sources from which you can find research topics. However, without strong curiosity and high sensitivity, you are not able to get any inspiration from such valuable experiences. But curiosity and sensitivity are not in-born qualities; they have to be fostered. Therefore, I suggest that whenever you have a question about language teaching or learning, you jot it down in a research-ideas journal. These ideas are like sparks, which are transient. If you do not catch them, they will disappear quickly. Only when you get hold of them, can they possibly develop into a full-fledged research project. For example, it is not uncommon for successful English learners in China to confuse "he" with "she" in spoken English. It is not surprising either to find their mistakes in the use of articles. One question you can ask is what are the most difficult forms in the English language for Chinese learners to master. Another question you can pose is what the variations among Chinese learners are in grasping these difficult forms. So long as you are curious enough about your L2 learning experiences, surely you can find innumerable topics to investigate.

### *Attending conferences and talking with other researchers*

The second source from which you can identify a research topic is attending conferences and talking with other researchers. Published papers usually report studies carried out two or three years ago. However, at a conference, presentations usually report recently-finished studies or on-going projects. By attending a conference, you are able not only to

obtain the latest information about the studies in the field concerned but also to meet a group of distinguished scholars as well as active researchers themselves. The discussion section after each presentation is often the most stimulating. You can ask the presenters questions personally and get great help in a very convenient way. You should take every opportunity to attend academic conferences related to your research and try not to miss any guest speaker's lectures in your own institution. Furthermore, you should try to have a conversation with them when they finish their presentation or public speech and ask them what questions they think need further research in their areas of expertise.

## Problem 2: Topic being too broad

Topics selected by researchers with limited experience tend to be broad without a clear focus. This is not a problem at all if it occurs at the initial stage. It is a problem only when the researcher does not know how to deal with such a situation, or worse, when s/he does not know the topic needs further narrowing down. For example, one of our students said that he wanted to investigate to what extent affective factors influence non-English major graduate students' English learning. The question appears to be appealing in the sense that most of the previous studies focused on cognitive factors and only a few of them investigated the effects of one or two affective factors. In this sense, designing a study of the learner's affective factors is desirable. The problem he had is that he was not able to see why the topic was too broad.

From his point of view, he had already tried to narrow the topic down to some extent. For example, he selected affective factors rather than all the factors related to learners, and he only intended to study non-English major graduate students rather than all university students. But he did not think of the fact that there might be as many as a dozen affective factors and it would be difficult if not impossible to cover them all in a single study. Nor did he realize that there are very few valid instruments available for measuring the affective aspects. Eventually, with the help of

his supervisor, he was able to narrow down the topic and focused on one affective factor in his research, i.e. motivation.

## Problem 3: Questions not worth investigating

Some questions raised by researchers may not be worth investigating if the questions are trivial and have neither practical nor theoretical value. You might wonder how this can happen. It is true that no researcher would like to invest time and effort in a meaningless way. However, an inexperienced researcher has difficulty in detecting triviality in the questions put forward. For example, one student tried to see how first-year English majors and third-year English majors differ in summary writing. The general question sounds interesting and seems not to be insignificant at all. Let us look at the specific questions: (1) Do these two groups of students show differences in their overall scores on their summary writing? (2) Do these two groups of students show differences in the number of important points included in their summaries? (3) Do these two groups show any difference in the total number of grammatical errors in their summary writing? All these questions appear to be well constructed with a clear focus. My question is: why she wanted to find out the answers to these questions? Without her empirical data, I am sure, you can provide the answers without too much thinking because the third-year students should get higher scores, include more important points and make fewer grammatical errors than the first-year students. Otherwise, there must be something wrong with our education system. Then what implications could we draw from such findings? Do these findings have any implications for improving our teaching or learning or do these findings have any contribution to theory-building? Absolutely not. Therefore, the specific questions mentioned above are trivial and are not worth investigating at all.

From the above example, you may notice that by saying questions are trivial, I mean that the answers to questions do not have practical value or

theoretical value even if the process of research is very complicated or laboratoryor-intensive.

## Problem 4: Questions like a ehopping list

It is very common for beginning researchers to form a set of questions like a shopping list. By saying so, I mean that the questions do not display a strong logical link and there is no distinction between general questions and specific ones. The reason for such a problem is mainly due to insufficient thinking on the part of the researcher. For these researchers, the process of question-formation is somewhat like shooting bullets in the sense that questions are formed one after another without considering the internal link at all. Obviously, these questions are not matured enough to serve as the final set, which need further time and effort to work at.

If you are in this situation, what should you do? I suggest that you list general questions first and then establish a link among them based on your thorough understanding of your topic. Such a link is extremely important since it enables your study to be an organic whole. Once general questions are internally connected, you move to the stage of developing specific questions which is comparatively easier than the previous one. Let's look at the following example which can well illustrate what kind of questions I refer to as a shopping list and what questions have a hierarchical structure. The first set of questions is the one which was formed at the initial stage of research. It does not display a strong logical link. The second set is the final one that resulted from a substantial revision.

The first set in the early draft of one student's thesis:

1. Is there any use of first language in the second language composing process of university students? If yes, how much L1 is used in L2 writing process?
2. When does first language occur in second language composing process?

3. What functions does first language have in second language writing process?
4. Why do writers use their native language in the process of composing in second language?
5. Does L1 use in L2 writing vary with the writers' L2 proficiency development?
6. Does L1 use in L2 writing vary with the type of writing task?
7. Is there any correlation between L1 use in L2 composing process and the quality of the composition?
8. What do L2 writers think about the use of their native language in second language writing?

The second set is the final version:

1. How is L1 used by tertiary-level English majors in China in the second language composing process?
   1) How much is L1 used in the L2 composing process?
   2) When is L1 used in the L2 composing process?
2. Does L1 use vary with such factors as types of writing tasks and the development of L2 proficiency?
   1) Does L1 use in the L2 writing process vary with the type of writing task?
   2) Does L1 use in the L2 writing process vary with the writers' L2 development?
3. What roles does L1 play in the second language writing?
   1) What function does L1 have in the second language writing process?
   2) Is there any correlation between L1 use in the L2 composing process and the quality of the composition?
4. What do second language learners think about the use of first language in the second language writing and learning in general?
   1) Why do writers use their native language in the process of compesing in a second language?
   2) What attitudes do L2 writers hold towards the use of their uative language in the second language writing process?

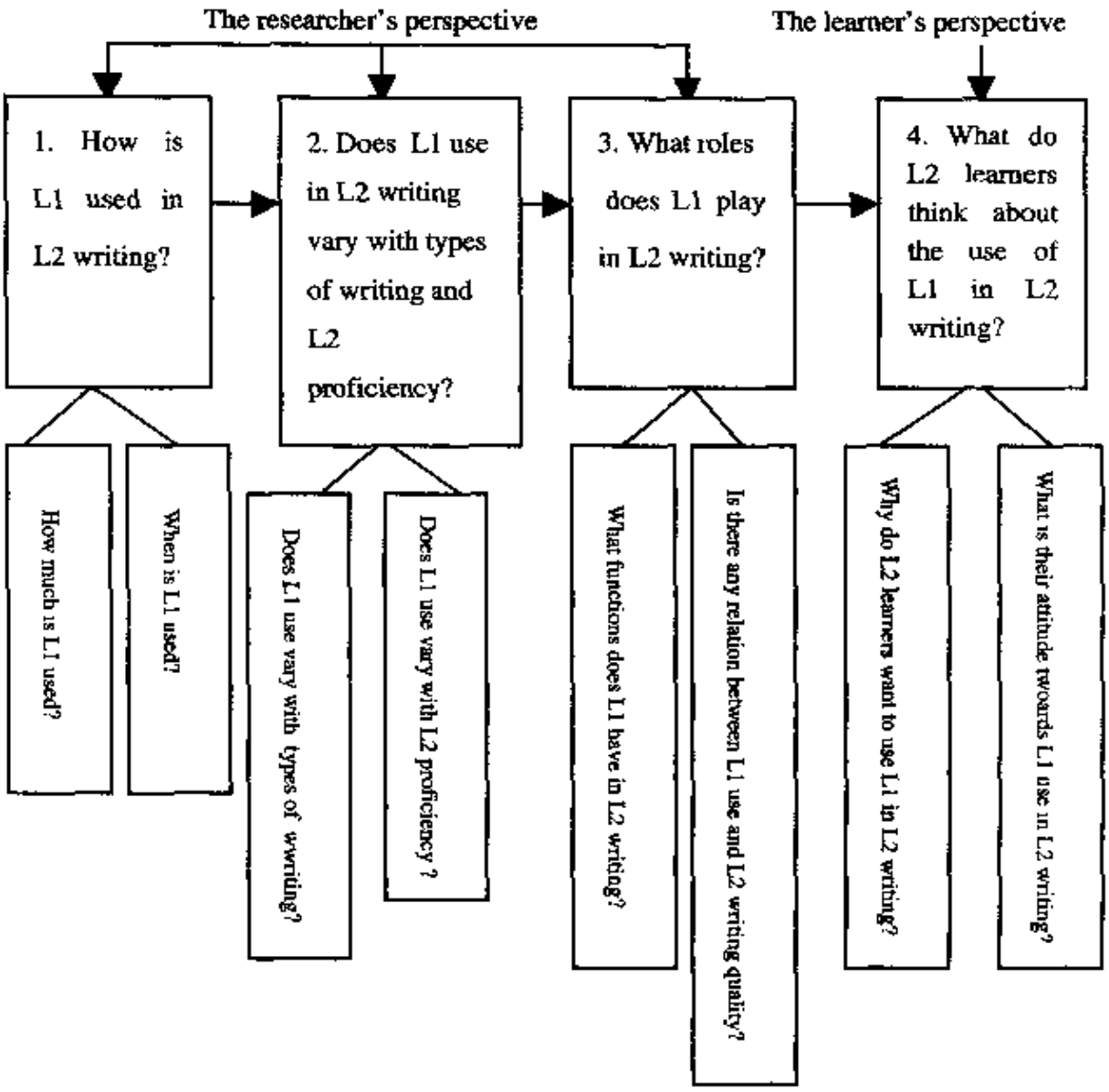



Figure 3: A graphic description of the second set of questions

In the first set, the researcher addressed eight different issues concerning the use of L1: how much, when, what functions, why, the effect of L2 proficiency on L1 use, the effect of types of writing task on L1 use, the relation between L1 use and L2 writing quality and L2 writers' belief about L1 use. These eight issues are arranged in such a way that it is difficult for the reader to figure out the internal link among them. By contrast, the second set of questions are reorganized into four general questions and each

general question contains two specific questions. The first question is to obtain an overall picture of the use of L1 in L2 writing, the second question to find out to what extent the use of L1 is affected by other factors such as types of writing task and L2 proficiency, the third question to identify specific functions of L1 in L2 writing and the relation between L1 use and L2 writing quality, and the last question to examine L2 learners' view about the use of L1 in L2 writing. The link among the four general questions has thus surfaced (See Figure 3.3).

## SUMMARY

Developing research questions starts by identifying a research topic that needs narrowing down to obtain the focus of research. The strategy that can help you narrow down the topic is to keep on asking questions beginning with "who", "what" and "how". Reviewing the literature while discussing with other people is necessary in both obtaining a research topic and determining the focus. Once the focus of research is specified, general questions are expected to develop together with a sub-set of more specific ones. Would-be researchers are very often troubled by having no research ideas at the initial stage. They are advised to reflect on their own learning experiences, and attend conferences or guest lectures to gain inspiration. They may also be troubled by the topic being too broad or the questions being trivial. The last problem they often encounter is that their questions are presented as a shopping list without a hierarchical structure.

## DISCUSSION QUESTIONS

1. What are the general procedures by which questions are developed?
2. When you identify a research topic, what should you consider first?
3. What strategies can you use in narrowing down your topic to get a focus?

4. What kind of contribution can the literature review make to the development of questions?
5. Why do we need to formulate general as well as specific questions?
6. Please try to narrow down the following topics:
   a) The learning of English articles
   b) The mastery of modal verbs by Chinese learners
   c) The teaching of English tense system
   d) The teaching of L2 argumentative writing
   e) The relationship between the use of L1 and L2 learning
7. Classify the questions in each of the groups given to show the logical relationship between general and specific questions. (Note: these questions are from students' theses with some modifications).
   a. A study of the risk-taking beliefs and behaviors of English majors in China
      (1) What do Chinese students think of risk-taking in English learning?
      (2) What is their average rate of risk-taking?
      (3) Is there any difference between their beliefs and behaviors?
      (4) What do high, moderate and low risk-takers think about risk-taking respectively?
      (5) What are the respective behaviors of high, moderate and low risk-takers in learning English?
      (6) What are the differences in the beliefs and behaviors of high, moderate and low risk-takers?
      (7) Is there any correlation between risk-taking and L2 learning outcomes?
      (Wang, 1999)
   b) A study of error correction in university EFL classrooms
      (1) How do intensive reading teachers correct their students' errors?
      (2) Is there any difference between English-major teachers and non-English-major teachers in error correction?

(3) What are the students' attitudes towards error correction?

(4) Is there any difference between English majors and non-English majors in their attitudes and preferences?

(5) As far as error treatment is concerned, to what extent can the teachers' behavior match the students' expectations?

(Hu, 1999)

# 4. Reading the literature

In the previous Chapter "developing research questions", you were asked to read the literature for drawing inspiration and checking whether your chosen topic is worth investigating. In this chapter, you can find out how to review the literature for other purposes as well.

## SOURCES OF LITERATURE

At the most general level, materials for the literature review can be roughly divided into two groups: one is within the field of applied linguistics and the other is in the related fields such as linguistics, sociolinguistics, psycholinguistics, education, psychology and sociology. For a thorough and extensive literature review, the sources of literature should not be confined to the field of applied linguistics since the research on second language teaching and learning often borrows ideas and research methods from related fields. The question as to how much you should review within your own field and others will be discussed in the following section, "DECIDING THE SCOPE".

In each field, there are various kinds of sources for locating references, such as indices, conference papers, journal articles and books. In the following section, I will introduce these different types of resources one by one.

### Indices

Indices or bibliography lists are published quarterly, monthly, annually, or bi-annually. They are a good source for activating your literature review. Table 4.1 lists some examples which are most commonly used by applied linguists.

The indices offer readers a large number of references on a variety of topics, which are presented according to an author's name and subject.

They provide various types of useful information, for example, the year of publication, the place of publication and the name of the publishing house, where and how it can be obtained, and the form in which it was published (paper, article, or report). Most indices also contain abstracts or summaries of the material concerned.

| Name of index | How often was it published | Where was it published | Information provided |
|---|---|---|---|
| Language Teaching | Quarterly | London: Cambridge University Press | References to journal articles, books and reports |
| Resources in Education | Monthly | Washington, DC: Government Printing Office | References to articles, abstracts of reports, conference papers, and other studies |
| Current Index to Journals in Education | Monthly | Phonenix, AZ: Oryx Press | References to articles related to education appearing in about 780 journals |

Table 4.1: Examples of indices

## Unpublished papers

Unpublished papers[3] refer to conference papers or manuscripts circulated among colleagues for discussion. They are valuable sources

[3] Unpublished papers should be listed in your references just like published papers; only instead of publisher, specify "manuscript" and where and when it is circulated (e.g., course handout, which course; conference paper, which conference; Master thesis, which university). Same goes for papers published on the web; then the website URL should be given. Otherwise you will be regarded as plagiarized

for finding out about ongoing or recently completed work. Very often in such papers, in addition to the titles of papers and abstracts, you will find the names of presenters and their mailing addresses as well as email addresses. Thus you can easily contact the author if you want any additional help from him/her. Another advantage of the unpublished materials is that they can provide you with up-to-date references while published materials are usually a report on a project finisbed one-to-three years ago.

## Journals

Journals (or periodicals) contain a collection of papers which are published regularly. Different journals target at different readers. In second language acquisition, some journals are primarily for researchers; some more oriented towards practitioners; some for both types of audience.

The following list of journals covers major ones in the field of applied linguistics: *Applied Psycholinguistics*, *Applied Linguistics*, *Annual Review of Applied Linguistics*, *Canadian Modern Language Review*, *ELT Journal*, *English Teaching Forum*, *Foreign Language Annals*, *Interlanguage Studies Bulletin*, *Language Testing*, *Language Acquisition*, *Language Learning*, *Modern Language Journal*, *Second Language Acquisition Research*, *Studies in Second Language Acquisition*, *System*, *TESOL Quarterly*, *English for Specific Purposes*, *TESOL Journal*, *English Today*, *Asian Journal of English Language Teaching*, *World Englishes*, *Pragmatcics*, *The Journal of Pragmatics*, *Written Communication* and *Language in Society*.

Among the journals listed above, for example, "Applied Linguistics" and "Language Acquisition" are typically oriented towards researchers; "English Teaching Forum" and "ELT Journal", primarily for practitioners; "System" and "TESOL Quarterly", for both types of readers. Generally speaking, the journals oriented towards researchers are more valuable for your literature review than those targeting practitioners.

However, if you read papers for practitioners with a critical eye, you can get a sense for the current problems in teaching methodology which may be quite relevant to focusing a research question.

## Books

There are four kinds of books as a source for locating references: 1) textbooks, 2) dictionaries and encyclopedias of linguistics, 3) research monogrphs, 4) an edited collection of papers.

Textbooks provide a good introductory survey and explain concepts more systematically than research papers do. They provide the kind of overview of a topic area which research students lack and which is invariably assumed by writers in journals. Furthermore, papers in journals never have the space to give background.

Reference sources like encyclopedias and dictionaries of linguistics are particularly good for the defnitions of key concepts and the origins of the research into these concepts. Working with recent literature, you might get the most recent controversy or defnition but not the original source. The most useful encyclopedias and dictionaries for you, for example, are David Crystal's Encyclopedia of the English Language, Tom McArthur's Oxford Companion to the English Language and Jack Richards, Platt & Platt's Longman Dictionary of Applied Linguistics.

Research monographs give a single author's (or joint authors') theory or viewpoint about a topic, and can be quite important for the development of the field, e.g. Krashen's *Language Two*.

An edited collection of papers arises either out of a conference, or by invitation from the editor, or out of the need for collaborative research on an interdisciplinary topic. They pool together the expertise of many people to shed light on the central topic or topics. They have a clear focus and deal with topics from a particular perspective.

Compared with indices, conference papers and journals, books as a source for locating references may have their own weakness. For example, the references in a book are usually not recent ones since it takes

much longer to write a book than a paper. Remember that books give you the background and breath while journals bring you what is new.

## PROCEDURES FOR REVIEWING THE LITERTURE

Beginning researchers often have the experience of feeling overwhelmed once they enter the library, because of the vast amount of materials surrounding them.

In this section, I will recommend to you a set of procedures which can help you get done the job of reviewing the literature effectively (See Figure 4.1). However, this is not the only possible approach.

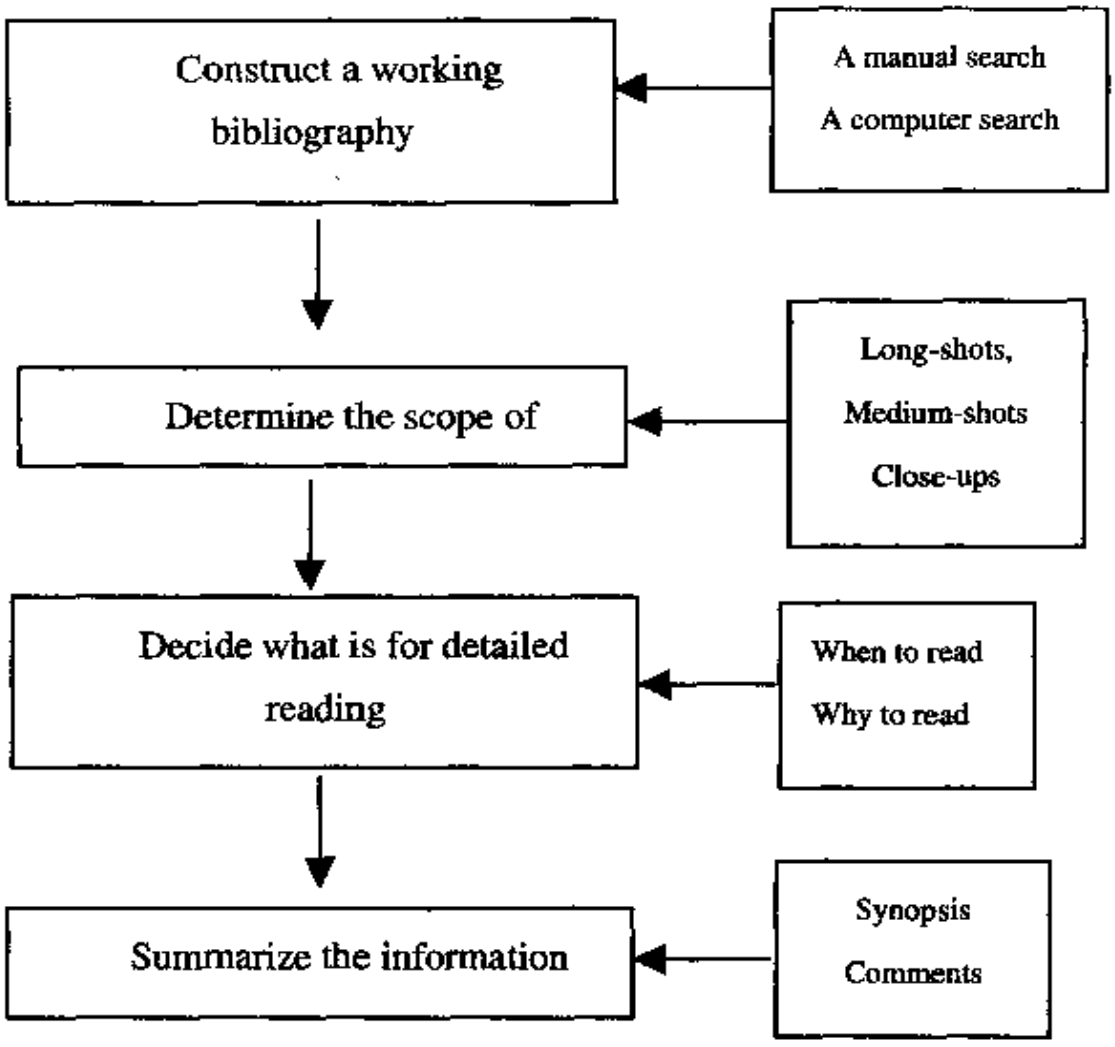



Figure 4.1: Procedures for reviewing the literature

## Constructing a working bibliography

### *What is a working bibliography?*

A working bibliography is a tentative list of references constructed by the researcher for the preparation of reviewing the literature. It may serve two purposes. First, it can be used as a blueprint to guide your review of the literature. Secondly, it can be taken as a resource bank from which you construct the section of references for your thesis or dissertation in the end. Normally, you keep revising your thesis or dissertation. As a result of revision, the materials you refer to are changing. You might refer to a paper today, cross it out tomorrow and add it again the day after tomorrow. If you change the section of references after each revision, that will be very troublesome. The more economic way is to produce a reference list once you finish all the revision of the thesis by selecting the cited works from your working bibliography.

The references in a working bibliography should be arranged alphabetically and each entry should contain the name(s) of the author(s), the title of the report, facts about publication, page numbers, and so on. If the reference is to a journal article, it should include the name of the journal, volume, and inclusive page numbers. If it is to a book, it should contain information about the publisher, and the place and date of publication. Each entry should be double checked to make sure all the information in the entry is accurate. Tolerating a mistake at this stage will lead to a huge trouble afterwards. It might take you an enormous time to retrieve the page number of a reference paper, or the correct spelling of an author.

How do you locate the resources to construct such a list? Primarily, there are two kinds of search through which you can locate them: (a) a manual search and (b) a computer search. A computer search is obviously more advanced and more efficient than a manual one. However,libraries in some universities are not computerized, or the databases for applied linguistics and its related fields are not available. In

this situation, we have to search the sources manually without other choices. Even in those universities where a computerized system is available, a manual search cannot be totally avoided. Thus in the following section, these two kinds of search will be described in turn.

### *A manual search*

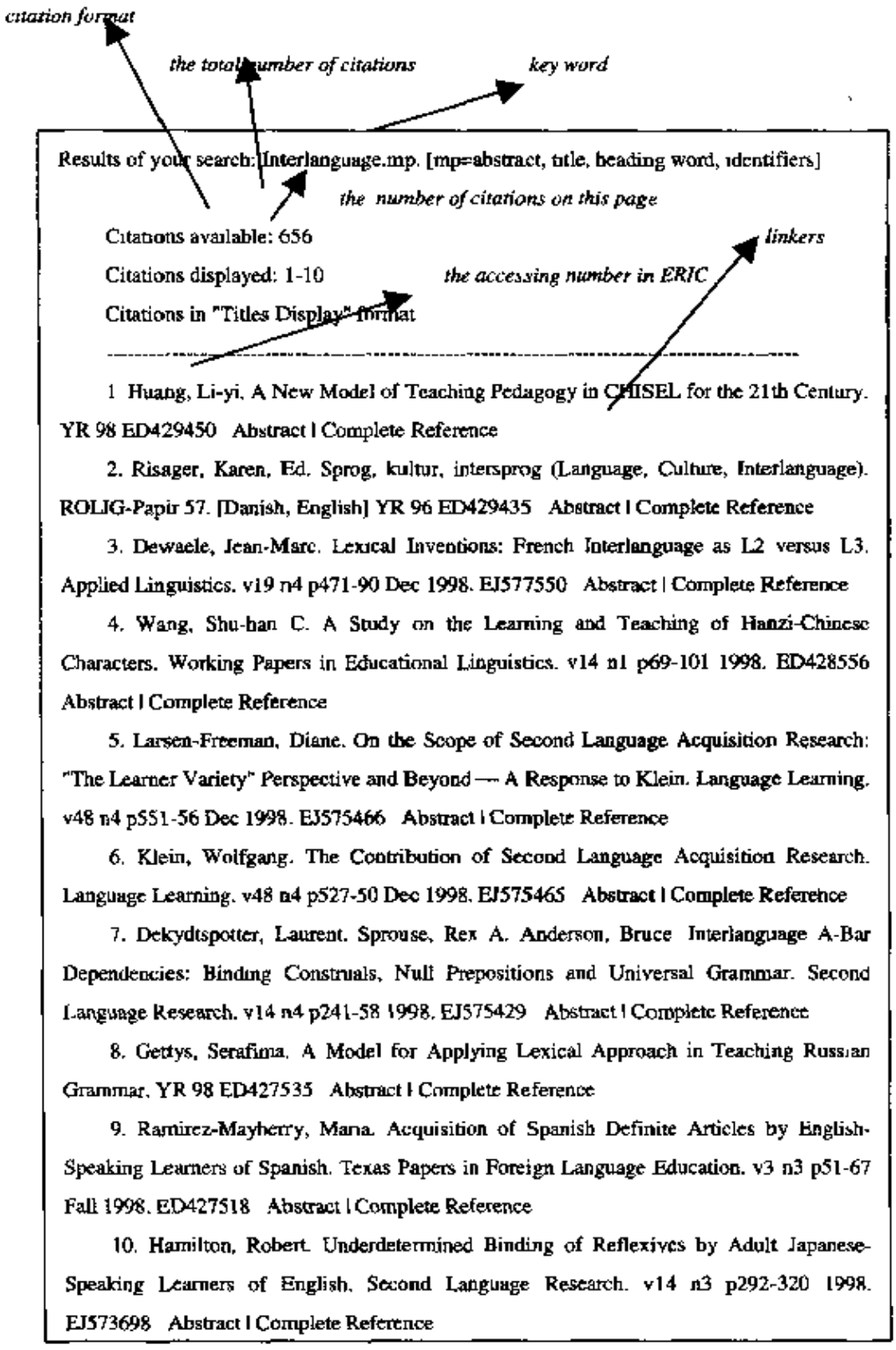
In a manual search, the researcher has to locate the materials physically without the help of computers. From my own experiences and those of our graduate students, we feel that a snow-balling method works rather effectively in this case, i.e. you start with a few references each of which can generate a few more and by repeating the process, thus your reference list grows. To be specific, you may begin with the references in one paper that you think is most closely related to the topic. By screening through the whole list of references, you weed out those that appear to be irrelevant. Through such a screening, if you are lucky enough, you may have 10 to 15 references left, among which you can find 5 to 10 in your library. At the second stage, you read the abstracts of these 5 to 10 papers to decide whether they are related to your topic or not. Suppose among them only five papers are valuable. As a resnlt, these five are kept and the other five are removed from the list. Then you repeat the procedures of the first stage to go through each of these five papers: screening through the reference list at the end of each paper to get rid of the irrelevant ones and searching those apparently relevant references. Gradually, the number of references grows exponetially till the list can well form a working bibliography.

Remember at this stage, you only need to read the abstract of each paper rather than the paper in full. Reading the full papers will be too time-consuming and will get you bogged down in details, whereas comparing the abstracts allows you to see a bigger picture of the range of references which are relevant to your topic.

### *A computer search*

In a computer search, we can easily and efficiently construct a working bibliography based on the information provided by computer databases. The most important database in second language acquisition is provided by ERIC (Education Resources Information Center) within the US Department of Education. It contains over 700,000 references and abstracts; its indexes cover 800 journals and 13,000 documents (including selected books) annually.

A list of relevant references can be generated from the database once you input the key word. Figure 4.2 is what was immediately shown on the screen from ERIC once you key in "interlanguage". It consists of two parts. The first part reports general information about the search, i.e. altogether there are 656 citations which are in "titles display" format and the first page only contains the first 10 citation. The second part is a list of citations each of which provides information about the author, the title, where it is published, in what language the material is written and the assessing number in ERIC, etc. At the end of each citation there are two linkers, i.e. Abstract and Complete Reference each of which can be accessed to by double clicking the mouse.




(Note: The italicized words are explanatory notes made by the author.)

Figure 4.2: Results generated by ERIC as an exemplar

Different from a manual search, a computer search can generate a list of references in a few seconds which, however, is not what we call a working bibliography. Such an initial list of references has to be screened through at least three times to get rid of those irrelevant ones. In the first screening, you simply delete the citations which are obviously not useful. The second screening can be made after you read the abstract concerned by activating the linker "Abstract". Finally, you have to search the computerized library system to see whether the materials are available or not. The references that can survive the three times' screening form a working bibliography.

A working bibliography can be produced simply by sitting in front of computer and clicking the mouse. In this sense the computer search can help save time and energy. Furthermore, the resources in the database are abundant and up-dated. Thus a working bibliography built up through a computer search is more comprehensive than one yielded by a manual search. However, the problem with ERIC and other computer-based sources is too much quantity but too little quality. You need to spend enormous time winnowing out irrelevant literature.

## DECIDING THE SCOPE

Once a working bibliography has been constructed, you have to decide the scope of reading, i.e. drawing a boundary for reviewing. If the scope of reviewing is too broad, you may be overwhelmed by too much reading and thus lose the right perspective on the research; if it is too narrow, you may overlook studies which contain important and relevant information. Therefore, the task of delimiting the scope of reviewing cannot be skipped over. In this section, I will provide you with some suggestions on how to set up such a tentative boundary.

### *Long-shots, medium-shots and close-ups*

To draw a boundary is to decide what should be included in the review and what should not. As we know, when drawing a picture of a

landscape, instead of concentrating on the details at the beginning, the painter sketches a draft in which he/she decides what is placed as a foreground, what is set as a background and what is in between. Similarly, for the delimiting of the boundary of a review, the researcher does not need to pinpoint the specific function of each reference or to give a detailed account on every paper or book. What is required instead is to decide which reference should be used in which type of review: a background review, a foreground review or something between the two categories.

For the sake of having separate terms for each type of review, I would like to borrow from film-making the terms "long shots", "medium shots" and "close-ups" (Rudestam & Newton, 1992). These three kinds of shots in film-production differ in the distance between the camera and the subject matter. Now their metaphorical difference in literature review lies in their degree of detail.

A long shot or background review usually provides your project with a general framework or theoretical definitions of certain variables; a close-up or foreground review focuses on the details of some specific studies on which the design of your study is based; a medium-shot gives a description of empirical studies on the topic in general. Nevertheless, the differences among these three kinds of reviewing are relative. They can only be understood with reference to concrete cases. In other words, without the context, it is difficult to clarify what is a long shot, a close-up or a medium-shot.

## *Examples*

The following are examples which can serve to illustrate the differences among the three types of reviewing.

Example One

A study of the relationship between second-year English-majors' risk-taking and L2 performance in the Chinese context

Long-shots: Materials in the field of education concerning the relationship between risk-taking behaviors and academic achievement in general;

Medium-shots: Materials in the field of second language acquisition that address the issue of the relationship between risk-taking behaviors and L2 performance both theoretically and empirically;

Close-ups: Empirical studies on this issue in and outside the Chinese context pertinent to the proposed study.

Example Two

A study of the effects of modifiable learner factors on L2 achievement of second-year English majors in the Chinese context

Long-shots: Theories in the field of education about factors affecting learning outcomes,

Medium-shots: Theoretical models on this issue in the field of second language acquisition; materials about English teaching and learning in the Chinese context

Close-ups: Empirical studies on this issue in and outside China

Example Three

How is L1 involved in the process of L2 composing?

Long-shots: Theoretical claims on the use of L1 in L2 learning;

Medium-shots: Empirical studies on the same issue;

Close-ups: Empirical studies confined to L2 writing with a focus on the use of L1.

According to the above three examples, it is clear that "long-shots" are usually related to the materials that are theory-driven; "close-ups" are

only concerned with empirical studies that are most relevant to your own design; "medium-shots", can be theory-driven or data-driven. We may arrange them along the two continua: a) Abstract ⇒ concrete; b) General ⇒ specific.

## Decide what is for the detailed reading

Before you embark on the detailed reading of the selected materials, you should know clearly which aspect of the paper should be focused on in reviewing. The answer is not straightforward since the content for the reading depends on when to read and why to read as shown in Figure 4.1.

Literature review conducted at different stages of research serves different purposes. At the very beginning, you read the research papers to identify a research area, or to find out whether the topic is worth researching. In other words, this kind of literature reviewing is to define the topic and the need for it. Once you have identified the research topic, the review of literature will help you to form appropriate research questions and design your study effectively. Finally, you review the literature to discuss implications of the findings. As I pointed out above, literature reviewing occurs frequently throughout the process of research. Usually the researcher goes to the library at least once a month to read the newly published journals and update the working bibliography.

## SUMMARIZING THE INFORMATION

When you have selected sufficient relevant materials, you will start reading them. I suggest that you read backwards, beginning with the most recent one and moving back to the old ones. What should you do during and after reading? You are supposed to synthesize the information from reading and record it on cards. It is best to prepare three sets of cards: (a) cards for abstracts; (b) cards for critical comments and (c) cards for comments concerning the potential use of the materials. Having separate cards for three kinds of information might appear to be a waste but such a filing system can facilitate retrieving afterwards. Its advantage

can be recognized when you are writing up your thesis. Finally, these three sets of cards need to be cross-referenced so that you can easily search the information you want to retrieve.

## Cards for abstracts

An abstract should include the most important information in a paper. The cards are usually arranged in alphabetical order and/or according to different subheadings related to the research topic. Each abstract is entered on one card or preferably in the database of a computer, and should contain the following information:

(1) The reasons for the proposed study;
(2) The underlying assumptions of the study;
(3) The procedures used for collecting the data;
(4) The subjects and instruments involved in the study;
(5) The procedures for analyzing the data;
(6) The major findings of the research;
(7) The specific new contribution the study made to the existing literature.

In most cases, the papers themselves include abstracts, which however are not written to serve your purpose and thus need a lot of modification. If a paper does not have an abstract, you have to write it. To save time, such an abstract need not be a coherent summary, but in note-form. In other words, it may contain incomplete sentences. The basic requirement for your notes is that the information recorded should be adequate for your subsequent literature review.

To write an abstract is not that difficult but it is time-consuming. Some graduate students tend to be conscientious at the beginning but grow impatient and lazy gradually since they feel their memories are good enough to remember the information and it is not worth spending so much time writing. However, they will soon realize their memories are not so

reliable, and they have to go back to the piles of papers and read them again. In order to avoid such subsequent troubles, you had better form a good habit, i.e. write an abstract as soon as you finisb the reading of each paper.

## Cards for critical comments

Critical comments are formed through the evaluation of a study. You write down strengths and weaknesses of the study, especially the flaws you have identified. You are advised not to wait to write them down until you finish the reading. Such comments may occur in your mind while reading, and are often instantaneous and transient. Note down the ideas before they disappear and organize them on a card once the reading is over.

To write evaluative comments is not as easy as writing a descriptive abstract since you have to examine reading materials with a critical mind. You might find flaws in subject-selection, or in instrument-design, or in data-collection, or in data-analysis, or in the interpretation of the findings. A few flaws are self-evident but many more are covert. The flaws can be easily found by relating the study you are reading to other similar studies you are familiar with. Your appraising ability can be improved if you strive to be critical all the time. To be critical is not necessarily to be cynical. Apart from searching for weaknesses, you are also sensitive to others' strengths and unique features which might be mentioned in your review.

## Cards for comments concerning the potential use

When you read a paper, you are interested in its important information, its strengths as well as its weaknesses. More important than this is that you are interested in to what extent the paper you read can be related to your own study. For this reason, we have a third kind of card, i.e. the card for comments concerning the potential use of the material. These comments indicate to you how the material can be used at the stage

of designing a study and at the stage of thesis writing. For example, you might write: this part can be referred to in developing instruments, or this point can be used in describing the conceptual framework of your study.

Such comments do not necessarily form a coherent paragraph, or even a complete sentence. Notes can serve the purpose well so long as they make sense to you and are easy to put in a file.

According to my own experience, we need to read most relevant papers at least twice. For the first reading, you produce those different cards mentioned. The second reading occurs when you are writing up your thesis/dissertation.

## SUMMARY

Reading literature is a task frequently undertaken by a researcher through the whole process of research. There are various kinds of literature available in the field of applied linguistics and in related fields, such as indices, unpublished papers, journals and books. We may locate reading materials either by a manual search or a computer search. To begin your literature review, you construct a working bibliography, followed by delimiting the boundary of reading. The next task is to decide what to choose for detailed reading. Finally, you are expected to produce three sets of cards by synthesizing the information from reading: cards for abstracts, critical comments and comments for potential use.

## DISCUSSION QUESTIONS

1. Make a list of journals that can be useful for your literature review according to the resources available in our library.
2. Use a snow-balling method to produce a tentative working bibliography for a study on the risk-taking beliefs and behaviors of English majors in China.

3. Please choose two papers from an influential journal to see what is reviewed as background, what as foreground and what as something in between.
4. Produce the three kinds of cards while reading one self-selected paper of your own choice.

# 5. Selecting research designs

Once you formulate your research questions, you will move to the next stage, i.e. selecting designs. Generally speaking, there are two general opposing camps in the fields of social sciences that include applied linguistics, i.e. quantitative and qualitative. In this chapter, I will first present to you a general picture of quantitative and qualitative designs, then discuss the link between research questions and designs, and finally some important issues involved in design-selection.

## QUANTITATIVE AND QUALITATIVE

The general description in this section will start with definitions of quantitative and qualitative designs followed by their general differences, and then move to the historical development of the two designs. The section will end with the advantages of using the two designs in combination.

### Definitions

It is difficult to define quantitative and qualitative research in one or two sentences since they do have a series of differences. To make the whole issue simpler at the beginning, let's first look at Punch's definitions (1998:4):

> Quantitative research is empirical research where the data are in the form of numbers.
> Qualitative research is empirical research where the data are **not** in the form of numbers.

According to Punch's definitions, both quantitative and qualitative research are primary research but they differ in the form of data.

Quantitative data are numeric while qualitative are not. Thus, the best and simplest way to differentiate these two kinds of research is to check whether the data collected are presented in numbers or not in numbers. Other researchers such as Blaxter, Hughes & Tight (1996) and Tashakkori and Teddlie (1998) also suggest the numeric-narrative[4] contrast to capture their essential difference although they admit that it is oversimplified to some extent.

Let's now look at some examples of numeric data:

Example One

You wanted to find out your students' vocabulary size and thus you designed a test. The test consisted of 100 words randomly selected from the syllabus, and the students were asked to give a Chinese equivalent to each English word given. If an answer was correct, one point was obtained. Once the test paper was corrected, each student got a test score representing the vocabulary size. Here the test scores are numeric in nature and are thus quantitative data.

Example Two

Suppose you intended to find out the strategies used by students in increasing their vocabulary. You constructed a questionnaire that contained 20 statements about the strategies, such as "I pay attention to the new words used by my teachers and classmates" and "I try at every opportunity to use the words I have just learned". Each statement was followed by three choices from which the students were only allowed to choose one. These three choices form a three-point scale: Often - Sometimes - Rarely. Once the students' questionnaires were collected, you converted

[4] Narrative means describing something in words.

their responses into numbers: "1" being used for "rarely", "2" for "sometimes" and "3" for "often". Eventually, all the selected responses are changed into numbers. Hence, the use of strategies is converted into quantitative data.

In contrast to quantitative data, qualitative data are narrative in most cases, if not in all. For example, to investigate vocabulary expansion strategies, instead of using a questionnaire, you may interview 20 students and ask them how they increase their vocabulary. They answer this open-ended question verbally and their answers are bound to be varied. Such diversified verbal answers are hence narrative data.

Apart from the critical difference in the form of data, the two types of research design have many more important differences which affect various phases of a research process. In the following section, these important differences will be described.

## More differences

Table 5.1 adapted from Malhotra (1993:159) summarizes the differences between quantitative and qualitative designs in addition to the numeric-narrative contrast.

According to Malhotra, quantitative and qualitative designs are used to achieve different objectives. For the former, its goal is to examine variables that have already been established in the literature and to see to what extent hypotheses regarding these variables can be confirmed or refuted; whereas for the latter, the aim is to identify variables for further research and to formulate hypotheses for testing in the future. In other words, quantitative research is more powerful when the research is to validate the findings in previous studies; qualitative research is more useful when the research is exploratory in nature.

| | Quantitative | Qualitative |
|---|---|---|
| Objective | Examining variables identified; Testing hypotheses | Identifying variables; generating hypotheses |
| Questions | Specified before data-collection | Gradually specified in the process |
| Sample | A large sample | A small sample |
| Data-collection | Structured | Less-structured |
| Data-analysis | Statistical | Non-statistical and statistical |
| Outcome | Generalizable | Ungeneralizable |

Table 5.1: Differences between qualitative and quantitative designs

Questions for the quantitative design are prespecified in the sense they are constructed ahead of empirical study. The prespecified questions help decide the procedures for subject-selection, data-collection and data-analysis in advance. The researcher thus has fewer uncertainties and ambiguities once the research starts. In contrast, questions for the qualitative design are rather general before data-collection and become focused gradually along with the progress of the research. Consequently, it is quite common in qualitative research that the focus of data-collection shifts and the way of data-analysis changes over time.

Quantitative research often needs a large sample to satisfy statistical requirements when it tests hypotheses. By contrast, qualitative research usually involves a small sample because an in-depth study requires enormous time and energy.

For data-collection, qnantitative data are usually collected through more structured procedures than qualitative data. By contrast, qualitative research procedures for data-collection are more flexible and dynamic.

Quantitative data can only be analyzed statistically while qualitative data can be analyzed both non-statistically and statistically although non-statistical analysis is more common for qualitative data. Since there are

two options available for qualitative data-analysis, the researcher has to decide when and how to use them to best serve the purpose.

Due to the exploratory nature of qualitative research that involves a few cases, the outcomes are less generalizable than the findings yielded by quantitative research. After all, qualitative research in the first place is not aimed at establishing universal laws that can be applied beyond the cases in question.

## Historical development of the two designs

The debate on the quantitative-qualitative issue has been on-going. Reviewing the history, we may roughly identify two stages: (1) the stage of mono-design and (2) the stage of mixed designs (Tashakkori & Teddlie, 1998).

### *The mono-design stage*

From the 19th century to the middle of the 20th century, either quantitative or qualitative design was employed in a single study although some researchers tried to collect data by different techniques within either quantitative or qualitative design. Thus we designate this period as the era of mono-design.

Before the 50s, quantitative design was in a dominant position while qualitative one was peripheral. The popularity of the quantitative design was primarily due to its remarkable success in natural sciences in the 20th century. Social sciences, in order to justify their status as scientific, followed the natural sciences. They tried to quantify what they studied and establish the relation or pattern between two or more variables. Quantitative findings were believed to be more powerful and convincing than verbal accounts since it was widely assumed that without numbers, the scientific nature of a study would be called into questions.

However, in the 60s, quantitative design started to be challenged and qualitative design gradually moved into the mainstream of social science research. Many social scientists argued that social reality is different

from natural reality and thus transplanting the design for the natural sciences into social science is not appropriate. Obviously, human beings and the societies in which they live cannot be studied in the same way as natural objects such as rocks and trees. Furthermore, natural scientists can study physical objects in a detached way but social scientists who are part of a society can not be totally detached from the society even if they want to. Starting in the 60s, the debate on the quantitative-qualitative issue became more and more intense. Supporters of the two designs were often intolerant and inflexible. Each side criticcized the other harshly and strove to argue for its own superiority. Naturally, studies conducted by the researchers from the two different camps were either quantitative or qualitative.

### *The mixed design stage*

Since the 80s, many social scientists, having realized that such a debate on the superiority of designs is unproductive, have tried their best to make peace between the two opposing camps. They advocate that the world is so complex that both quantitative and qualitative designs are needed if the eventual purpose of research is to have a whole picture of the world. In their opinion, these two desigus each have strengths as well as weaknesses. They are complementary to each other rather than in conflict. A quantitative design can be used in a large-scale study which can provide a succinct and parsimonious pattern. The results are generalizable, but are often oversimplified and show poor ecological validity[5]. A qualitative design can produce a more realistic picture of reality and reveal more complexities, but it is time-consuming and the results are not generalizable (Cohen & Manion, 1991; Keeves & Sowden,

---

[5] Ecological validity refers to what extent the findings can be applied to real life. Very often the quantitative study, in order to strictly control the factors not investigated in the study, is conducted in a laboratory. Therefore, the findings obtained from the laboratory can be hardly applied successfully to a real situation.

1992). Merton and Kendall who promoted the combination of these two designs said,

> Social scientists have come to abandon the spurious choice between qualitative and quantitative data: they are concerned rather with that combination of both which makes use of the most valuable features of each (Cited by Cohen & Manion, 1991:42).

Marton (1981) emphasizes that the results obtained by the two designs can present a better picture of the object investigated. They are more reliable and of higher validity in comparison with a single-design approach. Markee (1994) maintains the same view by saying, "qualitative and quantitative studies are in reality complementary ways of creating new knowledge. Recently, more researchers, instead of relying on one design exclusively, use whatever design is appropriate for their studies (for example, Wang, 1999; Wu, 1998; Hu, 1998.).

## Advantages of using two designs together

As said before, the use of the two designs together, compared with the use of a single design, can produce more powerful and more convincing results. The following is an example used by Patton (1990) to illustrate this point.

> In the early 70s, a new accountability system was implemented in some schools in Michigan. The new system was rather complicated, and included comprehensive and systematic procedures to evaluate the teachers, for example, standardized achievement tests, criterion-reference tests, teacher peer ratings, student ratings of teachers, parent ratings of teachers, principal rating of teachers, and teacher self-rating. The school authorities made a very positive assessment of such an accountability system while teachers' association and teachers themselves had rather

different views about it. In the spring of 1976, the Kalamazoo Education Association, with assistance from the Michigan Education Association and the National Education Association, conducted a study to find out the teacher's evaluation of and their own attitude towards the accountability program. The teachers were asked to answer a questionnaire that included both multiple-choice questions and open-ended questions. The quantitative data from multiple-choice questions indicated that almost all the teachers "felt the accountability system was ineffective and inadequate" (1990:20). 70% of the teachers who responded to the questionnaire also answered one of the open-ended questions and their answers filled 101 pages.

In this case the quantitative finding informed the school authority of the overall evaluation given by the teachers about the accountability system. The disadvantage of numbers was "detached" and "frozen" since they were not able to evoke readers' emotions. Furthermore, school board members could easily regard the quantitative findings as a predictable attempt of the union to discredit school officials. However, once the school officials read through a few pages of the teachers' own personal comments and the verbal descriptions of their anguish, fear and frustration with the accountability system, they understood the problems and also thought about what they should do to deal with the problems.

The above example vividly shows the strengths of combining quantitative with qualitative research. Quantitative research in this case involves a set of standardized statements which are difficult to design but can easily be answered by the teachers. Furthermore, the dataanalysis is simple and straightforward. However, the findings presented in numbers usually can not touch upon people's emotions. By contrast, the questions for qualitative researcb are simple in design but troublesome to answer.

The qualitative data are more detailed and more diversified in content, and thus analysis is much more difficult. Yet, the open-ended responses enable the researcher to understand and capture the points of view of other people which otherwise are not available.

## LINKS BETWEEN QUESTION AND DESIGN

In this section, I will try to answer an important question in design-selection, i.e. what is the relation between questions and design?

### Questions first or designs first?

What is our starting point when we select a research design? Should we start with research questions or with the research design? Logically speaking, questions should always come first. In other words, the question dictates the design rather than the other way around (Punch, 1998; Tashakkory & Teddlie, 1998). In this sense, the formulation of research questions must be finished before the choice of the design. To put it simply, we should first think clearly what we are trying to research and then consider how we are going to do it. If we are not sure what we aim at, then, there is no point of thinking of the design.

However, some researchers tend to put research designs before research questions. They may first decide what design they would like to use and then find research questions that can fit into the design. In this way, their research questions are guided by designs. Such a reversed sequence is very dangerous and can "lead to polarized thinking and intolerant and exclusionary attitudes" (Johnson, 1992: 228). That is to say, they may perceive different designs in conflict with each other which cannot be reconciled; or they may maintain that one type of design is superior to another without considering their strengths in addressing particular questions.

Actually there are no good or bad designs as such. Each design is effective to answer certain types of questions but not others. Each

research design has definite strengths and weaknesses. What counts is whether the design fits a particular question.

## How to match questions with designs?

Different questions need different designs to tackle them. Generally speaking, all research questions can be put into two big groups. One group of questions must be answered by quantitative designs while the other must be answered by qualitative designs. The common practice in today's research is to include two groups of questions in a single project. Thus, matching questions with designs becomes more essential. Often the wording of a question has implications for the designs to be chosen. For example, words like "variables", "the correlation between X and Y" , "factors affecting L2 achievement", etc. imply a quantitative design, while "How" and "Why" might imply a qualitative design. However, beginning research students often make a mismatch between a question and a design. One of the reasons is being unfamiliar with the implications of these words.

Let's look at some questions and see how the wording of a question may offer us information about the research design. Suppose we are doing research on motivation. The first question might be: What is the relation between students' motivation and their L2 learning outcome? (Or, to what extent can the students' motivation predict their L2 learning outcomes?) The wording like "relation" and "predict" clearly implies that the questions are quantitative in nature and thus a quantitative design such as a survey can well serve the purpose. A second question might be: Which approach (the Traditional Approach or the Communicative Approacb) is more effective to motivate the middle students to learn past tense in an L2 class? To make a comparison of the effectiveness of the two teaching approaches need a quantitative design such as an experiment. A third question might be: Why are some students better motivated than others? Obviously the reasons that can account for motivation might be various and difficult to quantify. Most likely, we

will employ a qualitative design such as a case study to find out such reasons.

Although certain words and phrases may imply designs, in many cases the wording of questions does not give us any clues to which design should be used. For example, our question might be "What are the differences between well-motivated and poorly-motivated learners?" The same question can be tackled by both quantitative design and qualitative design depending on how much previous research has been done in this area and what is the focus of the study. If there are a lot of similar studies available and the area is rather well-researched, a quantitative design might be a better choice to confirm or dispute the previous findings. If you are interested in detailed and vivid differences between these two kinds of learners rather than quantitative differences, a qualitative design certainly can do a better job.

## A MONO-DESIGN OR A MIXED DESIGN?

By using a mono-design, I mean only one design is employed, either quantitative or qualitative. Within the same design, the researcher may collect data from a single source or multiple ones. If employing a mixed design, both the quantitative and qualitative designs are adopted. These may be arranged in a different sequence and may differ in their contributions to answering the questions. The following section will discuss these two types of designs in turn. Since in most cases, the study conducted for writing an MA thesis or a Ph.D. dissertation demands a mixed design, the focus of this section will not be on a mono-design.

### A mono-design

A mono-design can be simple or complex. In a simple design, data are collected from one single source. Let's start with a simple quantitative design first. Suppose one researcher intends to find out what are the best ways to motivate students to learn an L2. She asks the students to answer a questionnaire in which various ways are listed and

the students are asked to rank order them. In this case, the data are from a single source, i.e. the students' responses to the questionnaire. The same study can result in a complex design if you want to. For example, in addition to the questionnaire for the students, the researcher might also ask the L2 teachers in 10 different schools to answer a questionnaire in which they are asked to rank order the same set of motivating strategies. The study could become further complicated by adding an experiment in which different motivating strategies are compared in terms of their effectiveness.

Similarly, a qualitative design can range from simple to complex. Let's look at the same question: What are the best ways to motivate the students to learn an L2. Suppose you interview 15 students who represent three levels of L2 proficiency in a class. The interview is conducted individually and each student is asked to list methods their teachers have tried to motivate them and then to make comments on these motivating strategies. If the study is limited to the students' interviews, then it is a simple qualitative design. You may make the design complex by including interviews of the teavhers or by adding the component of observing the classes to see how students respond to the teachers' motivating strategies, and what strategies teachers are actually using.

Let us look at two examples. The first example is for a pure quantitative design and the second one for a pure qualitative design.

> Example One: A pure quantitative design
>
> The study intended to examine the relationship between learner factors and English learning outcomes. In March 1995, more than 1,700 non-English majors who were from three universities located in Heilongjiang, Shangdong and Jiangsu responded to the "Learner Factors Questionnaire" and in June 1995, they took College English Test-band 4. The results yielded by maltiple regression analysis indicate which learner factors have

predicting power in relation to the students' scores on CET (Wen & Wang, 1996).

Example Two: A pure qualitative design

The study attempted to find out whether students with higher L2 proficiency used reading strategies differently if compared with students with lower L2 proficiency. The subjects involved were four students who were majoring in Russian and Japanese, and studied English as their second foreign language. All of them were good students in their major study but in English learning, two of them were at the lower-intermediate level, and the other two, at the upper-intermediate level. The researcher selected two reading passages to match their English proficiency respectively. Each student was asked to read the passage individually while speaking-aloud what was going on in his/her mind. The whole process was recorded and transcribed subsequently. The strategies used by the students were categorised and then presented. A comparison was made between these two groups in terms of the use of reading strategies (Lu, 1997).

## A mixed design

A **mixed design**, as its name suggests, always involves both quantitative and qualitative designs. However, the ways the two designs are arranged may be various. In terms of the time, we may have **sequential** or **parallel** designs. In a sequential design, the quantitative design is used before the qualitative one or the other way around. The sequential order of the two designs is logically decided by the research questions. The order cannot be changed unless the research questions are revised. In a parallel design, the quantitative or the qualitative are implemented independently. They can be operated simultaneously if sufficient human resources are available. Or they may be taken one after another since one is not dependent on the other. With reference to the

contribution either design can make to answering the research questions, we may have a **balanced** design or an **unbalanced** one. In the balanced one, the two designs (quantative and qualitative) have equal status while in the unbalanced one, one design is more dominant than the other. Usually a parallel design is a balanced one but a sequential design could either be a balanced one or an unbalanced one. Combining these two parameters, namely, time and importance, we may have various types of designs (Tashakkori & Teddlie, 1998). The following section will describe each of them.

### *Balanced mixed designs*

By a balanced mixed design, I mean that both a quantitative design and a qualitative one are used and furthermore, these two designs are equally important to answer the questions. We may have three possibilities as shown in Figure 5.1. The first possibility as shown in Figure 5.1a is a parallel design in the sense that the quantitative and the qualitative designs are used independently. The second one and the third one differ in their sequence. In the second one displayed in Figure 5.1b, the quantitative design is implemented before the qualitative one while in the third one as shown in Figure 5.1c the sequence is reversed.

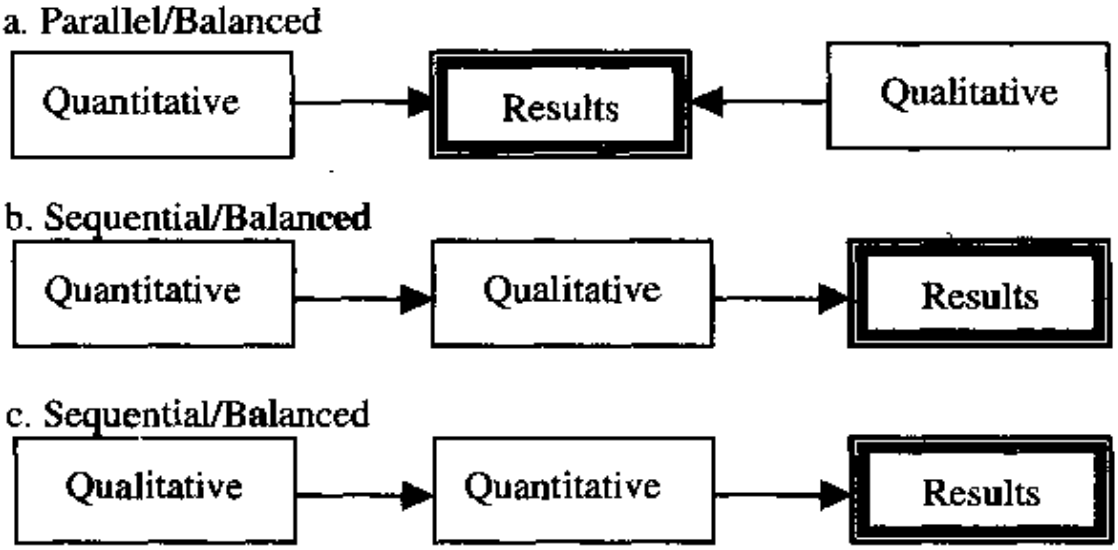



Figure 5.1: Variations within a balanced mixed design

Now let's look at three examples which show the three types of a balanced mixed design respectively.

Example One: Balanced/Parallel

The study (Wen, in press) intended to seek an answer to the question: Should English learning start from Primary One in the Chinese context? The subjects involved in the study included students, parents, teachers and administrators. The study consisted of two parallel components: quantitative and qualitative parts. In the quantitative part, Primary One students and their parents were asked to answer two different questionnaires respectively. In the qualitative part, Primary school English teachers and the school/district administrators were interviewed individually. The two sets of data were equally important since they answered the same set of questions from different perspectives.

Example Two: Balanced/Sequential (quantitative-qualitative)

The study attempts to address two questions: (1) to what extent L1 was involved in L2 writing and (2) how L1 was used in L2 writing. First of all, 100 students responded to a questionnaire in which various kinds of questions concerning the use of L1 in L2 writing were asked. Based on the students' responses to the questionnaire, 15 students who reported using L1 most frequently were selected to join the qualitative part in which the students were required to write an English composition while speaking out what is going on in their minds. The results from the quantitative part address the first question and the results from the qualitative part, the second question (A hypothetical study).

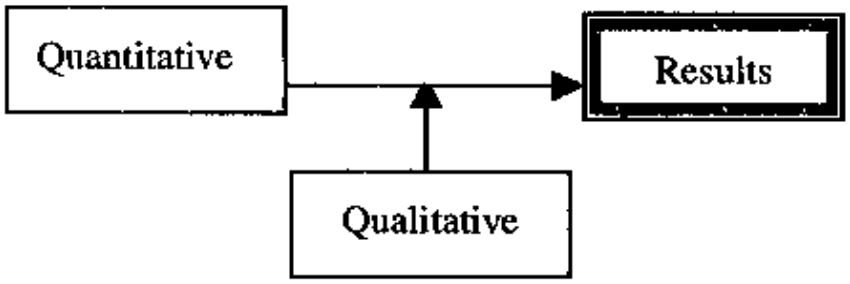
Example Three: Balanced /Sequential (Qualitative-quantitative)

In Gu's study (1997), he intended to find: (1) what strategies were used by non-English major undergraduates and (2) to what extent their use of strategies was related to their vocabulary learning. To answer the first question, he employed a qualitative design in which a group of second-year undergraduates representing three levels of English proficiency were interviewed individually. Based on the findings from the interviews, he constructed a questionnaire which was subsequently used in the quantitative part subsequently to answer the second question.

### *Unbalanced/sequential mixed designs*

In an unbalanced/sequential mixed design, both quantitative and qualitative designs are used but they differ in their roles in answering the questions. In the case of Figure 5.2a, the quantitative results play a dominant role while the qualitative findings are used to illustrate and illuminate the quantitative results. In the case of Figure 5.2b, the qualitative design is more important than the quantitative one since the qualitative findings are answers to the major research questions while the quantitative findings are not.

a. Quantitative dominating Qualitative




b. Qualitative dominating Quantitative

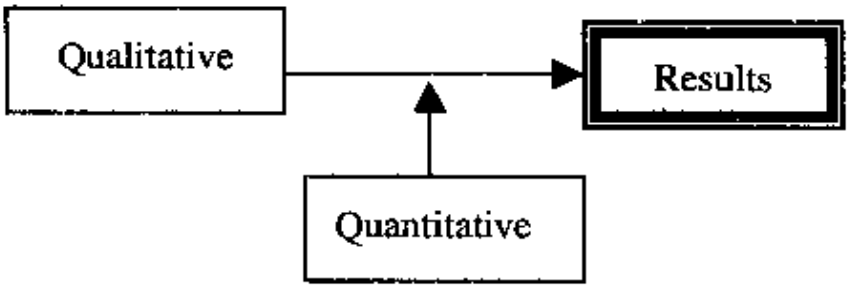



Figure 5.2: Variations within an unbalanced mixed design

Let us consider the following two examples:

Example one: Quantitative dominating Qualitative

In 1990, I carried out a study on the relationship between learner modifiable variables and English achievement, which used both quantitative and qualitative designs (Wen, 1993). In the quantitative part of this study, the subjects were asked to answer a questionnaire on learner factors. The results presented us with an overall picture of the extent to which learners' English achievement was influenced by the modifiable learner factors. Such results helped me narrow down the focus of the qualitative part of the study. Specifically speaking, the quantitative part of the study indicated that the areas worth further researching for advanced level English language learners were Vocabulary Strategy, Mother-tongue-avoidance Strategy, Tolerating-ambiguity Strategy, Management Strategy, the relation between beliefs and the choice of strategies, etc. However, the results obtained from the quantitative data were general and oversimplified because the quantitative study didn't show how successful learners differed from less successful learners. Without the qualitative part of the study, the interpretation of the statistical figures would be pure speculation. The qualitative part of this study includes interviewing, observation and diary. The subjects involvud in the qualitative part were five high-achievers and five low-achievers. The data yielded uncovered a lot of information on the strategies used by successful and less successful learners, which form a more complex and more realistic picture of the relations of the modifiable learner factors to English achievement. However, without the general framework provided by the quantitative part of the study, the focus of the qualitative part of the study may not be easily narrowed down.

Example Two: Qualitative dominating Quantitative

The project studied the differences in risk-taking between high L2 achievers and low L2 achievers. The researcher first identified 16 students from second-year English majors based on their previous scores on English tests and then interviewed them individually. She found a lot of interesting differences among which she decided to single out four striking differences related to L2 production for a survey study to validate these four differences. Thus a questionnaire was constructed and 100 students were asked to respond to the questionnaire.

I often encourage graduate students to use a mixeddesign in their research. However, in individual case, such a combination may not always be necessary.

## COMPLEXITIES IN CLASSIFYING DESIGNS

In the views of Punch (1998), Hughes and Tight (1996), and Tashakkori and Teddie (1998), there is a clear-cut division between quantitative and qualitative designs. Following their argument, we have to say a design is quantitative if the data are in the form of numbers. Otherwise, it is qualitative. However, in reality, the situation is much more complicated. In this section, I would like to use one example to illustrate such complexities and discuss possible solutions to the problem.

### An illustration

The following is a live example taken from our students' research.

Example One

In Chen's study (2000)[6], she intended to address the following three questions:

[6] Chen's study presented here has been modified to serve my purpose.

(1) How frequently do Chinese children adopt strategies in remembering the spelling of new English words?
(2) Are there any differences in the use of vocabulary strategies by good and poor child learners?
(3) How can learner factors (i.e. age and length of learning) affect the use of vocabulary strategies?

100 school children were involved in her study who were asked to remember 20 new words while speaking out what is going on in their minds. She first read the students' verbal accounts repeatedly to identify the strategies used by children in remembering the spelling of new words. Altogether she set up 10 categories, that is, Rote learning, Form association, Familiar word, Phonetic alphabet, Syllable-cutting, Word-building, Pronunciation rule, Visual frame, Chinese sound and Arbitrary cutting. All these categories were not available in the existing literature and were thus not preconceived ahead of her research. These categories emerged only in the process of analyzing the data. By repeatedly checking, she made sure that the 10 categories were well-grounded in the think-aloud data. She then counted the frequencies of each category occurring in their verbal accounts and then carried out Independent-samples t-test to examine whether there was any difference in the use of vocabulary strategies by good and poor child learners. Finally, she used correlation analyses to see the relationship between the use of strategies and the two learner factors. In the chapter of results and discussion, ahe answered each of the three questions one by one based on the quantitative results she got frem the analysis of the deta. Specifically speaking, the answer to the first research question is a table of frequencies, the answer to the second question is the results yielded by Independent-samples t-test and the answer to the third one is a set of correlation coefficients.

Consider the above example. Did the research use a quantitative design or a qualitative design? Following the definition given by Punch, we have to say the study was qualitative because the original think-aloud data are in words rather than in numbers. However, all the research questions were answered by figures rather than verbal accounts. Many researchers would intuitively feel it unacceptable that such a study be classified as a qualitative one. Let's closely examine the study:

Step One: collecting the data
The children produced verbal protocols that were not numeric in nature = **Qualitative data**

Step Two: identifying the categories
The researcher identified 10 categories by reading the verbal protocols repeatedly = **Qualitative analysis**

Step Three: Counting the frequencies
The researcher counted the frequencies of each category occurring in the students' verbal account = **Quantitative analysis**

Step Four: Comparing the average frequencies of strategy use by good and poor child learners
The researcher calculated the means of each strategy and then made a comparison between the good and the poor child learners in terms of the means = **Quantitative analysis.**

Step Five: Examining the relationship between the use of strategies and the two learner factors
The researcher first calculates the relation of the frequency of each strategy occurrence to the age factor and then to the factor of length of learning = **Quantitative analysis.**

There is no doubt that the original data were qualitative in nature and the initial analysis was qualitatively made. However, the subsequent analyses were all computed statistically. Above all, the answers to the research questions were presented in numbers. Evidently, it is an oversimplification to say Chen's study is either qualitative or quantitative.

## Suggestions

In my opinion, it is not appropriate to make a decision on whether a design is quantitative or qualitative simply by looking at the data form. We should also take data-analysis into consideration. Instead of using quantitative or qualitative as a cover term to describe a research design, I suggest that we specify the data form and the way of data-analysis separately. In this way, a study may have quantitative or qualitative data, or both, and data may be analyzed either quantitatively or qualitatively or both. Quantitative data are normally processed quantitatively while qualitative data have more choices. It is clear that complexities typically arise when the collected data are not in numbers.

## SUMMARY

Research designs are classified into two opposing camps: **quantitative** and **qualitative**. Some researchers regard the data form as the most critical feature for differentiating these two contrasting designs, i.e. whether data are in numbers (**Numeric**) or in words (**narrative**). In addition to this critical feature, there are many other differences between them. These differences are shown in the objectives, research questions, sample size, data-collection, data-analysis and outcomes. Generally, we can identify two historical stages: the **mono-design** stage (from the 19th century to the mid of the 20th century) and the **mixed-design** stage (from the 60s to the present). The most prevailing view now is to make use of the advantages of the two designs. The selection of research designs is primarily decided by research questions which very often give you some

clues in wording for what kind of design to choose. Once a mixed design is determined, you may have a series of choices. With reference to time, you may have **sequential** or **parallel** designs; with reference to the importance of contribution made by each design to answer research questions, you may have **balanced** or **unbalanced** designs. By combining the two dimensions, i.e. time with importance, you have a **balanced/parallel** design, and **balanced/sequential** design (quantitative-qualitative or qualitative-quantitative), or unbalanced/sequential designs (quantitative dominating qualitative or qualitative dominating quantitative). In the end of this chapter, the author analyzed the complexities in classifying designs, and suggested that instead of using a cover term (quantitative vs. qualitative) to describe a research design, it is better to specify a study along two dimensions: (1) **data form** and (2) **data analysis**.

## DISCUSSION QUESTIONS

1. What are the important differences between quantitative and qualitative designs?
2. Use an example to illustrate the advantages of combining the quantitative and qualitative designs.
3. What is the relationship between research questions and designs?
4. What are the possible combinations of the quantitative and qualitative designs?
5. What is the difference between the designs illustrated in 6.1b and 6.2a?
6. What is the difference between the designs illustrated in 6.1c and 6.2b?
7. Find examples from international journals to illustrate variations within a balanced mixed design and variations within an unbalanced mixed design (Note: you may form a group where each member is responsible for one example.)

# 6. A survey study

In Chapter 6, I will introduce to you how to carry out a **survey study** that primarily collects quantitative data. First I will give you a very brief description of a survey study followed by detailed procedures for data-collection in the survey. A paper that reports a survey study is presented as Appendix One.

## A BRIEF DESCRIPTION

A survey study usually involves a relatively large sample and is descriptive in nature. It either describes the features of a few variables or the relation between two or more variables. When your interest is confined to the relation between two variables only, the study is a correlation design. However, people nowadays have realized that understanding a phenomenon in terms of two variables is an over-simplification. A phenomenon is usually a result of interaction between a series of variables. Thus a more common design is tuultivariate in which the relation between a set of independent variables and dependent variables is examined. A survey study may be cross-sectional, when data are collected from different people at one point in time; it may be longitudinal, when data are collected from the same group of people at different points over a period of time.

To collect data in a survey study, you are expected to undertake several tasks: (1) constructing an instrument for data-collection; (2) selecting subjects for the study and (3) collecting data either personally or through correspondence. For the first two tasks, it does not matter which one is carried out first. In the following sections, I will describe these three tasks one by one.

# INSTRUMENT DESIGN: QUESTIONNAIRE

The questionnaire[7] is one of the most widely used techniques for collecting quantitative data. When compared with interviews and observation, it requires less time and less expense. A questionnaire consists of a series of questions which a subject is required to answer. The great weakness of questionnaire design is lack of theory in the sense that there are no scientific principles that guarantee an optimal or ideal questionnaire. Questionnaire design is a skill acquired through experience, and is an art rather than a science. What I can do here is to give you some guidelines. Although these rules can help you avoid major mistakes, the fine tuning of a questionnaire comes only from your creative and rich practice.

In the following section, I will first introduce to you the criteria for preparing a sound questionnaire, then the process of constructing a questionnaire and finally scaling techniques in a questionnaire.

## Criteria for e good questionnaire

Any good questionnaire should at least meet the following two criteria: (1) having high internal validity; and (2) taking a professional outlook.

### *Having high internal validity*

By saying a questionnaire should have high internal validity, I mean that items in the questionnaire must measure the variables you really want to investigate. You might say this criterion is unnecessary since no researcher would be so stupid as to fail to do so. It is true that no researcher would like to violate this requirement intentionally. However, they might fail to meet the criterion because of lack of skill or experience.

---

[7] The questionnaire can also be used to collect qualitative data if the questions are open-ended. In this chapter, I will only focus on structured questions that are used to collect quantitative data.

Actually, having lower internal validity is a common flaw occurring in a questionnaire constructed by a novice researcher. Let's look at some examples. One of my BA students intended to investigate whether senior middle school students used L1 in the process of L2 reading. Besides many other items, her questionnaire included the following ones:

(1) I often read reference materials in Chinese in order to help me understand the context in which an English story took place.
(2) I read English novels, newspapers, journals, etc. in my spare time.
(3) I only have English-Chinese bilingual dictionaries.

Suppose some of the students responded to the first item by choosing "This statement is always or almost always true of me". Did this response imply that they used L1 in the process of L2 reading? It is not clear at all. So far as the second item is concerned, frequent reading in English cannot guarantee the avoidance of the use of L1 in L2 reading. The third item might not be working at all since almost all the senior middle school students would give the same response. Furthermore, even if there were variations in the responses, the responses could not give us any information directly concerning the use of L1 in L2 reading.

One of my MA students wanted to investigate students' degree of anxiety in speaking L2. The following three items were part of her questionnaire:

(1) I try to avoid answering questions in class.
(2) I try to sit in the corner of the classroom so that I will not be asked questions frequently by the teacher.
(3)I do not like to speak in public.

*What do you think of* the above items? If a student did admit that the above three statements were true of him, could you say that this student

showed a high degree of anxiety in speaking L2? Obviously not. The avoidance of answering questions in class may be due to a variety of reasons, such as the student's poor preparation for the lesson, the subject's lower L2 speaking ability, the subject's introverted personality, in addition to the student's anxiety in L2 speaking. Therefore, the above three statements would measure many other factors in addition to the student's degree of anxiety in L2 speaking.

From the above two examples, you might have a better understanding of what is internal validity in questionnaire items. To make a self-evaluation, you may ask yourself from time to time the questions "What does the student's response to a particular item mean? Does it have one or multiple interpretations?" If there is only one interpretation which is directly linked with the variable you are examining, then the item concerned is of high validity. Otherwise, you have to revise it or delete it.

### *Taking a professional outlook*

How a questionnaire looks can affect the attitudes of respondents which in turn influence the quality of their responses. For instance, if a questionnaire is printed on poor-quality paper or looks shabby in appearance, the respondents will think the research project is not important and, as a result, they do not take it seriously. A good questionnaire with a professional outlook should not contain the following problems.

First of all, you should avoid clipping or stapling papers together if a questionnaire contains more than 5 pages. In this case, it is better to take the form of booklet since booklets do not easily fall apart like clipped and stapled papers. Furthermore, booklets allow the use of double-page format for questions, and look more professional. Secondly, one question (including response categories) should not be printed across two pages since split questions may lead respondents to think that the question has ended at the end of a page. Thirdly, you should avoid crowding questions together to make the questionnaire look shorter. Overcrowded

questions with little blank space between questions can lead to errors in data collection. Moreover, they give the impression that the questionnaire is complex and this can result in a lower cooperation and response rate. Finally, the paper for printing should be of high quality and the printing itself must be clear and easy to read. Remember that a questionnaire with a professional look can not ouly minimize respondent's fatigue, boredom and effort, but also increase the response rate.

## Questionnaire design process

The design of a questionnaire can be described in terms a series of steps that include: (1) selecting the mode for collecting questionnaire data; (2) specifying what kind of data you intend to collect; (3) determining whether questionnaire items will be processed individually or as categories; (4) deciding the content of individual items; (5) choosing the question structure; (6) selecting the question wording; (7) arranging the questions in proper order; (8) deciding the format of the questionnaire and (9) pretesting the questionnaire (Malhotra, 1993). In the following section, each of the above steps will be discussed in detail.

### *Selecting the modes of administration*

By the mode of administration, I mean in which way questionnaires are distributed to the respondents. In the Chinese context, there are two common modes: (1) by person and (2) by mail. By person it means that questionnaires are distributed to the respondents by an organizer rather than through correspondence. The organizer can be the researcher herself/himself or other people the researcher has asked for help. By this mode, a group of subjects usually answer the same questionnaire together. The organizer gives the respondents instructions on how to respond to the questionnaire and clarifications if there are any ambiguous items. The organizer can also check whether the respondent has responded to all the items and thus the rate of missing data can be greatly reduced.

If the respondents are spread out in different cities or even in different countries, it is difficult to employ the first mode. What you can do is to mail questionnaires to the targeted people. By doing so, the researcher cannot control the overall response rate and the rate of missing items. Even if a questionnaire is returned, s/he cannot be sure that the questionnaire is filled out by the people s/he expected and under what circumstances the questionnaire is completed.

The instructions of and the length of a questionnaire vary depending on whether questionnaires are distributed by person or by mail. The mailed questionnaire should contain more explicit and simple instructions than the person-administered questionnaire. Furthermore, the mailed questionnaire should be shorter and easier to answer than the person-administered questionnaire because it is next to impossible to get respondents to fill out a mailed questionnaire for one or two hours unless you are prepared to pay well for their time.

Each of the modes of administratiou has its own strengths and weaknesses. Your choice of a mode will depend on your own consideration of cost, convenience and the nature of the questions you are asking. I highly recommend the person-administered mode since it allows you to clarify ambiguities and check whether all the questions are answered by the respondent.

### *Specifying what kind of data you intend to collect*

At this stage, you have to understand what kind of data you really want to collect. At the very general level, there are basically five types of information:

(1) Experience/behaviour (What does a respondent do?);
(2) Opinion (What does a respondent think?);
(3) Feeling (How does a respondent feel?);
(4) Knowledge/abilities (What does a respondent know and what is a respondent able to do?

(5) Background (such as the date of birth, age, gendar, marital status)

For example, in my own study on the relation between L2 modifiable learner variables and English achievement, the questionnaire I constructed included three types of information: (1) background; (2) opinion, i.e. beliefs; (3) behavior, i.e. learning strategies. The following is a shortened version of the original questionnaire (Wen, 1993: 294-302).

Language Learner Factors Questionnaire

The following questionnaire is designed for research on the way Chinese students study English. The questions have been carefully selected to cover the more important aspects of English learning. Please answer each question honestly and frankly according to your own opinion or learning experience. There are no "correct" answers. All the data collected will be highly confidential and will be used for the research only.

Part A Background information

1. Name _____; Date of birth _____; Sex _____

2. Scores of entrance examination: English _____

3. Scores of entrance examination: Chinese _____

Part B Beliefs

Below there are some beliefs that people have about learning foreign languages. There are no right or wrong answers. We are simply interested in your opinions. Please write the number which best indicates your opinion in the bracket at the end of each statement. The numbers stand for the following responses:

1 = I strongly disagree with this statement

2 = I disagree with this statement

3 = I neither agree nor disagree with this statement

4 = I agree with this statement

5 = I strongly agree with this statement

B01 Planning your study time is important for success.

B02 Learning a foreign language requires painstaking effort.

B03 It is important to repeat a lot.

....

Part C Learning strategies

Below there are some strategies that people use when learning a foreign language. Please read each statement and write down the number that best describes you in the bracket at the end of each statement. Please indicate **what you really** do, not what you think you should do, or what other people do. Remember there are no right or wrong answers. The numbers stand for tha following responses:

1 = This statement is never or almost never true of me

2 = This statement is usually not true of me.

3 = This statement is somewhat true of me

4 = This statement is usually true of me

5 = This statement is completely or almost completely true of me

C01 When reading a text, I try to understand everything in it.

C02 I memorize texts.

C03 I talk to myself in English outside of class.

....

You can relate the above five types of information to the time frame: past, present and future. For example, an item about one particular behavior may refer to the respondent's action in the middle school, or in the university now, or in the future when they take a postgraduate program.

Very often, novice researchers cannot clearly differentiate belief items from behavior items. The structure of a belief item is "I think/believe that...". For example, the statement "I think learning an L2 well requires painstaking effort" is to find out the respondent's opinion or view rather than the behavior. In contrast, the statement like "I put a lot of effort into learning an L2" is to investigate the student's behavior rather than their view. Unfortunately, a novice researcher more often than not mixes up the two types of information and that certainly confuses the respondents. When respondents take beliefs for behaviors or behaviors for beliefs, the validity of the data is then called into question.

### *Determining the way you process the questionnaire data*

Once you have decided in which mode to administer the questionnaire and what type of information you intend to collect, then you need to determine whether your questionnaire data are eventually analyzed in terms of individual items or in terms of categories. If the basic utit in the data analysis is an individual item, such as describing the frequency and the mean of one isolated item, then such a questionnaire is called an individual-item-based questionnaire. The construction of such a questionnaire is comparatively simple and easy because it does not involve setting up conceptual categories. However, this kind of questionnaire has very limited value because its results are less generalizable.

If the data-analysis is to be based on categories, each of which contains a few questionnaire items, you must establish conceptual categories either by a top-down approach or by a bottom-up approach. By a top-down approach you construct conceptual categories based on logical arguments or existing theories before you design specific items. By a bottom-up approach, you simply write down whatever items occur in your mind. You then try to classify these items into categories. Obviously, constructing a category-based questionnaire is very demanding and challenging on the part of designers. Yet all MA students

need to grasp this skill. For example, I designed a category-based questionnaire to find out second-year English Majors' beliefs about and strategies for L2 learning. The conceptual categories constructed are presented in Table 6.1.

| Approach | Traditional | Non-traditional |
|---|---|---|
| Beliefs | Form-focused beliefs<br>Mother-tongue-reliance beliefs<br>Accuracy-focused beliefs | Function-focused beliefs<br>Mother-tongue-avoidance beliefs<br>Fluency-focused beliefs |
| Strategies | Form-focused strategies<br>Mother-tongue-reliance strategies<br>Accuracy-focused strategies | Function-focused strategies<br>Mother-tongue-avoidance strategies<br>Fluency-focused strategies |

Table 6.1: Categories of beliefs and strategies

The above two types of learning approaches: Traditional versus Non-traditional are based on three major controversies in L2 learning proposed by Stern (1975, 1983, 1992): the L1-L2 connections, the explicit-implicit option and the code-communication dilemma. Each controversial issue could be perceived as a continuum. Let's take for example the issue of the L1-L2 connection, which is about the role of the learner's mother tongue in L2 learning. Generally speaking, there are two contrasting views. One is the view that learning a second language should exploit the first language and thus the learner is encouraged to use translation in learning the target language. The other is the view that learning the target language should be entirely within and through the target language and so the learner is exhorted even to think in the target language. Centering around the issue, I set up two contrasting categories about beliefs and strategies respectively: Mother-tongue-reliance beliefs and Mother-tongue-avoidance beliefs; Mother-tongue-reliance strategies and Mother-tongue-avoidance strategies.

Furthermore, an approach is regarded as a combination of beliefs and strategies. Such a conceptual definition is based on the theory of

educational psychology proposed by Marton and Saljo (1976) who advocate that students' actual learning behaviors are influenced by their conceptions of learning. In other words, students' beliefs about learning will affect their choice of learning strategies.

One of my B.A. students (Chen, 1996) used a category-based questionnaire to investigate senior middle school students' degree of anxiety for the purpose of discovering to what extent the teacher's attitude affects the student's degree of anxiety. When he first constructed questionnaire items, he did not have any preconceived categories. He just wrote down all the items he could think of and then tried to classify them into categories. Through repeatedly examining the items in relation to the research questions, he decided to group them into two categories: environmental anxiety and inherent anxiety. The first type of anxiety refers to anxiety induced by the L2 learning environment which includes the classroom atmosphere and teachers' attitude towards the students; the second type refers to anxiety caused by students' personality.

### *Deciding on the content of individual items*

Once you are clear about what type of information is needed, you start with thinking of the content of each individual item. You should not sit in the room thinking alone. The sources of questionnaire items are varied. One source is the items occurring in previous studies which are related to your research topic. For example, if you want to investigate university students' strategies used in learning L2, you may refer to Oxford's Strategy Inventory for Language Learning (1990) or Wen's Learner Factors Questionnaire (1993). Or you may start by interviewing some typical university students and asking them how they go about their L2 learning. The interview data may provide you with a lot of details on their strategies.

If a questionnaire is intended to contain several categories, you have to consider whether the items designed cover all the categories and whether each category contains an adequate number of items. In order to

answer these two questions, you have to make up a table which has two columns. In one column, you write down the categories and in the other column, you put down the sequential number of each questionnaire item (See Table 6.2).

Then you have to read the items within the same category again and again to make sure that all the items logically go together. If you find some items do not fit the category, you don't need to feel frustrated since it is a very common problem at this stage. However, you cannot ignore it. What you should do is to revise them or construct new ones.

| Category Name | Item No. |
|---|---|
| Form-focused beliefs | 2, 4, 5, 7, 9 |
| Meaning-focused belief | 1, 3, 6, 8, 10 |
| Form-focused strategies | 11, 13, 16, 18, 20 |
| Meaning-focused strategies | 12, 14, 15, 17, 19 |

Table 6.2: A table for checking questionnaire items within each category.

Finally you have to check question wording very carefully. If a question is worded poorly, the respondents may fail to understand the item or misunderstand it. Their answers under such circumstances tend to contain more missing values[8] or to be misleading. To avoid poorly-worded questions, the following are some of the suggestions made by Malhotra (1993) and Bernard (1994):

**Use ordinary words** Ordinary words should be used in writing a question and they should be suitable to the vocabulary level of the respondents. Suppose the respondents involved in your research are primary school students. The words used in the questions must be within

[8] Missing values result from the questionnaire items that were not answered by students.

their range. Otherwise, they are not able to answer or the answers given are of low validity.

**Avoid leading questions** A leading question is one in which a hint is given so that the respondent is biased towards a particular choice. For example,

The unpopular course of Intensive Reading should be revised.

| Strongly disagree | Disagree | Uncertain | Agree | Strongly agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Obviously, the above question would bias respondents to a " strongly disagree" answer since the word "unpopular" is a biased word.

**Using dual statements: positive and uegative** By using dual statements, I mean that positive and negative statements should be roughly balanced in a questionnaire. In other words, the total number of positive statements should be more or less the same as the total number of negative statements. For example, in one of my own studies, I intended to investigate the students' attitudes towards using a language laboratory to teach spoken English. The following are the statements from the original questionnaire[9], and three are positive and five, negative marked with a star:

1 = This statement is never or almost never true of me

2 = This statement is usually not true of me.

3 = This statement is somewhat true of me

4 = This statement is usnally true of me

5 = This statement is completely or almost completely true of me

1. I like to have a speaking class in a language laboratory.

[9] The original questionnaire items were in Chinese and the English items presented here result from translation.

2. I am not clear about why we have a speaking class in a language laboratory.*
3. In a language laboratory, I cannot concentrate on what I want to say because so many people talk to microphones at the same time and they make a big noise.*
4. I don't like the teacher to monitor my conversation with others.*
5. I think it is easier to combine listening with speaking when a speaking class is undertaken in a language laboratory.
6. I think in a speaking class listening practice should be kept to minimum.*
7. I think talking to other people through a microphone in a language laboratory lacks communicative authenticity *
8. I feel talking to people through a microphone in a language laboratory is the same as talking to people over the phone in our daily life.

From the above examples, you may find negative statements often contain explicit negative markers such as "not clear" in No. 2, "cannot" in No. 3, and "don't like" in No. 4 but sometimes they use implicit negative terms such as "lack" in No. 7.

Why do we need to do so? Because research shows that the way respondents answer the questionnaire is affected by whether statements are positive or negative.

**Being cautious in translation** The last suggestion is made for cross-cultural studies where in most cases two or more different languages are involved. Very often apparent equivalents of two different languages actually contain different meanings. How can we make sure our translation did not distort the original meaning? The best way to do this is through back translation. For example, I once conducted a study on Chinese and Swedish students' conceptions of learning. The first questionnaire was in English which was a common language shared by

my partner and me. But both of us needed to translate the questionnaire into our native languages for data collection. What we did was back-translation. First we asked bilinguals to translate the English version into Chinese and Swedish respectively. We then asked another two bilinguals to translate the Chinese and Swedish versions back into English. By comparing the initial and final English versions, we could easily detect the problems and make further modifications until a satisfactory back satisfaction was achieved.

### *Choosing question structure*

Basically, there are two types of questions: open-ended questions or unstructured questions versus closed questions or structured questions. Now we are going to deal with these two types of questions one after another.

Open-ended questions refer to those whose answers are in the respondent's own words and are difficult to predict. The following are some examples:

Why don't you like to read simplified English novels?
When do you consult a dictionary?
What do you read outside class?
Why do you study English?

From the above examples, we can see that open-ended questions begin with "Wh-words" such as what, when and why. In fact, this type of question is typically used to collect qualitative data. Therefore, open-ended questions are not further discussed in this chapter.

In closed-questions or structured questions, the researcher provides respondents with a set of response alternatives. In other words, the respondents have to select one response out of the alternatives provided. They do not have freedom to use their own words to express their

response. The closed-question may be multiple-choice, dichotomous or a scale.

**Multiple-choice questions** In multiple-choice questions, the researcher offers a set of answers and the respondent is expected to choose one or more of the alternatives given. Let's look at the following two examples that are taken from Wu's questionnaire (1998) with some modifications.

1. How do you learn the cultural differences in your daily life? Please check as many as you like.
   (1) Read English magazines and books on Western customs and etiquette
   (2) See English movies and videos
   (3) Actively interacting with native speakers of English

2. When you talk with native speakers of English, which aspect do you attach the most importance to? Please choose only one.
   (1) Accuracy of language forms
   (2) Explicitness of meaning
   (3) Fluency of speaking
   (4) Politeness of speech

**Dichotomous questions** A dichotomous question has only two alternatives to choose, such as yes or no; important or unimportant; disagree or agree, etc. Often, the two choices provided are supplemented by "don't know" as a neutral alternative. Consider the following examples:

(1) Do you talk to yourself in English outside class?
   —Yes
   —No
(2) Do you often read English newspapers?

—Yes

—No

(3) Do you think people with talents for language learning can learn an L2 without effort?

—Yes

—No

—Don't know

**Scales** A scale is a continuum upon which a set of alternative choices are placed. The following are the examples of scales:

(1) Do you guess the meaning of new words when reading English novels for pleasure?

| Never | Occasionally | Sometimes | Often | Very often |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

(2) Do you think the best way to understand the English text is to translate?

| Strongly disagree | Disagree | Uncertain | Agree | Strongly agree |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Scales are apparently similar to multiple choice questions since both provide a set of response alternatives but they are different in nature. You see, the alternatives in the multiple choice question cannot form a continuum (Note: Detailed information about scales will be presented in the next section "Scaling techniques" ).

### *Determining the order of questions*

The order of questions often affects respondents' answers. Therefore, we need to arrange them with great caution. The following are some tips:

**Background information first** Questions concerning background information in the Chinese context are usually placed before questions about basic information such as the respondents' opinion, behavior, knowledge, etc. However according to Malhotra's view (1993), in the western context, basic information should be obtained before background information.

**General questions before specific questions** General questions usually go before specific questions. This can help prevent specific questions from influencing the responses to general questions. For example, if you want to find out what strategies students use in their reading, a general question like "How do you go about reading?' should be asked before the specific question "How do you deal with new words in reading?"

### *Deciding the format of a questionnaire*

Before you print out questionnaires, you need to decide the format, space and position of questions. This is particularly important for self-administered questionnaires. Research findings show that questions placed at the top of the page receive more attention than those placed at the bottom. Questions that are divided into several sections tend to increase the response rate. Furthermore, all the questions must be numbered and the way the answers are recorded should be convenient for the respondents to write.

### *Conducting a pilot study to test the questionnaire*

A pilot study is one in which a questionnaire is tested on a small sample of respondents to detect and overcome potential problems. As a normal practice, no questionnaire should be used in a formal study without testing it since even a questionnaire designed by an experienced researcher can always be improved through the pilot study. The potential problems may lie in all aspects of a questionnaire, such as question content, wording, sequencing, question difficulty, instructions and format. The respondents selected for the pilot study should be similar to those

who will be involved in the actual survey. The best pilot study is done through personal interviews in which you can observe the respondents' reaction and attitudes, and check whether their interpretation of the question is the same as you intended. Finally, the results of the pilot study should be analyzed to see whether your research question are set up properly and whether the conceptual categories, if there are any, are supported by empirical data.

## SCALING TECHNIQUES

Scaling involves creating a continuum along which people's responses can be recorded and eventually each response can be assigned a number. For example, you might like to label peoples' attitudes towards L2 learning as unfavorable, neutral or positive; or you might like to classify the frequency of using strategies into frequently, sometimes and occasionally. Here, unfavorable-neutral-positive and frequently-sometimes-occasionally are two continua. The process of establishing these two continua is scaling. The scaling techniques used in applied linguistics include comparative and noncomparative scales. In the following section, these two types of scales will be described in detail.

### Comparative scales

Comparative scales involve the direct comparison of choices. For example, respondents might be asked whether they prefer speaking in public or speaking in a small group. Or respondents might be asked to rank a list of items in order. For example,

> How do you improve your reading skills? (Rank order the following items according to the frequency of your actions)
>
> (1) Read English magazines
> (2) Read English newspapers
> (3) Read simplified novels

(4) Read English novels without simplification

Comparative scales can only yield ordinal data and thus they are interpreted in relative terms. As shown in Figure 6.1, comparative scales include paired comparisons and rank order.

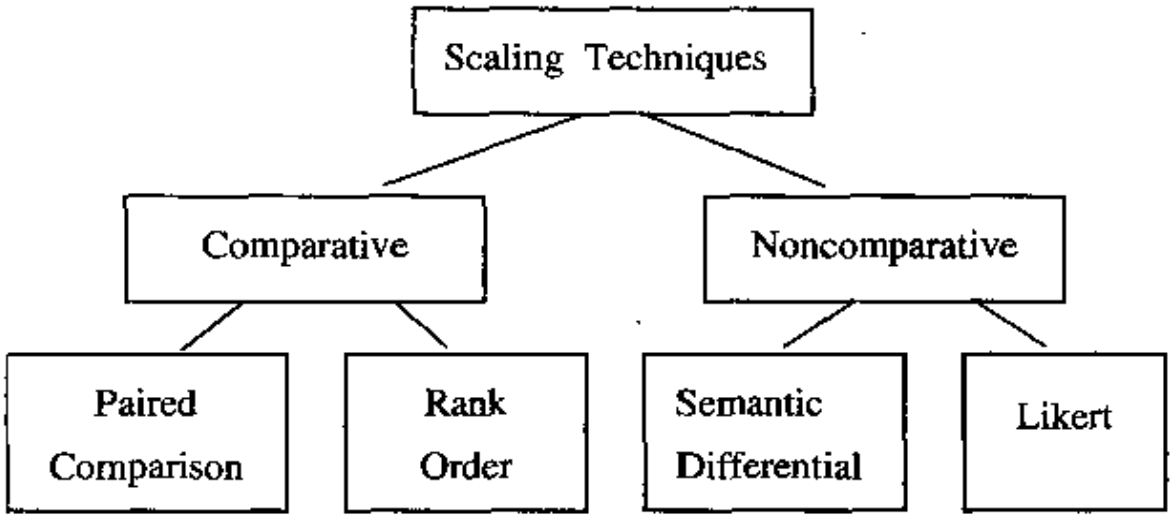



Figure 6.1: Classification of scaling techniques (Adapted from Malhotra, 1993: 282)

### *Paired comparison scaling*

Paired comparison scaling, as its name suggests, only offers two choices to the respondent who is asked to compare them according to some criterion. Por example,

1. When you speak English, which of the following two aspects do you pay more attention to?
   (1) Accuracy
   (2) Fluency

2. When do you feel more nervous?
   (1) Talking to native speakers of English
   (2) Talking to your English teachers

3. When do you want to speak more?

(1) Speaking in a group

(2) Speaking in pairs

### *Rank order Scaling*

Different from paired comparison, rank order scaling expects respondents to order or rank more than two choices with reference to some criterion. For example,

1. Why do you want to learn English? (Rank order the following items according to your perceived importance.)
   (1) I am interested in learning a foreign language.
   (2) I want to get a high-paid job after graduation.
   (3) I plan to study abroad.
   (4) I like to know more about the foreign culture.
2. When do you want to put more effort into learning English?
   (1) When I get higher scores on the English test
   (2) When I am praised by my English teacher in class
   (3) When I obtain a prize in an English speaking competition

According to Malhotra (1993), the major advantage of comparative scaling is that small differences between two or more choices can be listed since respondents are forced to choose between them. Furthermore, respondents undertake the rating task from the same known reference point and thus comparative scales can be easily understood. Another advantage is that designing such items does not require any theoretical assumptions.

## Noncomparative scaling techniques

In noncomparative scales, each item is scaled independently of the others. The resulting data are generally assumed to be intervally scaled. There are two kinds of noncomparative scales: semantic differential scale and Likert scale, as shown in Figure 6.1.

### *Semantic differential scale*

The semantic differential scale is a five- or seven-point rating scale with end points associated with bipolar labels that have the opposite semantic meaning. The subjects mark the blank that best indicates their views or behaviors. For example, Gardner and Lambert (1972: 156-157) employed such a scale to investigate parental encouragement:

How much do your parents encourage you to study French?

Not at all __: __: __: __: __: __: __ Very much

They also used a similar method to find out the subject's impressions of French people from France:

| | | |
|---|---|---|
| Interesting | __: __: __: __: __: __: __ | Boring |
| Prejudiced | __: __: __: __: __: __: __ | Unprejudiced |
| Brave | __: __: __: __: __: __: __ | Cowardly |
| Handsome | __: __: __: __: __: __: __ | Ugly |
| Colorful | __: __: __: __: __: __: __ | Colorless |
| Friendly | __: __: __: __: __: __: __ | Unfriendly |
| Honest | __: __: __: __: __: __: __ | Dishonest |
| Stupid | __: __: __: __: __: __: __ | Smart |

One thing that has to be emphasized here is that the negative adjectives or phrases sometimes appear at the left side of the scale and sometimes at the right. This controls the tendency of some subjects, particularly those with very positive or very negative attitudes, marking the right- or left-hand sides without reading the labels.

### *Likert scale*

The Likert scale was first introduced by Rensis Likert. As an example, in one study, the subjects were asked to evaluate their reasons for learning English on a four-point scale (1=not important; 2=somewhat important; 3=important; 4=very important). This is called a Likert Scale.

To conduct the analysis, each response is assigned a numerical score, ranging from 1 to 4. When using this approach to determine the total score for each subject, it is important to use a consistent scoring procedure so that a high (or low) score consistently reflects a favorable response. This requires that the numerical value assigned to the negative statements by the subjects be scored by reversing the value. Note that for a negative statement, an agreement reflects an unfavorable response, whereas for a positive statement, an agreement reflects a favorable response. Accordingly, a "strongly agree" response to a positive statement and a "strongly disagree" to a negative statement would both receive the same score.

## SELECTING SUBJECTS

Usually, people do not study the whole population. The reason is very straight forward: "You cannot study everyone everywbere doing everything" (Miles and Huberman, 1994: 27). The common practice is to select a sample from the population to study, hoping the findings from the sample can be applied to the whole. In the following section, I will explain to you how to select subjects.

Selecting subjects for a survey study is neither less difficult nor less important than constructing an instrument. If the subjects are not properly selected, the data collected will certainly be of poor quality. In the worst case, where the subjects selected are not cooperative or even do not turn up, the data-collection will either be spoiled or be jeopardized. In this section, we will first discuss standardized procedures for selecting subjects, and then focus on some practical concerns.

### Random sampling techniques

There are three basic random sampling techniques: simple random sampling, systematic random sampling and stratified random sampling.

### *Simple random sampling*

Simple random sampling is the basis for the other two random sampling techniques. By simple random sampling, each unit is numbered from 1 to $N$ (where $N$ is the size of the population). Next, a table of random numbers is used to select $n$ items to the sample. Table 6.1 presents some random numbers (The whole set of random numbers is presented as Appendix Five). These numbers are random in the sense that for each number, any of the ten digits (0-9) is equally likely to occur and so getting the same digit twice or more is possible. In the process of selecting, a number that is bigger than the population size $N$ or a number that has been chosen for the second time has to be ignored.

The following is an example in which the use of random sampling technique will be illustrated.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 91567 | 42595 | 27958 | 30134 | 04024 | 86385 | 29880 | 99730 |
| 46503 | 18584 | 18845 | 49618 | 02304 | 51038 | 20655 | 58727 |
| 34914 | 63976 | 88720 | 82765 | 34476 | 17032 | 87589 | 40836 |
| 57491 | 16703 | 23167 | 49323 | 45021 | 33132 | 12544 | 41035 |
| 30405 | 83946 | 23792 | 14422 | 15059 | 45799 | 22716 | 19792 |
| 09983 | 74353 | 68668 | 30429 | 70735 | 25499 | 16631 | 35006 |
| 85900 | 07119 | 97336 | 71048 | 08178 | 77233 | 13916 | 47564 |

Table 6.3: A brief table of random numbers (Taken from Black, 1992: 256)

Suppose you want to select 10 out of 20 universities by simple random sampling for your study. First of all, you number every university of the population (See Table 6.4). How many digits you use for the first number depends on the digits of the population size. That is to say, if the population size is two digits, you start with 01; if the population size is three digits, you start with 001.

| | |
|---|---|
| 01 Beijing University | 11 Liaoning Normal University |
| 02 Qinghua University | 12 Sichuan University |
| 03 Beijing Normal University | 13 Wuhan University |
| 04 Fudan University | 14 Nankai University |
| 05 Huadong Normal University | 15 Zhongshan University |
| 06 Nanjing Normal University | 16 Henan University |
| 07 Shanghai University | 17 Shantou University |

| | |
|---|---|
| 08 Dongnan University | 18 Zhejiang University |
| 09 Nanjing University | 19 Shangdong University |
| 10 Hehai University | 20 Xiamen University |

Table 6.4: Numbered population of 20 universities

Secondly, select as many digits as there are in the sample. In this case, since the sample size is 10 universities out of the 20, what we need to do is to look at two digits as a unit each time in the table. All the numbers greater than 20 must be ignored. Now let's look at every two-digit unit of all the numbers in Table 6.3. In the first row of digits in the table, the first two-digit unit is 91. This number is out of range, so it has to be ignored. The next two-digit unit is 56. Next is 74, followed by 25, 95, 27, 95, 83. All these numbers are beyond the range. The two-digit unit after 83 is 01 which is useable. Continuing the process, skipping over 34, we will get 04 and 02, both of which are within the range of 20. Continuing along the first row, we have to ignore 48, 63, 85, 29, 88. The next number is 09 which can be used. The remaining two numbers 97 and 30 are not usable, either. By the same method, we will move to the second row and all the numbers 46, 50, 31, 85, 84, **18**, 84, 54, 96, 18, 02, 30, 45, **10**, 38, **20**, 65, 55, 87, 27, except 18, 10 and 20, cannot be used either because they are out the range of 20 or because they have already been used before such as 02. Up to now, we have 01, 02, 04, 09, 10, 18, 20. Continuing the process, we will get another five numbers: 11, 03, 16, 13. The following universities are the ones chosen by simple random sampling.

| | |
|---|---|
| 01 Beijing University | 10 Hehai University |
| 02 Qinghua University | 11 Liaoning Normal University |
| 03 Beijing Normal University | 13 Wuhan University |
| 04 Fudan University | 18 Zhejiang University |
| 09 Nanjing University | 20 Xiamen University |

Table 6.5: The universities selected by simple random sampling

In the above example, we checked the random numbers in the table row by row. You can also do it column by column. Such a decision is made purely based on your personal preference. But once you have made a decision, you have to stick to it all the way through. Furthermore, you are not allowed to skip over any number unless they are out of the range or they are repeatedly chosen.

### *Systematic random sampling*

If a population is large but the intended sample size is small, simple random sampling is not suitable because the sample selected might not be evenly distributed among the population. Suppose the total number of a population consists of 900 students and you would like to select 30 students out of them as a random sample. Following the technique of simple random sampling, we might get 30 students when we only reach half of the 900 subjects. To avoid this problem, we use another random technique, i.e. a systematic random sampling technique which can make sure that the subjects selected are evenly spread out among the population. As with simple random sampling, the systematic random sampling technique also requires the numbering of all the suhjects. Then, the following formula is used to calculate the interval:

$$\text{Interval} = \frac{\text{The total number of the population}}{\text{The size of the sample}}$$

Once the interval is decided, the subjects are selected according to the interval. How can we get 30 out of the 900? The interval is 30 resulting from 900/30. Then one subject is chosen at every 30$^{th}$ interval. What number do you start with? You have to use a table of random numbers to select a number between 1 and 30 as a starting point. According to Table 6.1, the first number is 915 that is bigger than 900 and thus is skipped over. The next nnmber is 674 that can be used in this case. The second

number is 704 (i.e. 674+30), and the third one is 734 (i.e. 704+30). Following the same procedure, you will get 764, 794, 824, 854, 884. Once you get 884, you will have a difficulty. Adding 30 to 884, you will get 914 which is beyond 900. What should you do now? The solution is simple: just subtracting 900 from 914, you will get 14 which is the ninth number in the following list. The tenth number is 44 resulting from adding 30 to 14. Eventually, the subjects which are chosen are the following:

674, 704, 734, 764, 794, 824, 854, **884**, **14**, 44, 74, 104, 134, 164, 194, 224, 254, 284, 314, 344, 374, 404, 434, 464, 494, 524, 554, 584, 614, 644

If the internval happens to be a number with a decimal, you need to round it off into a whole number. For example, 29.4 will be rounded off into 29, and 29.5 into 30.

### *Stratified-random sampling*

Suppose there are 900 first-year students majoring in science in a national university. Among them, there are only 90 female students. Suppose you want to get a sample of 30 students in which males and females are balanced. If using systematic random sampling, the female students may be far less than 10%. In this case, you may try out another random sampling technique, that is, stratified-random sampling. First of all, you need to group students into females and males. These two groups of subjects are called subpopulations or strata. And then you extract a random sample from each subpopulation by a systematic random sampling. In the above example, you may choose 15 out of the 90 females and 15 from the 810 males.

If you want your sample of 30 students to represent the original gender distribution of the population, i.e. 10% females and 90% males. In this case, stratified-random sampling can be employed, too. First of

all, you divide the students into males and females. Then by systematic random sampling, you get 3 out of the 90 females and 27 out of the 810 males. Stratified-random sampling is widely used in applied linguistics. For example, you might want to obtain a sample of students to represent different levels of L2 proficiency, or with different family backgrounds, or with different motber tongues. Remember that in stratified-random sampling, each unit of the population must be assigned to a stratum before the random selection process begins. Therefore, it is more complicated than the other two techniques.

Subjects for a survey study are normally not less than 30. A smaller sample will cause difficulty in statistical analysis. What kind of sample do you select and how many times do you collect data from the subjects? These two questions will be answered in the following part.

### *Convenience sampling*

Despite our best efforts, it is often impossible to do strict random sampling in our actual research especially for individual researchers. Often, instead of strict random sampling techniques, we simply use convenience sampling, i.e. elements are selected for the sample for the convenience of the researcher. In other words, the researcher tends to choose subjects that are readily available. Convenience sampling is useful for exploratory research, and for pretesting questionnaires to make sure that the items are unambiguous and not too threatening. In other situations, however, convenience sampling tends to have inherent limitations.

## What needs to be considered in subject selection?

Once you move to the stage of selecting subjects, you should first remember that subject selection always depends on the other elements of the research, particularly the research questions. Therefore, whatever decisions you made for selecting subjects, you need to justify them in

relation to the whole project. Generally speaking, you have to address the following three questions:

(1) How many subjects will be involved in the study, and why?
(2) How will the subjects be selected and to what extent can the chosent subjects represent the population, and why?
(3) How many times will the subjects be involved in the study, and why?

### *A large sample size or a relatively small one?*

If the survey study is a major part of your project and the main purpose of your study is to reveal a general pattern or a tendency in L2 learning or teaching, then you have to choose a large sample. However, if the survey study is only part of your project and it is exploratory in nature, you can select a relatively small sample. A legitimate question is what is meant by saying "big" or "small". It is true that these two adjectives are used in a relative sense. Based on my understanding of the statistical requirement, the smallest sample for a survey is no less than 30 while a large sample, is no less than 500. If your study needs to compare two or three groups, then each group requires to be no less than 30.

### *A random sample or a convenience sample?*

When representiveness is a compulsory requirement for your research, you must have a random sample which may be obtained through one of the standardizing sampling techniques: simple random sampling, systematic random sampling or stratified random sampling. As a greduate student, you more often than not cannot obtain a random sample due to many practical constraints. Very often, you have to take a convenience sample "where the researcher takes advantage of an accessible situation which happens to fit the research context and purposes" (Punch, 1998: 105). For a convenience sample, you need to decide on the procedures by which the subjects are chosen. For example, is your sample obtained on a voluntary-basis or based on the subjects' L2

proficiency levels or through the teachers' recommendation? Obviously, the subjects who voluntarily join the study won't have any cooperation problem in data-collection but may be less representative because the students who volunteer are likely to be among the more motivated ones. Taking an intact class as a sample will be better than the volunteers if you want the study to be more generalizable. Furthermore, in the selection of subjects, you have to consider the demographic attributes of the subjects. For example, how many subjects are males or females? If all the subjects are the same gender, you have to justify such a decision. Some studies, for example, are interested in the age difference of learning. Then the age will be a key variable in subject-selection.

### *Longitudinal or cross-sectional*

If a study is to detect changes or reveal developmental patterns in certain variables, then the subjects need to be studied at several points of time. In this case, the subjects form a longitudinal sample. Suppose you want to find out whether the students' motivation for learning English has changed or not through tertiary education and thus you ask them to respond to the same questionnaire twice: the first time when they just start their undergraduate studies and the second time upon their graduation. The responses to the questionnaire at the two points of time are then compared. However, sometimes, the researcher cannot afford the time to follow up the same group of subjects for several years, you choose a cross-sectional sample in which the subjects who are selected from Year One through Year Four are studied only once. Take the study on motivation change for example again. Instead of studying the same group twice, you may ask a group of the first year students and a group of the fourth-year students to respond to the same questionnaire together. Then the responses given by the two groups are compared. The advantage of such a cross-sectional sample is economical. The disadvantage is that the changes detected from the two groups may be due to other differences between the two groups rather than just the tertiary education experience.

## ADMINISTERING THE QUESTIONNAIRE

Very often, we are willing to spend as much time and effort as is needed for constructing a good instrument but the same effort and the time have not been put into the actual data-collection. Obviously, constructing the instrument and administering the instrument are equally important in determining the quality of the data. This section is about practical suggestions on administering a questionnaire.

### Approaching the subjects professionally

It is very important to approach the subjects in a professional way. When the subjects are called to meet together in a classroom to get ready to answer a questionnaire, the researcher should give them a brief but clear introduction before the questionnaires are administered. Such a talk can make them feel that they are fully informed about the study and their cooperation is important and meaningful.

The introduction consists of two parts. The first part is about the background of the project and the second part about the requirement for responding to the questionnaires. In the first part, you should try to give the subjects a very clear explanation of the research purpose, the context and the way the data will be used. The research purpose, however, in some cases is a sensitive issue and must be explained in a very tactical way. Suppose the real purpose of the research is to find out the relationslrip between the use of L1 and L2 proficiency. In this case, the purpose cannot be revealed to the subjects in a frank manner since it may directly affect the subjects' responses to the questionnaire. Thus, you bave to talk about the purpose in a very general way by saying the study wants to find out how L1 functions in learning an L2. Furthermore, you should not only assure the subjects of the confidentiality of the data but also ensure that the data are in fact confidential. Finally, you need very sincerely to express your deep gratitude to them for their time and effort exerted in answering the questionnaires.

For the second part, you need to give them specific instructions. You might think it is not necessary to repeat them if the written instructions are simple and easy to understand. However, I still suggest you emphasize some important points since some students are rather careless or simply don't read the instructions. Finally don't forget to remind them that they should answer all the questions and ask them to put up their hands, if they have any difficulty in understanding the items.

If your resources permit, you had better give the subjects each a small gift as a sort of appreciation of their work. The gift of a pen or a folder, for example, can be an incentive for the subjects to answer all the items in the questionnaire seriously.

## Trying to have face-to-face administering

The way a questionnaire is administered does make a difference to the quality of data. If possible, I strongly recommend face-to-face administering instead of the mailing of questionnaires. Furthermore, I strongly suggest if possible, the researcher should administer the questionnaire personally. By doing so, you can guarantee that the subjects are professionally approached.

If your survey is a large-scale one in which the subjects are spread over several universities that are located in different cities, you usually cannot afford to go personally from one university to another. What you have to do is to ask other people for help. When this is necessary, these helpers need training so that they can administer the questionnaire in a standardized way.

## A good environment and suitable timing

Where and when a questionnaire is administered will also affect the quality of the data. For example, the quality of questionnaires answered in a classroom is different from that of those responded to in a dorm. Similarly, the quality of questionnaires responded to near the lunch time is different from that of those answered just after their breakfast.

Obviously, a quiet environment and suitable timing is important if you want the data to be reliable and valid. The best arrangement, according to my previous experiences, is to have the subjects answer the questionnaire in normal class time with the consent of their course teacher. With their teacher's presence and in normal class time, the students are usually more cooperative and more serious about the questionnaire.

## SUMMARY

A survey study usually involves a relatively large sample and employs a questionnaire with closed questions to collect quantitative data. A good questionnaire must be of high internal validity with a professional look. To design a questionnaire, you have to undertake a series of tasks: (1) deciding the mode of administration; (2) specifying what kind of data to collect; (3) determining how to process the data; (4) deciding the content of individual items; (5) choosing the question structure; (6) selecting the questionnaire wording; (7) arranging the questions in proper order; (8) deciding the questionnaire format and (9) pretesting the questionnaire. Scaling is used as a technique in a questionnaire to create a continuum upon which people's responses can be recorded in numbers and eventually can be calculated statistically. The scaling techniques can be divided into comparative and non-comparative ones. Comparative scales involve direct comparison of choices. When a comparison is made between two items, it is called paired comparison scaling; otherwise, it is called rank order scaling. In non-comparative scales, each item is scaled independently of the others with an equal interval between every two adjacent items. There are two kinds of non-comparative scales: semantic differential scale and Likert scale. The semantic differential scale is usually a five- or seven-point rating scale with end points associated with bipolar labels that have the opposite semantic meaning. A Likert scale contains a few choices which can form a continuum with every two adjacent choices having an equal interval. If the number of subjects who are asked to answer a questionnaire is less than the whole population, they

must be selected through randomization. There are three random sampling techniques: simple random sampling, systematic random sampling and stratified random sampling. When you are administering a questionnaire, you must approach the subjects professionally and have them in a comfortable environment at a suitable time.

## DISCUSSION QUESTIONS

1. What are the basic criteria for a good questionnaire?
2. Select one questionnaire from an international journal to evaluate its internal validity.
3. Briefly describe procedures of designing a questionnaire.
4. Construct questionnaire items by using different scaling techniques.
5. What should be considered in selecting subjects?
6. Decide which random sampling technique will be better used in the following studies:
   (a) Researcher A wants to divide 106 Senior Two middle school students into two random groups for a comparative study.
   (b) In 1999, 180 students were enrolled in the School of Foreign Studies at Nanjing University. Among them, 70 were majoring in English, 25 in French, 24 in German, 26 in Japanese, 20 in Russian and 15 in Spanish. How can we get a representative sample of 30 to represent the 180 students?
   (c) In one university, there are 1,000 freshmen in the science stream and 860 in the arts stream. Among the science students, 80% are males and 20%, females while among the arts students, 52% are males and 48% females. Now how can we get a random sample of 100 in which the science and arts students are evenly divided and so are the females and males?

# 7. An experimental study

Like a survey study, an experimental study is typically to collect quantitative data. Compared to the survey study, however, it is much more difficult to implement but much more powerful for establishing the cause-effect relation. In this chapter, I will start by explaining what an experimental study is, and then discuss two important concepts in an experiment: causality and validity. Next, I will describe various types of experimental studies and procedures for conducting an experiment. A paper that reports on an experimental study is presented as Appendix Two.

## WHAT IS AN EXPERIMENTAL STUDY?

Experiments were first used in the natural sciences. All of us have had the experience of undertaking either a physical or chemical experiment in a laboratory. What have left in your memory by these experiments, I am afraid, are tubes, chemicals, wires and other materials. What is an experiment in a social science? I will answer this question first by a conceptual definition and then by examples.

### Definition

An experimental study is a study in which the researcher manipulates one or more independent variables and measures their effect(s) on one or more dependent variables while controlling the effect of extraneous variables. The basic elements of the experimental study include:

(1) A treatment in which one or more independent variables are manipulated;
(2) A comparison which involves at least two groups of people or two conditions;

(3) The measurement of one or more dependent variables as the result of the treatment;
(4) Measures that are used to control for the effects of extraneous variables.

If the comparison is carried out between two groups of people, then one group is called a control group and the other one, an experiment group. These two groups are assumed to be alike in all respects except for differential exposure to treatment. The former does not receive treatment while the latter does. Once the treatment is accomplished, the dependent variable is measured. The differences that are found between the two groups can be attributed to the independent variable(s).

The essential difference between a survey study and an experimental one is that the former one is carried out in a natural surrounding but the latter one is undertaken in a human-manipulated environment. The manipulation typically involves a treatment to some subjects but not to the others, and makes the treatment meet the specified conditions. The more manipulation the researcher engages in, the more convincing the inferred causal relation is.

## Illustrations

The following are examples of experimental studies.

Example One: a comparison between two groups of people

Bejarano (1987) intended to determine whether cooperative learning approaches were more effective than traditional, whole-class learning approaches. An experiment was conducted in Israel where 665 seventh-grade students were involved. Classes were randomly assigned to using either cooperative learning approaches or "whole-class" learning approaches. After four and a half months' instruction, their performance on listening and reading comprehension was measured respectively. The results of the

comparison of the two groups' performance indicated that the cooperative learning method was more effective than the whole-class learning method for the development of listening comprehension but not for the development of reading comprehension.

In the above example, the treatment is the use of different learning designs and the result of the treatment is the students' performance on listening and reading comprehension tests. The control group is the classes that stick to the traditional approach and the experimental group is the classes that employed the cooperative learning method. Randomization is used to control the effects of the extraneous variables.

Example Two: a comparison between two conditions

In the study conducted by Pica, Young and Doughty(1987), they intended to find out whether negotiated interaction would facilitate L2 comprehension. They constructed a small-scale experiment in which two ways of providing linguistic input were compared. Altogether 16 adults at the intermediate level were involved. Their comprehension was checked by placing objects on a board under two conditions. In the first condition, a native speaker of English read instructions that were modified by simplifying syntactic structures, using more words to paraphrase each instruction and reading each content word several times. While listening, the students were not allowed to interact with the native speaker. In the second condition, the native speaker read an instruction without any modification. However, the subjects were encouraged to communicate with the native speaker to obtain comprehension. It was observed that the input was modified through interaction. The results showed that the subjects place more objects correctly on the board in the second condition than in the first condition. It

is concluded from the results that negotiated interaction can facilitate L2 comprehension.

In the above example, a treatment involves two ways of providing linguistic input that was, however, tried out on the same group of people. The results of the treatment were measured by placing objects on a board according to instructions. A comparison was made between the results produced by two types of treatment. Many extraneous variables were controlled since only one group of people were involved.

## CAUSALITY

The concept of causality is very essential for experimental studies. However, this concept can be interpreted differently. Therefore, the following section will first discuss the meaning of causality and then the conditions for causality.

### The meaning of causality

A statement such as "X causes Y" is interpreted by scientists as:

X is only one of many possible causes of Y;
X can account for the variance in Y;
The cause-effect relation between X and Y can never be proved but inferred.

The above interpretations are different from our daily understanding in a number of ways (Malhotra, 1993). An ordinary person tends to think of a causal relation in an absolute and deterministic sense. The statement "X causes Y" would mean that X is the sole cause that can determine the occurrence of Y and such a causal relation can be observed and testified.

Evidently, the scientific interpretations of causality are more appropriate to applied linguistic research. L2 teaching and learning is

complex and it is bound to be linked with multiple variables. The cause-effect relationships existing in the field of L2 teaching and learning tend to be probabilistic rather than deterministic. Moreover, we can never directly perceive or verify causality. What we can do at most is to infer a cause-and-effect relationship. Consequently, it is possible that the genuine causal relation, if it exists, may not have been identified. The following section will further clarify the concept of causality by identifying conditions for a cause-effect relation.

## Conditions for causality

Three identifying conditions are generally proposed, i.e. temporal precedence, necessary connection and the absence of spuriousness within the cause-effect relationship (Maxim, 1999; Punch, 1998). The three conditions should be present simultaneously if a causal relation is to be proposed since none of them alone is enough alone to define causality.

Temporal precedence is regarded as essential for the occurrence of causality. It means that X must occur before Y but not the other way around if we want to claim that X is a cause of Y. This view seems to be clear enough and can account for many instances of causality. For example, language aptitude can account for the differences in L2 achievement; L1 proficiency can affect L2 performance. Temporally speaking, language aptitude and L1 proficiency evidently occur before L2 learning. However, this condition is easily challenged. For example, summer always follows spring, and winter always follows autumn, yet we don't want to say that spring is a cause of summer, autumn is a cause of winter. Therefore, the temporal precedence alone does not seem to be enough to define causation.

The second identifying condition is the necessary connection. It means that the two variables concerned must show a necessary link if they form a causal relation. In other words, the changes in X should be related to the changes in Y. Or we say that the two variables must be covaried. The problem with this condition is that many co-varied

variables possess reciprocal relations in which the cause and the effect are not easily teased out. For example, L2 motivation and L2 achievement are such a pair of variables. However, if we can establish the time order between these two co-varied variables, we are in a better position to identify causality since logically speaking, something that occurs later cannot become a cause of something that happened earlier. Even if the above two conditions are satisfied, we still cannot conclusively establish a causal relation since it is possible that variables other than the ones under investigation are the real cause that however has not been noticed.

The third requirement is that other plausible causes can be ruled out. That is to say, the researcher should be able to explain why causes other than the one(s) proposed are not possible. There are two ways to eliminate the plausible causes. One is called physical control in which the experimental group and the control group are divided through randomization on the assumption that the differences between the two groups are not systematic and they can cancel each other out. The second way is called statistical control by which the effects of the extraneous variables are removed in the process of data analysis. However, the justifiable explanations are not always easy to provide since phenomena in applied linguistics are complex in nature. This topic will be further discussed when we deal with the measures for controlling extraneous variables.

## VALIDITY IN EXPERIMENTATION

The basic criterion for evaluating an experimental study is validity. What is validity in an experiment? What factors may affect validity? How do we ensure high validity in experimentation? These questions will be dealt with in the following part.

### Internal and external validity

There are two kinds of validity in experimentation: internal validity and external validity. The internal validity concerns about the question "To

what extent can the claimed cause-effect relation be justified?" In an experiment, if the researcher succeeds in constructing an environment where only the independent and dependent variable(s) are working while the other variables are controlled, the changes in the dependent variable can thus be attributable to the treatment only. Such an experimental design shows high internal validity. In other words, it ensures that the independent variable(s) in question in fact induce the effects in the dependent variable rather than some other variables that are not investigated. Those variables that are not targeted for investigation are called extraneous variables or intervening variables. Controlling the extraneous variables is a key to success in achieving high internal validity.

External validity is the question of the generalizability of research findings. Bracht and Glass (1968) indicate two major aspects of generalizability: people and environment. The question for "people validity" is "How far can the findings be applied to other populations?" The question for "ecological validity" is "To what extent are the findings transferable to other settings or contexts?"

## Factors affecting internal validity

The factors which can influence the internal validity of research are varied. Among them, some are related to the environment, some related to subjects themselves, some related to the way variables are measured, some related to the selection of subjects and some related to experimental conditions. In the following section, each type of variable will be described.

### *Factors related to the environment*

Environmental factors refer to the factors occurring in the environment when the research is progressing. For example, Brown (1988) once conducted research in China where part of the study was to compare two groups' performance on a listening comprehension test. The

first group took the test with windows open, but it was reasonably quiet outside. However, when a second comparison group was tested several hours later, there arose a big noise produced by diesel tractors outside. Even with the windows closed, the noise was still disturbing. Obviously, any comparison of these two groups would be meaningless. In addition to noise, the variables such as temperature, time of day, adequacy of light, ventilation, comfort of seats, etc. are also important environmental factors that should be considered by the researcher when designing an experiment. Otherwise, they will function together with independent variables to confound the results.

### *Factors related to subjects*

The important factors related to subjects include maturation, mortality and Hawthorne effect.

Maturation refers to changes in the subjects themselves that are, however, not caused by the impact of independent variables but by natural changes over time. In L2 learning, maturation takes place as L2 learners become older and more experienced. Such experiences might include other simultaneous learning, or psychological and physical changes. In other words, their L2 proficiency will be improved as their learning experiences accumulate. If you undertake a study that lasts for several months or several years, you should keep maturation in mind when you interpret the results.

Mortality refers to the loss of subjects while the experiment is in progress. This happens for many reasons. For example, the subjects refuse to continue the experiment because the experiment is boring, time-consuming, or too demanding, or it is held in a time that is not convenient to the subjects. Mortality confounds results because it is difficult to determine if the lost subjects would respond in the same manner to the treatments as the remaining ones.

The Hawthorne effect was first found when Mayo, Roethlisberger, and Dickson carriad out their study at the Hawthorne branch (Chicago) of

the Western Electric Company (as reported in Brown 1954). They noted that whenever they were present, the productivity was increased although working conditions were varied from time to time. In social science, the Hawthorne effect refers to a situation where the subjects are so pleased to be chosen for the study that they behave better than usual. In this case, the findings may be related more closely to this pleasure than to the independent variables you intend to investigate. Consider, for instance, a hypothetical study of the effectiveness of a task-based approach in which you tell the students that they are selected for an experiment in order to see whether this new method is better than the traditional one. At the end of the study, you find the students in the experimental group outperform the students in the control group. It is not clear whether the results are due to the greater effectiveness of the method or to the Hawthorne effect. Consequently, this will pose a problem: how do we separate the Hawthorne effect from the effects of the task-based approach?

### *Factors related to measurement*

Another set of factors which might affect the internal validity of a study is related to measurement. For example, we use tests, questionnaires, interviews or any other means to measure one or more variables. The way we measure the variables may affect the internal validity through testing effects and the instability of the measures.

Testing effects occur in the process of data-collection. They may happen in various ways. For example, if the time interval between pre-test and post-test is too short, the subjects may remember the answers they gave in the pre-test. Some of the subjects might try to select the same responses in the post-test in order to show their consistent views, while others might try to select different answers in the post-test when they think they have figured out what the study is about and try to "help" the researcher. In this way, the validity of the responses in the post-test is affected.

Secondly, testing effects may happen in a situation where the items in the pre-test might make the subjects become aware of the variable you intend to investigate. As a result, their sensitivity to the variable will lead to better learning results. For example, once Brown (1988) attempted to examine the effectiveness of teaching reduced forms. Two equivalent groups took a pre-test on reduced forms before the teaching started. Then, one group (the experimental group) was explicitly taught reduced forms while the other (the control group) was not. At the end of the study, both groups were again tested on reduced forms. It was not surprising to find that the experimental group outperformed the control group in the second test. However, he was not certain whether the high scores obtained by the experimental group were also affected by the subjects' high degree of awareness of the reduced forms through the pre-test.

Thirdly, testing effects may be found when the same test of language proficiency is administered repeatedly to determine how much progress students have made in their performance through a certain period of teaching. The progress in the second test may not be solely caused by teaching. For example, if the same grammar test is given to students twice, some of the problems occurring in the first test may be cleared up in the second test simply because the students learn from their mistakes. Thus, the results are ambiguous and difficult to interpret.

In addition to testing effects, inconsistency or instability of instrumentation is another factor related to measurement. Consider two different tests which are administered to the same group of students. One test is used before the experiment and the other one after the experiment for the purpose of checking whether there is any progress in L2 performance. If the second test is actually much more difficult than the first one, then it is not clear whether the differences in the scores on the two tests are a real failure of the experiment or are simply due to the different degree of difficulty of the two tests.

### *Factors related to the selection of subjects*

Last but not least, subject selection may influence the results of research studies. Suppose you selected two groups to investigate the effectiveness of two teaching designs: the task-based approach and the traditional method. Actually they were not equal at the beginning of the study since one group had already received instruction in that teaching method for quite a long time. Thus the differences between the two groups' performance on the post-test might not be caused by the instruction but by preexisting differences. Sometimes, the subjects selected are volunteers who are most likely better motivated in learning than those who do not want to participate in the experiment. The findings are then ambiguous because they might be also affected by the subjects' strong motivation.

### *Factors related to a treatment*

The last set of factors that may affect the internal validity of an experimental study is related to treatment. For example, how long does the treatment last? Suppose you intended to measure the effects of the peer-correction on the development of L2 writing. Your treatment lasted four weeks and the result was that there the experimental group did not do better than the control group in their post-writing task. Based on such a result, you concluded that the peer-correction is not more effective than the teacher-correction. However, it can be argued that the ability of L2 writing develops slowly and gradually, and the drastic improvement in L2 writing cannot be made within the four weeks. Along this line of argument, the internal validity of your study is called into question.

In addition to the length of treatment, the way the treatment is given can also threaten the internal validity of an experiment. For example, you wanted to find out whether students' annotating their compositions is an effective method in improving their L2 writing ability. The experiment lasted one year. These two groups of students were taught in the same way except for one difference. That is, the experimental group was asked

to write each composition with their annotations while the control group was asked to write each composition without annotations. The result was that the experimental group did outperform the control group in their post-test. Therefore, you attributed the superior performance of the experimental group to the method of annotations. However, the students in the experimental group actually spent more time on each composition than the control group because it took time for the experimental group to annotate their compositions. By adding all the time spending in annotating in a year, we would immediately find that the experimental group had spent much more time in learning L2 writing than the control group. It is thus uncertain whether the better performance produced by the experimental group was caused by the method of annotations or by having more learning time or both.

Finally, the internal validity of an experiment can also be affected by the people who implemented the treatment. Suppose you asked two teachers to carry out your experimental study, and one teacher was teaching the experimental group while the other the control group. How about these two teachers? Did they have similar teaching experiences? Did their L2 proficiency reach the same level? Consider that the experiment where a more experienced teacher taught the experimental group and the less experienced teacher the control group. The final result was that the experimental group did better than the control group. Obviously, it is not sure whether the superior performance of the experimental group was the effect of the experiment or due to the teacher's rich teaching experiences.

## Factors affecting external validity

When you run an experiment, you hope that the results will be generalisable to other students or settings. However, factors related to the environment and the subjects may affect generalizability.

Suppose you are investigating the effectiveness of a particular method of teaching /r/ vs. /l/. If you conduct the study in a language

laboratory with highly sophisticated equipment and tightly controlled procedures, you won't be able to interpret the results of your study in terms of teaching those items in an ordinary classroom. The reason is that the setting in which you carried out the research is not that of the real world.

Selection of subjects is another factor influencing external validity. For example, if you want to find out about English reading strategies of Chinese middle school students, you should not select the students from a foreign languages' school as a sample. If you want to be able to generalize your findings, you need to choose your sample carefully.

## Relationship between two types of validity

Internal and external validity are extremely important if you hope your results will be useful to you and to others in the same research field. However, there is a trade-off between maximizing internal and external validity. In order to have the most valid results, you restrict the experimental procedures as carefully as possible, often to laboratory procedures. Consequently, the results are not generalisable to the real classroom situations. Similarly, maximizing external validity will be at the expense of reducing internal validity. Once you conduct your experiment in a natural environment to enhance external validity, many extraneous variables coexist with the variables you are examining. Thus, the internal validity is threatened. What you are expected to do is keep the best balance you can, selecting procedures that will maximize both types of validity. However, in reality, it is not easy to balance the two. I suggest that you give priority to internal validity since without internal validity, external validity will be meaningless.

## Controlling extraneous variables

Extraneous variables represent alternative explanations of experimental results. They pose a serious threat to the internal and external validity of an experiment. Unless they are controlled, they affect the dependent

variable and thus confound the results. For this reason, they are also called confounding variables. There are two general ways of controlling extraneous variables: physical ways and statistical ways (Punch, 1998). In physical control, variables are controlled in the process of designing an experiment while in statistical control, variables are controlled in the process of analyzing data. Table 7.1 shows various strategies (Punch, 1998: 83).

| Physical control | Statistical control |
|---|---|
| Randomization | Stratification |
| Restriction | Partial correlation |
| Matching | Analysis of covariance |

Table 7.1: Strategies for controlling extraneous variables

### *Physical control*

If you want to control for extraneous variables in designing an experiment, you employ any one of the three strategies: randomization, restriction and matching.

Randomization refers to the random assignment of subjects to an experimental group and a control group by using one of the randomization techniques. This technique does not guarantee that the control group and the experimental group are identical but it does make sure that extraneous factors are represented non-systematically in the two groups. Their effects are assumed to cancel each other out. Therefore, randomization is the preferred procedure for ensuring the prior equality between the control and experimental groups. However, randomization may not be effective when the sample size is small because randomization merely produces groups that are equal on average. It is possible, though, to check whether randomization has been effective by measuring the possible extraneous variables and comparing them across the groups.

Restriction refers to the avoidance of the variance of one or more variables by keeping the variables concerned constant. Keeping a variable constant means that the variable has no variance in the

experiment. For example, if the researcher thinks that gender may affect the dependent variable as an extraneous variable, s/he may select the subjects who are the same gender to participate in the experiment. Thus, gender cannot have any relation with the dependent variable. However, by keeping gender constant, the research findings are not generalizable to both sexes. Furthermore, keeping one variable constant does not guarantee the findings are not affected by other extraneous variables.

Matching involves comparing subjects on a set of key background variables before assigning them to treatment. For example, in a study on the effectiveness of teaching reading strategies, two groups could be matched on the basis of motivation, L2 reading proficiency and gender. Then one group would be assigned to be an experimental group and the other to be a control group. Matching has two obvious weaknesses. First, subjects can be matched on only a few characteristics. As a result, the subjects may be similar on the variables selected but unequal on others. Second, if the matched characteristics are irrelevant to the dependent variable, then the matching effort has been wasted.

### *Statistical control*

Instead of making the differences between the groups non-systematic or keeping the variance constant, we may allow natural groups to participate in experiments. In this way, two groups may have systematic differences before the experiment. However, if we are aware of such differences, in the analysis of the data, we can use some strategies to measure the differences. We can then adjust for their effects through statistical adjustments. In other words, we may remove the effects of extraneous variables from the overall effects of the independent variable on the dependent variable. This process is called statistical control and is an alternative to physical control for extraneous variables. Specifically speaking, we may employ the following three strategies to exercise statistical control:

(1) Stratification (Rosenberg, 1968), where one extraneous variable is selected as a moderator variable and then the analysis is carried out in groups partitioned according to the moderator variable. For example, if you think the level of L2 proficiency is a possible extraneous variable, you select it as a moderator variable. The subjects are then separated according to their levels of L2 proficiency. Further analyses are carried out within the same level of L2 proficiency. It is no longer possible that the resulting relationships from such analyses between the independent and dependent variables are influenced by the level of L2 proficiency since all the subjects within each group are at the same level.

(2) Partial correlation where the effects of extraneous variables are partialled out from the overall effects of the independent variable on the dependent variable. For example, there are three variables involved in your study: Motivation, Learning strategies and L2 learning outcomes. You intend to find to what extent learning strategies can affect L2 learning outcomes. In this case, you believe that Motivation can work together with learning strategies to influence L2 learning outcomes. In order to tease out the effects of Motivation, you may carry out a partial correlation between Learning strategies and L2 Learning outcomes in which Motivation is controlled.

(3) The analysis of covariance where the effects of covariate(s) are removed from the dependent variable before mainstream analysis. Suppose your experiment was intended to find out the effectiveness of reading strategy training. The students who comprise a control group and an experimental group are two natural intact classes. The control group was having a reading lesson as usual while the treatment group was taught reading strategies in addition to the normal reading activities. They had a pre- and post-tests of reading comprehension. Since these two groups were not formed by randomization, they had significant differences in their pre-test scores

before the experiment. In order to control the extraneous variable of previous reading ability, their reading comprehension test scores were adjusted by removing the differences shown in their pretest reading test scores. The formal comparison was made between the two groups after this adjustment.

The above three strategies by nature are the same in the sense that they all aim at removing the effects of extraneous variables from the dependent variable. Therefore, the rival hypotheses or alternative explanations to the preferred causal relation are eliminated. The assumption in doing this is that through statistical control, we are able to approximate the ideal conditions for demonstrating causality.

## TYPES OF EXPERIMENTAL STUDIES

Experimental studies are classified as pre-experimental, quasi-experimental and true experimental, and can be placed on a continuum in terms of the degree of the researcher's control. The pre-experimental studies are least controlled and the true-experimental studies, most controlled. The critical feature that separates these three types of experimental studies is the extent to which comparison groups to look alike in all respects.

| | Pre-experimental | Quasi-experimental | True-experimental |
|---|---|---|---|
| Various designs | One-group posttest only study<br>X $O_1$ | Posttest only nonequivalent groups<br>EG: X $O_1$<br>$O_2$ | Posttest only equivalent groups<br>EG: R X $O_1$<br>CG: $O_2$ |
| | One-group pretest-posttest study<br>$O_1$ X $O_2$ | Pretest/Posttest nonequivalent groups<br>EG: $O_1$ X $O_2$<br>CG: $O_3$ $O_4$ | Pretest/posttest equivalent groups<br>EG: R $O_1$ X $O_2$<br>CG: R $O_3$ $O_4$ |
| Different features | No control group<br>No randomization | Having a control group;<br>No randomization | Having a control group.<br>Using randomization |

Table 7.2: Different types of experimental studies

The best strategy to make the comparison groups similar is random assignment of subjects. True-experimental studies are characterized by randomization while randomization is absent in the other two types of experimental studies. Quasi-experimental ones involve naturally occurring treatment groups that are fairly clear-cut. In this sense, although randomization is not present, the differences between the two groups can be identified, measured and eventually their effects can be extracted statistically. Such a statistical control may enable the treatment conditions to approximate the ideal experimental conditions. Pre-experimental studies are less controlled than quasi-experimental studies since there are no two identifiable groups to compare as such.

Table 7.2 shows three different types of experimental studies with their different features. In the following section, we will describe them one by one.

## Pre-experimental studies

Pre-experimental studies are characterized by an absence of randomization and of a control group. Two specific designs are described: the one-shot case study and the one-group pretest-posttest design.

### *One-shot case study*

The one-shot case study may be symbolically defined as

$X \quad O_1$

(where X represents a treatment and $O_1$ a single measurement on the dependent variable)

Example: Is a new method better than the traditional one?

One researcher hypothesized that a new method was better than the traditional one to improve scores on TOEFL. She ran a one-year training course for the TOEFL candidates for five years by using

the traditional method while she recorded the students' TOEFL scores accordingly. The yearly average scores fluctuated from 540 to 575 with the mean being 567. This year she used a new method to teach her students. Once the treatment was accomplished, all the students in her class took TOEFL and their average was 585. She compared their average scores this year with the students' average scores in the past five years. The results showed that this group did outperform the students in the past (A hypothetical study).

This type of experiment is useful when it is possible to compare a sample statistic with a known value or population parameter. By such a comparison, it is sufficient to determine whether the group under consideration differs from some established values. One may, for example, suspect that the students in a particular class are outstanding on their examination performance. This hypothesis can be examined by comparing the students' test scores with historical records.

The value or applicability of this design is limited, however, since it is useful only when the established value is available and the dependent variable can be measured by an authoritative test. If these two conditions are absent, the weaknesses in experiments of this type can be easily demonstrated. First of all, it is not possible to compare the level of the dependent variable to what would happen when X was absent. Secondly, the level of the dependent variable might be influenced by many extraneous variables such as History, Maturation, Mortality and Selection. Lack of control for these extraneous variables will pose a serious threat to the internal validity. For these reasons, the one-shot case study is more appropriate for exploratoty than for formal research.

### *One-Group Pretest-Posttest Design*

The one-group pretest-posttest design may be symbolized as

$0_1$ X $0_2$

(where X represents a treatment and, $0_1$ the first measurement and $0_2$ the second measurement on the dependent variable )

Example: Is reading strategy training effective?

To find out whether reading strategy training can improve students' reading comprehension ability, Researcher A decided to carry out an experiment in her class. She measured the students' reading comprehension before the training started. After 10-week training, they took another reading comprehension test what was equally difficult as the pre-test. The two test means were then compared. It was found that the students performed much better on the post-test than on the pre-test (A hypothetical study).

In this design, one group is measured twice without a control group for comparison. A pretreatment measure is made ($0_1$) before the group is exposed to the treatment (X). A post-treatment measure is made ($0_2$) when the treatment is finished. The treatment effect is calculated as $0_2$-$0_1$. Such a design can work well in natural sciences where inanimate objects remain stable over time, but it is questionable in social sciences since extraneous variables such as History, Maturation, Testing (both main and interactive testing effects), Instrumentation, Selection, and Mortality could possibly be present. Furthermore, the longer the time between $0_2$ and $0_1$, the greater the danger of having confounding effects on the dependent variable.

## True experimental etudies

The distinguishing feature of true experimental studies is the use of randomization.

### *Posttest only equivalent groups*

The posttest-only equivalent groups design and the posttest only nonequivalent groups design share one common feature, i.e. the absence of pre-measurement. The difference between them is that the former employs randomization but the latter does not. Formally, it may be described as

| | | | |
|---|---|---|---|
| Experimental group: | R | X | $0_1$ |
| Control group: | R | | $0_2$ |

(where R stands for randomization, X, a treatment; $0_1$, a single measurement on the experimental group and $0_2$, the same measurement on the control group).

Example: Is multiple revision more effective than writing multiple compositions?

The hypothesis that was tested in the experiment was that multiple revision was more effective than writing maltiple compositions to develop L2 writing skills. Two randomly assigned classes participated in the experiment in which one class was required to do multiple revision and another class, to do multiple writing. Specifically speaking, the students in the first class wrote six compositions with each revised three times while the students in the second class wrote 18 different compositions without revising. With the exception of the difference in this requirement, the writing lessons were the same for the two classes. At the end of the semester, the students in the two classes were measured on their ability in writing in English. Against the researcher's expectation, these two classes did not show any significant differences in their writing ability (A hypothetical study).

In this design, the treatment effects are calculated as $0_1$-$0_2$ , i.e. the results on Test A of the experimental group minus the results on Test A of the control group. Although Random assignment of group members

allows the strong assumption that the two groups are identical on all relevant features before the experiment starts. This identity is assumed and cannot be further checked in the absence of pre-measurement. Therefore, this design is still sensitive to selection bias, mortality, Hawthorn effects. Apart from its limitations, this design possesses significant advantages in terms of time, cost, and sample size requirements. It involves two groups and one measurement per group.

### *Pretest-posttest equivalent group design*

In the pretest-posttest equivalent groups design, subjects are randomly assigned to either the experimental or the control group and a pre-measurement is taken on each group. This design is formally described as:

| | | | | |
|---|---|---|---|---|
| Experimental group: | R | $0_1$ | X | $0_2$ |
| Control group: | R | $0_3$ | | $0_4$ |

(where R stands for randomization, X, for a treatment; $0_1$ and $0_3$, for a pre-measurement on the experimental and on the control groups respectively; $0_2$ and $0_4$ for a post-measurement on the experimental and control groups respectively).

In this design, the assumption of random assignment is logically justified if $0_1$ and $0_3$ are statistically equivalent, and it is not possible to have any systematic bias. While, in reality, such an assumption can hardly be perfectly realized, the use of pre-measurement is a good solution to further control for any differences between the two groups that may exist before the experiment. Suppose we have modified the experiment on the effectiveness of multiple revision just mentioned by measuring the writing abilities of the two groups before the treatment. The pretest and posttest scores are the following:

Pretest: $0_1 = 75; 0_3 = 80$
Posttest: $0_2 = 85; 0_4 = 85$

Without the pretest scores, no experimental effects ($0_2 — 0_4 = 85—85 = 0$) could be found. In contrast, by having the pretest scores, the experimental effects could be obtained as:

$$(0_2—0_1)— (0_4—0_3) = (85—75)—(85—80) = 5$$

## Quasi-experimental studies

What separates quasi-experimental studies from pre-experimental studies is the presence of a control group, and from true-experimental studies is the lack of random assignment of group members.

### *Posttest only nonequivalent groups design*

This design involves two groups: the experimental group and the control group. Measurements on both groups are made only after the treatment, and subjects are self-selected or selected arbitrarily by the researcher. This design may be symbolically described as:

| | | |
|---|---|---|
| Experimental group: | X | $0_1$ |
| Control group: | | $0_2$ |

(where X stands for a treatment, and $0_1$, for a single measurement on the experimental group and $0_2$, for the same measurement on the control group.)

The treatment effect would be computed as $0_1$-$0_2$. However, this treatment effect could also be attributed to a series of extraneous variables since without randomization, the two groups may differ before the treatment, and selection bias may be present.

### *Pretest/Posttest nonequivalent groups design*

Compared with the posttest nonequivalent groups design, this design requires pre-measurement. It can be symbolically described as:

| | | | |
|---|---|---|---|
| Experimental group: | $0_1$ | X | $0_2$ |
| Control group: | $0_3$ | | $0_4$ |

(where X stands for a treatment, $0_1$ and $0_3$ for a measurement before the treatment on the experimental group and on the control group respectively, and $0_2$ and $0_4$ for a measurement after the treatment on the two groups respectively.)

The experimental effect is measured as: $(0_2—0_1)— (0_4—0_3)$. In this way, the differences between the two groups before the treatment can be taken into consideration to a certain extent. However, one pretest score cannot crystallize all the differences between the two groups before the treatment. Suppose the treatment is gender-sensitive, or personality-sensitive. Obviously, the pretest score does not reflect such difference. Therefore, without randomization, the confounding effects cannot be fully controlled.

Quasi-experimental studies are inevitably a second-best choice to test hypotheses and should be used only where true experimental studies are clearly impossible to carry out.

## PROCEDURES FOR AN EXPERIMENTAL STUDY

If your research questions need to be answered in an experimental study, what procedures should you follow? Logically speaking, the researcher should make a plan for an experiment first and then implement it. In the stage of planning, s/he has to make a series of decisions concerning subject-selection, pretest, treatment and post-test. With the plan fully justified, the researcher can start implementing the experiment in which monitoring, modification and recording are the important tasks. In the following section, I will describe these two stages one by one.

### Planning an experiment

At the stage of planning an experiment, the researcher needs to answer a series of questions as follows:

(a) What is the number of subjects and how are they selected?

(b) How is the pre-test administered if there is any?

(c) How is the treatment implemented?

(d) How is the post-test administered?

We will use a hypothetical study as an example in our later discussion to illustrate how to answer the above questions. The bypothetical study runs like this: The researcher intends to find out whether L2 writing taught in a computer laboratory is more effective than one in a traditional classroom. She decides to conduct an experiment to compare the effects of these two ways of teacbing writing.

### *Subject-selection*

To have an experiment, you must decide the number of subjects and the number of comparison groups. The ideal number of subjects within one group should be no less than 30 according to the statistical requirement and the number of comparison groups, no less than two. Where can you get these subjects? Should you randomly select them from all the students, or from one level of students? Or should you simply take intact classes? If an experiment is carried out in normal teaching time, random assignment of subjects may disturb the normal teaching schedule, or may not be possible at all since the English lessons of different classes are usually not at the same time. Therefore, randomization is very often not a feasible decision in real life although it is the best solution for controlling confounding effects. In this case, what you can do is to take naturally occurring classes by matching a few important features. In this case, the comparison groups should be similar in L2 writing ability, L2 overall performance and major (in science or arts). Altogether two or more classes are involved as shown in Table 7.3.

| Control group | Experimental group |
|---|---|
| One arts class (30) | One arts class (30) |
| | *One arts class (30)(optional)* |

Table 7.3: Classes for the control group and the experimental group

Suppose you think science and arts students may have different gains from computer-assisted learning since science students are more familiar with computer. You may add one more dimension for comparison, i.e. comparing two more classes: one science class as a control group and one science class as an experimental group.

If your experiment is conducted in the summer holidays when there is no requirement to follow the national syllabus, teachers have more freedom to assign the students randomly into comparison groups.

***Pre-test***

The purpose of a pre-measurement is to identify differences in comparison groups prior to the introduction of intervention. Thus, the measurement should cover the essential differences that may confound the effects of treatment. In the case mentioned above, you need at least to test the students' overall English proficiency, including English writing ability. If an authoritative test already exists, you had better use it since it can avoid the trouble of justifying the validity and reliability of the test. If subjects are university students, the College English Test for non-English majors and the English Proficiency Test for English Majors are the best tests available in China.

The conditions for a pretest should be the same. Suppose one class has the test in the morning and the others in the afternoon, or one class in the first period in the morning and the others in the last period in the morning. Such variations could lead to differences in the students' performance on the test.

How pretest papers are scored could be another factor that affects the internal validity of the experiment. For multiple-choice questions that have only one correct answer, fair scoring should be beyond question. However, scoring a composition involves subjective judgement. To increase the reliability of scoring, you had better take the following measures:

(1) At least two raters score each paper independently.
(2) The compositions of the control class and of the experimental class(es) should be mixed up so that the raters cannot show bias in marking the students.
(3) A marking scheme should be provided so that raters can follow the same criteria to evaluate a composition.

If comparison groups are formed by strict randomization, a pre-test can be avoided. However, by having a pre-measurement, you have an opportunity to further check whether there is any difference between the compared groups. If you can show that there is no significant difference between groups on the pretest, differences obtained in the posttest gain far more credibility.

### *Treatment*

To design a treatment, you have to consider many things. Let's look at the example of the L2 writing experiment again. You should decide how an L2 writing lesson will be taught to an experimental group and how it will be taught to a control group. Should they finish the same number of take-home assignments? Should these two groups be taught by the same teacher or two different teachers? Do the two groups have the writing lesson under similar physical conditions (e.g. the time schedule, the environment)? No matter what kind of decision you will make, you have to notice that differences in learning content, the number of take-home assignments and the variations in teachers may all influence the changes in the dependent variable. If insufficient attention is paid to these differences, the final result is more difficult to interpret. For example, it could be questionable whether the difference in the two groups is induced by the use of computer-assisted learning or by other extraneous variables.

The issue of controlling for extraneous variables could be rather complicated when you come to make decisions. For example, if the same

teacher is asked to be responsible for the two groups, the apparent advantage is that there seems to be no variation in teachers. However, in this case, it may not be certain whether this teacher has any bias towards either way of teaching. Even without any bias at the conceptual level, it is difficult to guarantee that s/he can manage the two ways of teaching equally skillfully. After considering the problems, you will most likely decide against having one teacher for the two groups. What can you do now? The only alternative is to choose two teachers, each responsible for one class. It is usually difficult to find two teachers who are similar in all respects. The logical solution would be to choose two teachers each considered to be an expert in using their own teaching methods.

Another important thing you have to decide is how long the treatment will last. Some treatments can produce immediate effects but some can only produce delayed effects. Obviously, the improvement of the students' L2 writing ability cannot be achieved within a short period of time. Identifiable effects on L2 writing can be shown, according to our experiences, only after at least one semester's treatment.

### *Post-measurement*

Pre-measurement is optional but post-measurement is obligatory for all experimental studies. Planning a post-test in the study where a pre-test is absent is simpler than otherwise since the researcher does not need to consider testing effects. How do you measure the effects of measurement? Let us look again at the case of the experiment to compare the effectiveness of two ways of teaching L2 writing. Suppose that in the pre-test, the subjects were asked to write a composition based on the given topic "Should private cars be promoted in China?" Do you want to ask the students to write a composition on the same topic used in the pre-test or a different one? Using the same topic can eliminate variations in difficulty of writing tasks. The danger is that some students who have a good memory may simply write down what has been retrieved from their memory. Using a different topic, you have to make sure it is more or less

similar to the one used in the pre-test in terms of writing difficulty. The common practice is to use a different topic which is at the same degree of difficulty level as the previous one.

The scoring procedures, no doubt, should be the same as those followed in pre-measurement. Ideally, the raters who scored the pre-test papers also scored the post-test papers. This can increase comparability between the two tests. If possible, the pre-test and post-test papers should be scored together. This arrangement can increase consistency in scoring the two sets of test papers.

## Implementing an experiment

The more careful and meticulous your plan is, the fewer problems you may come across when you implement an experiment. However, no matter how perfect your plan looks, you cannot avoid dealing with unexpected problems since situations change. Furthermore, no plan can be perfect and it is common that you find out flaws in the plan afterwards. Therefore, making modifications in the process of implementation is normal rather than exceptional.

The difficulty in the process of selecting subjects is often related either to the size of the sample or to the assignment of the subjects into comparison groups. For example, you may find the two intact classes show great differences in the total number of students: one class with 20 students and the other one with 30 students. If you change one class into another one, you will have a problem in matching two teachers. In this case, you had better give up the idea of having a new class since matching two teachers is even more difficult. If possible, you had better try out a new method on a bigger class rather than on a smaller one. Then, in the process of analysis, you can use the strategy of randomization to make the two classes have the same number of students.

The pre-test should be taken before actual teaching starts. Furthermore, the pre-test should be beld for the two groups at the same time. By doing so, you want to avoid the case where information about

the test will leak before the second group takes it. The difficulty is that the actual teaching schedule is often not arranged according to your own needs. Then, you have to rearrange the class schedule before the pre-test. Meanwhile you should inform the students of such a change. Otherwise, a substantial number of students may be absent in the test simply due to being uninformed. The quasi-experiment cannot afford the absence of the pretest scores. Another problem you might come across is that the students may not be serious about the pre- and post-tests. To avoid this problem, you had better take the scores on the test as part of the final score on the course.

Once the treatment starts, the researcher should keep a close eye on progress. If possible, s/he should sit in on the classes and take field notes. Through such observation, the researcher may gain a better understanding of what is happening in the treatment. Furthermore, s/he could detect problems and make necessary corrections if the treatment does not strictly follow the original plan either due to an incomplete understanding of the plan or due to the occurrence of some unexpected difficulties.

In regards to the post-test, the researcher has to deal with the same problems as those in the pre-test. First of all, the experimental group and the control group should take the test at the same time. Secondly, the students' absence from the test should be prevented. In addition to the above requirements which control potential differences between the two groups, the post-test and the pre-test should be taken in similar environment. By doing so, the possibility of attributing differences between the two groups to factors other than the treatment can be reduced.

## SUMMARY

An experimental study is a study in which the researcher manipulates one or more independent variables and measures their effect(s) on one or more dependent variables while controlling for the effect of extraneous

variables for the purpose of establishing a causal relation. The researcher's manipulation is typically in the form of a treatment in which two or more comparisons are made to measure the power of the hypothesized cause on the assumed effect. The meaning of the term causality used in an experimental study is different from our daily understanding. It is characterized by probability. There are two kinds of validity in experimentation: internal validity and external validity. Both are extremely important for an experiment. The former validity concerns the question "to what extent can the claimed cause-effect relation be accounted for by the variables investigated?" The latter validity concerns "to what extent can the research findings be applied to contexts beyond the group investigated?". The factors that can reduce internal validity are various; some factors related to the environment, some to measurement, some to the selection of subjects, some to the treatment. External validity can also be affected by a variety of factors related to the environment and the selection of subjects. In order to increase the validity of an experimental study, you may take various measures to control for the effects of extraneous variables. They include physical control and statistical control. Experimental studies are classified as pre-experimental, quasi-experimental and true-experimental in terms of the degree of the researcher's control. The true experimental study is characterized by two features: having a control group and randomly assigning the subjects into control and experimental groups. In the case of the quasi-experimental study, randomization is absent although there is a control group. A pre-experimental study has the least control by the researcher. To implement an experiment, the researcher needs to consider with great caution the number of subjects and the way of selection, the format in which the pretest and posttest are carried out and the manner in which the treatment is given.

## DISCUSSION QUESTIONS

1. What are the differences between a survey study and an experimental study?
2. Under what conditions can we claim a causal relation between two or more variables?
3. What factors may threaten the validity of an experiment?
4. Find an experimental study from an international journal and evaluate it in terms of internal and external validity.
5. Use hypothetical examples to illustrate three different types of experimental studies.
6. When you carry out an experiment, what should you take into consideration?

# 8. A case study

Different from survey and experimental studies, a case study typically involves a small number of subjects and aims at qualitative data. In this chapter, I will first explain what is a case study, and then describe how to select the subjects and collect the data in a case study. A paper reporting on a case study is presented as Appendix Three.

## WHAT IS A CASE STUDY?

Case studies are extensively used by applied linguists to gain an depth understanding of L2 learning and teaching. In the following section, I will first give you the definition of a case study together with illustrative examples. Secondly, I will discuss with you the possibilities of having a case study as an independent study or as a dependent one.

### Definition and examples

What is a case study? Let's begin with a very simple definition: a case study is a study that examines one or more cases in detail by using multiple sources of data. Let's look at some examples first.

> Example One:
> Chen[10] (2000) studied one six-year old boy whose way of learning English words was particularly interesting. Different from other children, be tried to remember new words through pictures. Cben observed how he learned new words for a period of six months and took down a lot of notes.
>
> Example Two:

[10] She had later changed her mind and did not incorporate this part into her thesis.

Dong (1998) studied a case of a joint online course in which the students from Nanjing University and Randolph-Macon Women's College read selected writings of Pearl S. Buck and then communicated online about what they read. She observed the classes, and collected the e-mails and assignments for the purpose of examining the advantages and disadvantages of the course.

Example Three:
In order to illustrate and supplement the results from a survey study, Wen (1993) examined two students who entered Nanjing University with similar matriculation scores but differed greatly in English Proficiency Test for English Majors-Band 4. Through interview scripts, observing notes and the students' diaries, she intended to find out why these two students showed such great differences in their achievement after two years' study in university.

Example Four:
Wang (2000) intended to find out whether the use of L1 is decreasing or increasing in L2 writing along with the progress of students' L2 learning. 16 English majors across four university levels with four from each level participated in the study where they were asked to write two compositions in English individually while thinking-aloud.

From the above examples, you may find that a case study may involve a single case such as Example One or Example Two, or cover two or more cases such as Example Three and Example Four. We may call the first two a single-case study where the focus is within the case. The last two are a multiple-case study where the focus is both within and across cases.

What is a case? It is difficult to give a short and clear answer since any phenomenon can be qualified as a case (Miles & Huberman, 1994; Theodorson & Theodorson, 1969). It could be one L2 learner, one L2

teacher, or a class, or a school. It could be a decision such as a language policy, an L2 curriculum, a teaching approach and an L2 course. It could also be an attribute of an individual such as vocabulary learning strategies, motivation, etc. or an institution.

According to Stake (1988, p. 258), a case study is "a study of a bounded system, emphasizing the unity and wholeness of that system, but confining the attention to those aspects that are relevant to the research problem at the time." This definition reveals several important features of a case study. First of all, case studies do not investigate all the details of each case. Instead, the case study has a clear focus which serves as a lens through which the researcher studies the case. Secondly, the focused aspect should be examined within the context. Thirdly, the focused aspect has to be viewed as part of a system rather than as an isolated factor.

## SELECTING THE SUBJECTS

Unlike a survey study or an experimental one, the subjects for case studies are not required to be selected by randomization. Such a requirement is waived not because of difficulty in or impossibility of doing so, rather, it is simply because randomization is not meaningful for a small number of subjects. However, the absence of randomization does not mean that the subjects can be chosen without careful thinking. In this section, I will introduce to you how to select the subjects for case studies.

### Single case or multiple cases?

Do you select a single subject or several subjects for study? The decision to be made is no doubt based on research questions. In research on medicine, a single patient whose disease is extremely rare is worth documenting and reporting to other people. Similarly, in the field of applied linguistics, a single L2 learner or teacher is fascinating to investigate if s/he has something unique. However, for an MA thesis or a doctoral dissertation, it is common to have multiple cases rather than a single case.

## Considerations for subject-selection

When the subjects are to be chosen, you first have to think clearly about how to select them. Success in selection results from quite a few factors. In the following section, I will discuss with you two major factors: (1) the subjects' attributes and (2) the subjects' attitude.

In the selection of subjects, the first factor you need to consider is their attributes. To be specific, you have to decide in which attributes you want the subjects to be varied and in which ones to be similar. The simple answer to this question is except for the attributes you are to investigate, you should try to neutralize as many attributes as possible. Now I will take my own study that has been introduced at the beginning of this chapter (i.e. Example Three) for example. As was mentioned before, my purpose was to find out why some students with almost identical scores on their Entrance Examination showed great variations in their scores on TEM-Band 4 after two years' university learning. My speculation was that Strategy Use could account for the variations in L2 achievements in some cases if not in all. For such a research purpose, the subjects were deliberately chosen from the same class in order to keep the effects of teaching quality constant and furthermore, the two subjects were almost identical in family background, L1 proficiency, previous L2 proficiency and Learning purpose as shown in Table 9.1. The essential differences were displayed only in their scores on Band 4 and Efforts (i.e. the time spent learning outside class).

| ID | Sex | Age | Parents | L1P | L2P | LP | Efforts | Band4 |
|---|---|---|---|---|---|---|---|---|
| 01 | F | 19 | F: University teacher<br>M: University teacher | 75 | 95 | 2.67 | 20.5 | 90.50 |
| 06 | F | 20 | F: Doctor<br>M: Nurse | 75 | 96 | 2.67 | 40 | 64.25 |

Table 9.1: Information about the subjects involved in a case study

Note: L1P = Scores on the Matriculation English Test

L2P = Scores on the Matriculation Chinese Test

LP = Views about the importance of learning English

Efforts = Amount of time on self-initiated study per week

Band4 = Scores on English Proficiency Test-Band 4

First of all, you might have noticed that I tried to keep constant as many variables as possible except for the variable of scores on TEM-Band 4 and Efforts. Why should I do so? In this way, the differences identified later in Strategy Use could be regarded as a major cause for their varied L2 achievements without too much disagreement. Otherwise, even if the differences in their use of strategies did exist, we might bave strong reasons to believe that there were many other factors working together with the use of strategies. Secondly, you might be curious about difference in the variable of Efforts. You see that I intentionally selected the cases where the high achiever spent less time studying outside class than the low achiever rather than the other way around. The reason for doing so is that my interest was in the role of Strategy Use and I did not want other people to challenge my conclusion by saying that the more important reason was Efforts rather than Strategy Use which led Subject One to a better L2 learning outcome.

The second factor you have to take into account is the subjects' attitude towards the study. By contrast with survey and experimental studies, a case study only involves a few subjects and therefore we cannot afford to lose any of them. Furthermore, compared to survey and experimental studies, a case study is much more demanding on the time and effort of subjects who are often required to undertake a series of tasks individually. Obviously, without their sound cooperation it is not possible to conduct any case study. How can we make sure that the subjects to be chosen will have a positive attitude? First of all, all the candidates should be well iuformed of what they are expected to do in the study and in whicb way you will compensate for their time and effort, if

you plan to, before they make their own decision whether they are willing to join the study or not. It is not uncommon for a few of them to refuse to participate in it. The refusing rate is about 20% on the average, and it is higher from poor learners than from good learners. Therefore, you had better select more candidates for poor learners if a balance between these two types of learners is essential for your study. After all, all the subjects who are eventually involved in the study must be volunteers. To motivate them to cooperate with you, it is often best to find a way to show your appreciation of their work. The simplest way is to give them some gifts such as dictionaries, books or stationery. As an alternative or an additional incentive, you may promise to help them analyze their problems in learning and provide them with constructive suggestions once the study is finished. This is more effective sometimes, particularly for the poor learners who are anxious to improve their study.

## COLLECTING DATA

In case study, the researcher more often than not needs to collect the data from multiple sources by different techniques. Widely-used techniques include interviewing, think-aloud, diaries, etc. Very often the researcher employ two or more techniques in one study. In this section, I will introduce to you these common techniques one by one.

### Interviews

In this section, we will discuss a very useful technique for collecting qualitative data, i.e. interviewing. What is an interview? It may simply be defined as a talk through which the researcher asks the interviewee a series of questions to find out some information about the interviewee. As was mentioned before, questionnaires may contain closed questions or open-ended questions. Similarly, interviews may be conducted by open-ended questions or closed questions. However, only truly open-ended questions can lead to qualitative data. Therefore, the interviews discussed below only contains open-ended questions.

### *Advantages and limitations*

The advantages of interviewing are various. It can elicit information which is impossible to obtain by other data-collection techniques. For example, you cannot observe feelings, thoughts, and intentions. You cannot observe situations that took place at some previous point in time. You cannot observe situations that preclude the presence of an observer. You have to ask people questions about those things. The purpose of interviewing, then, is to allow us to enter into the other person's perspective. The assumption we have for interviewing is that the perspective of others is meaningful, knowable, and that it is possible to make such a perspective explicit. However, we have to be cautious about such self-reported data resulting from interviewing. In an interview, people may understate or overstate something due to their unreliable memory or to their personal bias. Furthermore, people are not able to report things which are not consciously noticed. Finally, poor interviewing skills of the researcher may greatly reduce the validity and reliability of the interview data.

### *Three types of interviews*

There are three types of interviews depending on the degree of freedom on the part of the interviewer: (1) unstructured interview; (2) semi-structured interview and (3) structured interview.

Unstructured interviews are also called open interviews. They provide interviewers with a lot of freedom. The interview questions are generated spontaneously in the natural flow of an interaction. Typically, the interviewee may not even realize the interview is being conducted. For example, the teacher may have a talk with her students individually, discussing the problems in their mid-term exams. Actually, one of the purposes of such a talk is to find out to what extent the students differ in their ability to identify their own problems.

The semi-structured interview is conducted according to an interview schedule which is prepared before the interview begins. The questions in

the schedule need not be taken in any particular order and the actual wording of questions is not determined in advance. The interview schedule presumes that there is common information that should be obtained from each person interviewed. The interviewer is thus required to adapt both the wording and the sequence of questions to specific subjects in the context of the actual interview. Such an interview requires the interviewer to have a high level of skills.

The structured interview consists of a set of open-ended questions carefully worded and arranged with the intention of taking each interviewee through the same sequence and asking each interviewee the same questions with essentially the same words. Flexibility in probing is more or less limited, depending on the nature of the interview and the skills of interviewers. The structured interview is particularly appropriate when several interviewers are needed to conduct interviews on the same topic. By controlling and standardizing the interview questions, the evaluator obtains data that are systematic and thorough for each interviewee but the process reduces flexibility and spontaneity.

The common characteristics of all three types of interviewing is that the interviewees respond in their own words to express their own personal perspectives. While there are variations in strategy concerning the extent to which the wording and sequencing of questions ought to be predetermined, there is no variation in the principle that the response format should be open-ended. In other words, the interviewer never supplies and predetermines the phrases or categories that must be used by interviewees in their response.

The classifications of interviews appear to be clear cut. However, in an actual study, it is not uncommon that the researcher moves back and forth from one type of interview to another depending on his/her purpose.

### *Preparing an interview guide/schedule*

For semi-structured and structured interviews you should prepare an interview guide or schedule which consists of a series of questions. The

interviewer must decide what questions to ask, how to sequence questions, how much detail to solicit, how long to conduct the interview, and how to word the actual questions. Specifically speaking, preparing the interview guide involves the following tasks: (1) decide what type of information you want to obtain; (2) determine the sequence of questions; (3) choose the wording of questions.

**Types of information**

Like a questionnaire, interviews can generally obtain five types of information: (1) experiences/behaviors; (2)feelings; (3) opinions; (4) knowledge/abilities; (5) background. You can put the above five types of questions in the time frame: past, present and future. Once you have decided the general type of information, you need to determine exactly which research questions your interview data intend to address.

**Sequence of questions**

There are no fixed rules of sequence in organizing an interview. Unstructured interviewing is flexible so that a fixed sequence is seldom possible. However, structured interviews must establish a fixed sequence of questions due to their structured format. I offer, then, some suggestions about sequencing:

(1) Easier questions first;

(2) Questions about "here and now" first;

(3) Interesting questions first.

**The wording of questions**

An interview question is a stimnlus that is aimed at creating or generating a response from the person being interviewed. The way a question is worded is one of the most important elements determining how the interviewee will respond. Asking questions is an art. For purposes of qualitative inquiry, good questions should be at least open-ended.

What questions can be regarded as open-ended? The truly open-ended question allows interviewees to take whatever direction and use whatever words they want in order to respond in their own way. You know in closed questions, the subjects are asked to choose one as their answer out of a limited and predetermined set of alternatives given. In other words, in closed questions, the response possibilities are clearly stated and made explicit in the way in which the question is asked. Many interviewers think that the way to make a question open-ended is simply to leave out the structured responses. Such an approach does not, however, guarantee a question is truly open-ended. What has been achieved is often to make predetermined response categories implicit and disguised.

Consider the following "open-ended" question: "How satisfied are you with the speaking class in a language laboratory?" On the surface, this appears to be an open-ended question. On close inspection, however, it is clear that the dimension along which the subject can answer the question has already been identified, i.e. the subject is being asked for some degree of satisfaction. It is true that the interviewee can use a variety of modifiers for the word satisfaction—for example, "pretty satisfied," "kind of satisfied," "mostly satisfied," and so on. But, in effect, the possible response set has been narrowly limited by the wording of the question.

The truly open-ended question does not presuppose which dimension of feeling or thought will be salient for the interviewee. The truly open-ended question permits the person being interviewed to select from among that person's full repertoire of possible responses. Examples, then, of truly open-ended questions would take the following format:

> How do you feel about the speaking class in the language laboratory?
>
> What is your opinion of the speaking class in the language laboratory?

Moreover, a truly open-ended question cannot be phrased as a Yes/No question. The purpose of an in-depth interview is to get the interviewee to talk — to talk about experiences, feelings, opinions, and knowledge. Far from encouraging the interviewee to talk, Yes/No questions create a dilemma for the interviewee because they frequently are not sure whether they are being asked for a simple response. Let's look at the following example which is given by Patton (1990: 297-298):

(Teenager returns home from a date)
Do you know that you're late?
Yeah.
Did you have a good time?
Yeah.
Was it a good movie?
Yeah, it was okay.
So, it was worth seeing?
I've heard a lot about it. Do you think I would like it?
I dou't know. Maybe.
Anything else you'd like to tell me about your evening?
No, I guess that's it.

Yes/No questions can turn an interview into an interrogation or quiz rather than an interactive conversation.

**Interviewing strategies**

In order to have data with high quality and use them effectively, you may employ various kinds of strategies listed below.

**Using a tape recorder** Don't rely on your memory in interviewing and use a tape recorder[11] in all cases. Tapes are a permanent record of primary information that can be kept and passed on to other researchers. It is often best to prepare two tape recorders in case one is out of order. You should use really good tapes. Don't use thin tape because transcribing involves listening, stopping, and rewinding-often hundreds of times per tape. Thin tape just cannot stand up to this kind of use.

You must test your tape recorder before every interview. There's nothing worse than a recorder that does not run at all. If your tape recorder uses batteries, you should prepare spare ones. Good tape recorders have battery indicators. When batteries are slightly low, throw them out. Otherwise, the quality of recording will be affected. If you plan to use house current, you need to prepare batteries too because the electric power may be suspended without notice.

**Probing** The key to successful interviewing is learning how to probe effectively—that is, to stimulate an interviewee to produce more information, without injecting yourself so much into the interaction. There are many kinds of probes that you can use in an interview. In the following part, I will introduce to you some probes recommended by Bernard (1994).

The apparently easiest yet the most difficult probing technique is the silent probe, which simply means remaining quiet and waiting for an interviewee to continue. The silence may be accompanied by a nod, or by a mumbled "uh-huh". The silent probe sometimes produces more information than does direct questioning. Some interviewees are more talkative than others and require very little probing to keep up the flow of information. Others are more reflective and take their time. Inexperienced interviewers tend to jump in with verbal probes as soon as

---

[11] You have to gain permission from your interviewer before you record the interview.

an interviewee goes silent. Meanwhile, the interviewee may be just reflecting, gathering thoughts, and preparing to say something important. You can kill those moments with your interruptions. On the other hand, the silent probe is a risky technique to use, and that is why beginners should avoid it. If an interviewee is genuinely at the end of a thought and you don't provide further guidance, your silence can become awkward. You may even lose your credibility as an interviewer. The silent probe takes a lot of practice to use effectively but it's worth the effort.

Another kind of probe is called the echo probe that is simply repeating the last thing an interviewee has said and asking them to continue. This echo probe is particularly useful when an interviewee is describing a process, or an event. Suppose you ask an interviewee to describe his/her reading process. Once the interviewee stopped describing, you might say "I see. You read the title first to predict the general idea about the text. Then what do you do next?" This probe is neutral and doesn't redirect the interview. It shows that you understand what has been said so far and encourages the interviewee to continue with his/her narration. If you use the echo probe too often, though, you'll hear an annoyed interviewee asking you, "Why do you keep repeating what I just said?"

The third kind of probe is called the Uh-huh probe. You can encourage an interviewee to continue with his/her talk by just making affirmative noise, like "uh-huh," or "yes, I see."

**Conversational style** By a conversational style, I mean that an interviewer and an interviewee should talk to each other in a natural manner. That is to say, a question you ask should be related to the previous answer given by the interviewee. This is a difficult task. On the one hand, as an interviewer, you have prepared a set of questions to ask and you have to keep the conversation on the track. On the other hand, you should talk in such a manner that the interviewee does not feel that s/he is being interrogated. What is required here is flexibility and spontaneity.

### *Common problems in interviewing*

By observing our graduate students' simulated interviews, I have found several common problems. The first one is that many students who cannot put up with a single second's silence interrupt the interviewee's thinking by more questions. The second problem is that they often ask Yes/No questions to which the interviewee only offers a short answer. They then do not know what should be asked next. The third problem is that the interviewer does not know how to draw the interview back to the right track when the interviewees digress. The last problem is that the interviewer talks too much and s/he basicially gave the interviewee the answer s/he was hoping to get.

These problems can not be overcome overnight. Developing an effective interview technique requires a lot of reflective practice. By reflective practice, I mean you have to make a self-evaluation of each of your simulated interviews and try to overcome the weaknesses identified in the next one. I am sure such reflective practice will enable you to become a skillful interviewer.

## Think-aloud

Think-aloud, as a technique of data-collection, has gained legitimate status since the 80's when cognitive research gained favor. Quite a few researchers (for example, Cohen, 1987; Goodman, 1989; Guo, 1997; Guo & Wen, 1998; Flower et al., 1990; Hosenfeld, 1984; Hudelson, 1989; Lauer & Asher, 1988; Lu, 1997; Raimes, 1985; Zamel, 1983, 1987) probe into the process of students' L2 writing and reading by this technique. The findings from these studies provide insights into students' cognitive processes. In this section, we will answer the following questions: (1) What is think-aloud? (2) How is think-aloud carried out ? and (3) What are the limitations of think-aloud?

### *What is think-aloud?*

When a child is at the age between 3 to 7, s/he likes to talk while playing a game. Through this talk, we may know what is going on in her/his mind. The talk the child does is think-aloud. Similarly, in second language research, think-aloud requires subjects to speak out their inner thoughts which are otherwise not accessible to outsiders. By contrast with children who do think-aloud naturally, second language learners have to be trained to do so. In other words, think-aloud is no longer a natural behavior for an adult L2 learner. Therefore, without proper training, think-aloud as a data-collection technique cannot work effectively on people who are no longer young children.

In the following section, I will describe two studies by think-aloud, which were both conducted by our MA students in 1997. Their theses are not puhlished.

Example One: Guo's study

Guo used a think-aloud procedure to investigate to what extent L1 is involved in students' picture composition. In his study, the subjects were 20 senior 2 and senior 3 students from a middle school, who were recommended by their teachers as students who are extrovert and talkative, and more likely to be cooperative in accomplishing the task of thinking aloud.

The training in think-aloud lasted two weeks. It started with the researcher's demonstration in which he spoke out what was going on in his mind while reading a piece of classical Chinese. Then all the selected subjects were asked to accomplish two tasks by think-aloud: (1)Read some classical Chinese prose pieces and show their understanding by thinking aloud; (2) Describe a process by which they understood a distorted or blurred picture or words. These two tasks were well selected because they did not involve the use of two languages, which would not make them think they had to use Chinese mixed with English in their later L2 writing.

Once everyone understood what thinking-aloud was and knew how to carry it out, they each were asked to write a story in English of about 100 words based on a series of pictures while trying to say aloud anything occurring in their minds in the whole process of writing. The language used in thinking-aloud data was not specified but they were told that their think-aloud should be natural in the sense that the language used should be the same as the form through which a thought was expressed. To put it simple, if inner thoughts are in Chinese, then Chinese should be employed; if in English, then English should be used. Altogether they were given eight pictures in a sequence describing how a student saved the life of a blind man on his way to school and he was late for his class. A tape-recorder was used to record each subject's composing aloud process and at the same time the stories written dewn were aslo collected from the subjects as the data of product.

Once the data-collection was finished, all the tapes were transcribed and these transcripts are called protocols. The researcher first made a statistical comparison between the quantitative use of L1 and L2. Secondly, he identified the different functions of the use of L1 by qualitative analysis. Finally, he bypothesized two L2 writing models with a focus on L1 use.

## Example Two: Lu's study

Lu intended to investigate the relationship between L2 proficiency and EFL learners' strategy use in L2 reading. Four junior non-English foreign language majors in Nanjing University were selected for this study. Among them, two are Japanese majors and the other two, Russian majors. One Japanese major and one Russian major formed a group of beginners who hed only studied English for one year and the other two formed a group of intermediate readers who had studied English for seven years. They were all successful learners in their own majors.

Two passages were both chosen according to the readers' proficiency levels. They were asked individually to read the given passage while

reporting verbally as much as possible about how s/he understood the text. Their verbal reports were recorded. The researcher remained silent and observed their performance. An interruption by the researcher only occurred when the subject kept silent for more than a minute. No time limit was set for reading.

To make sure they knew how to go about thinking-aloud, the subjects were asked to read a sample passage after the researcher's instruction. Since they were university students, it did not take long for them to understand the think-aloud procedure.

The verbal reports were transcribed and then the protocols were coded on the basis of Block's classification of strategy categories (Block, 1986) with some necessary modifications. The results revealed from the study are the following:

The relationship between L2 proficiency and comprehension strategy use is rather complicated. L2 proficiency, on one hand, does not prevent the beginners from making use of some higher-level comprehension strategies and using them effectively. On the other hand, L2 proficiency is indeed a constraint on the use of strategy. Compared to more proficient L2 learners, less proficient L2 learners tend to use fewer varieties of reading strategies and use some categories less frequently and/or with poorer quality. Furthermore, lower L2 proficiency more likely produces a higher level of anxiety and frustration in dealing with new words and difficult sentences.

## *How is think-aloud carried out?*

Think-aloud is much more difficult than it appears to be. To make sure think-aloud is working successfully, you have to at least meet the following three requirements. First of all, the subjects you selected should be very cooperative in the sense they are willing and able to think aloud as required. Secondly, the training you conduct must be sufficient and adequate. Thirdly, the quality of the tape recorder has to be guaranteed. Finally, the researcher should behave professionally when a

subject is undertaking the task by thinking-aloud. In the following section, we will discuss how the above criteria can be met.

**Selecting "ideal" subjects**

By saying "ideal" s[illegible]cts, I do not refer to the subjects who can provide you with the information that can support your assumption. In this context, "ideal" subjects are those who are talkative, and capable of thinking-aloud after proper training. In fact, not all the subjects feel at ease to think aloud even after good training. For example, some L2 learners who are not good at articulation are quiet and hardly utter anything to their classmates in daily life. Some L2 learners, although they are talkative, don't show a positive attitude towards the technique of think-aloud. Therefore, the "ideal" subjects are those who are talkative and are also able to think-aloud without too much uneasiness. This of course is a constraint on the researcher's data.

The subjects eventually involved in your study are usually chosen through three stages. At the first stage, you select the subjects for training. The initial criterion for the first selection is the subject's personality, i.e. being talkative. Hence, the students who are obviously introverted and reserved are not candidates for training. The second selection is made after well-constructed training. At this stage, you are only interested in the participants who feel at ease in think-aloud and are also willing to be cooperative with you. According to our previous experience, nearly 40% of trainees are screened out because they are not happy with think-aloud. The third selection is made once you have collected all the data by think-aloud. You try to eliminate the subjects whose noticeable pauses comprise more than 10% of the total amount of time of think-aloud. The third selection often reduces 10%-20% of the subjects. Since quite a few subjects are screened out from the second and the third selections, you have to choose more than the required number for initial training and for think-aloud data collection.

**Providing effective training**

To provide subjects with effective training, the normal procedure is that the researcher first demonstrates how think-aloud is carried out and then gives the subjects at least two opportunities to practice it. In order to save time, the demonstration can be made in a group and the first try-out thinking-aloud can be undertaken in a laboratory where their thinking-aloud process can be recorded. Furthermore, the subjects might be less nervous when they are doing it together and no one is able to hear the others. However, the second try-out is better done individually. By doing so, you can easily offer help if they have any misunderstanding about thinking-aloud. Moreover, you can easily identify the subjects who are not suitable for this kind of study.

One potential problem in the training is that the researcher's demonstration may be misleading in the sense that the researcher's think-aloud illustrates how a certain type of data are produced. For example, suppose you want to find out to what extent L1 is involved in L2 writing. In your demonstration, you speak out what is going on in your mind sometimes in Chinese and sometimes in English. The subjects might then think this is the way they should think aloud. Thus, to avoid this problem, your demonstration should be carefully designed. In this case, it is often best to choose a task that is not related to language learning at all, for example, to guess the missing part of a character, or to resolve a puzzle.

**Having a tape-recorder with high quality**

A tape-recorder used for think-aloud should be checked more carefully than those used in interviews where the researcher asks the subjects one question after another and the interview can be stopped if the machine does not work properly. Furthermore, the interview questions are usually discrete and are not necessarily linked logically. By contrast with the interview, in thinking-aloud, once the flow of the subject's thoughts is interrupted, it is difficult, if not totally impossible, to resume as it was before.

Secondly, it is often best to use new batteries for each subject to make sure the batteries do not run out of power in the process of think-aloud. Finally, you should make sure that the machine is set in the mode of recording when the subject starts the task. One problem I once came across is that the student felt unhappy with what had been said and she stopped the machine. If this happens, you have to ask him/her not to do it again and remember to push the recording button when s/he resumes the task.

**Behaving professionally**

How should a researcher behave when a subject is undertaking a task by think-aloud? In general, the researcher should behave professionally. By behaving professionally, I mean that whatever the researcher does should be conducive to the completion of the task by the subject. I suggest that you sit quietly in a corner that is out of the sight of the subject while observing the subject's performance. If necessary, you may take some notes. Remember that any movement and noise you make will certainly disturb the flow of thought of the subject.

How do you deal with the situation where the subject keeps silent for a while? Should you ask him/her a question such as "What are you thinking now?" to push him/her to speak? In my opinion, instead of asking him/her a question, you had better make a sound signal indicating that s/he has to stop pausing. Of course, you need to tell your subject about the signal before the data-collection starts.

## *Limitations*

Think-aloud has gained its legitimate status in research on second language learning in the last two decades. As a result, a growing number of studies using this method have gained some valuable insights into L2 learners' mental processes which are not accessible by other data-collection techniques. However, this method has some unavoidable weaknesses which, as a researcher, must be kept in mind when interpreting the results.

As was mentioned before, the think-aloud technique can only be effective for the subjects who are eventually able to accomplish the task of thinking-aloud successfully. Therefore, it is doubtful that the data reported by such a highly selective group are the same as for those who are screened out. As a result, we have to admit the subjects involved in the study are not typical.

No matter how effective the training is, thinking-aloud is not a normal behavior of an adult. When the adult is required to do it, his/her mental process is likely to be distorted in one way or another. It is not clear to what extent the data elicited by think-aloud can reflect natural thinking. However, the more effective the training is, the more natural the data will be.

Although the subjects are required to speak out all that is going on in their minds, some mental activities inevitably go unreported verbally since thinking is always faster than speaking. Also, you can speak out something while thinking other things. In this sense, think-aloud can never 100% report the inner thoughts. To compensate for these inadquacies, the researcher may conduct an immediate follow-up interview to elicit more information.

Although we are aware of the limitations of think-aloud, this technique is the best available one to date for us to gain access to the inner thoughts of the learner. Along with the development of cognitive research, people are more curious about the mental activities involved in L2 learning. So far, we have not reached the stage where scientific instruments are capable of examining human inner thoughts. Think-aloud, in my opinion, can at least partially satisfy our curiosity.

## Diary

Diary is also an important tool in the research on second language learning and teaching. It can record the learner's or the teacher's feelings, attitudes, behaviors, reflections and conscious awareness of cognitive

process in a language class or other second language learning/teaching contexts.

### *Definition*

According to Bailey's definition (1990:215 cited by Nunan, 1994: 120), the diary used in research is "a first-person account of a language learning or reading experience, documented through regular, candid entries in a personal journal and then analyzed for recurring patterns or salient events"(p. 215). Her definition emphasizes three important requirements. First of all, the subjects have to keep the diaries themselves. Secondly, the subjects need to keep their diaries regularly and honestly. Thirdly, the purpose of keeping a diary is either to find out similarities or unique features displayed by diary-keepers.

However, in my opinion, Bailey's definition does not capture the essential differences between a diary in our daily life and a diary used as a research tool. First, a diary kept by a person in daily life is an account of personal feelings, views, random thoughts, etc. which is not meant for any public purpose. Therefore, it can be fragmentary, illogical and ungrammatical. However, once a diary is used as a research tool, it has lost the characteristic of being private. The diary keeper must write in a way that other people can understand without much difficulty. Secondly, the diary in our daily life can record anything occurring in people's minds, while the diary for research has to contain information relevant for the research purpose. Finally, the diary in our daily life is primarily to keep an account of things that have just happened whereas the diary for research can record not only the activities that have happened now and here but also then and there. When the diary is used to record the then-and-there experiences, I will call it a reflective diary.

Based on the above differences, I would like to define a diary in research as "a comprehensible written account of one's own experiences in L2 learning and teaching which are however confined to aspects relevant to specific research problems".

## *Illustrations*

A few second language researchers have recorded their own experiences of learning languages other than the first language in an attempt to have a better understanding of the process of L2 learning. For example, Rivers (1983) in her diary described how she strove to learn a sixth language. Bailey (1983) also recorded through a diary her rich experiences in learning French.

However, many more researchers ask their subjects to keep diaries according to the given instructions so that they can study other people's language learning. In the following section, two examples are given to show how data-collection can be conducted through diaries.

### Example One: Wen's study

In my study (Wen, 1993) on the role of learner controllable factors in L2 learning, 10 subjects out of the 242 subjects were selected for case studies. As part of the case studies, they were required to keep one-week diaries to record their self-initiated English learning activities in order to find out the differences in the use of learning strategies between high and low achievers. The diary instructions were given as follows:

> Please keep a diary for a week. In the diary, you are required to record the English learning activities you undertake after class every day. Your writing should address the following aspects:
>
> 1. What type of activities are you undertaking? And why?
> 2. How much time do you spend on each activity?
> 3. What is your physical and psychological state when carrying out the activity (i.e. physical state: tired or energetic; psychological state: attentive or absentminded)?
> 4. What do you think of the activity? (Is it important for improving your language proficiency?)
> 5. What strategies have you adopted in undertaking the activity? (Please try to give details as much as you can.)

> Since my purpose in asking you to keep a diary this time is not to test your language proficiency, you may use any language (English, Chinese or mixed code) you feel that can best express yourself (Wen, 1993: 309).

By examining the questionnaire and interviewing data provided by the 10 subjects, the conclusion was that both low and high achievers were active strategy users. However, the diaries displayed quite a few qualitative differences between high and low achievers, which made me think why such differences were not present in questionnaire and interview data. The reason I found was that the high achievers tended to set up higher demands on themselves than the low achievers so that the high achievers were more likely blamed themselves for not doing well in responding to questionnaire items and interview questions. In contrast, the low achievers tended to overvalue themselves since their expectations were not that high. For example, through diaries, I found that the low achievers tend to be poorer in the use of management strategies than the high achievers. The following excerpts from the one-week diaries were written by one low achiever.

> Wednesday, Oct. 23, 1991
>
> "After class I went to the library, sitting in a section of 'Language and words'. There were many books in this section. I often chose different books to read. But today, I was interested in the book "Idioms for Everyday Use". The idioms were arranged in an alphabetical order. Most of them are useful, such as 'Nothing ventured, nothing gained'...So I copied down some sentences I liked and memorized them at the same time. I copied down more than 20 sentences. But I found I only reached the letter 'c'. It took me too much time and I did nothing else in this afternoon. *I often did things like this: I do not care whether it is helpful or* not, so long as I like it, I will spend time on it As with my English study, even though it seemed that I learned little this afternoon, 'Many a mickle

make a muckle'. As time goes on, I believe I will feel at home in English." (Subject 10)

Saturday, Oct. 26, 1991

This afternoon, I am reading 'Pygmalion'. The whole book has about 300 pages including the introduction to the author's life and his works. I spend three hours but only cover the first 10 pages of the introduction. While reading, I copy down some good sentences and some information about the author. I underline the sentences that are difficult for me to understand with a blue pen and underline the key words and useful expressions with a red pen. When I encounter new words, I look them up in a dictionary and write down the explanations beside the new words. It is good for me to read famous English works this way

Obviously, the above two reading activities recorded in her diary were ad hoc decisions. Furthermore, in her outside reading, she habitually or automatically employed the strategies that she used in reading prescribed texts. Since she followed the same method to continue these two reading activities, it would be extremely difficult for her to accomplish them because she could not afford the time. Even if she finished them, it could be imagined that she would not have fruitful results. Unfortunately, given such ineffective learning strategies, she still showed satisfaction with them.

**Example Two: The study by Schmidt & Frota**

Schmidt and Frota (1985) conducted a case study in which a subject, a native speaker of English, was asked to keep a diary documenting his learning experience of learning Portuguese in Brazil. The diaries indicate that the subject's learning could be divided into three stages: (1) learning through natural communication with local people, (2) learning througb formal instruction and (3) learning through natural communication with subsequent self-reflection. The following three extracts selected by

Nunan (1994: 121-23) can illustrate the subject's cognitive as well as affective experiences in these three stages of L2 learning.

Extract 1

Journal entry, Week 2

I hate the feeling of being unable to talk to people around me. I'm used to chatting with people all day long, and I don't like this silence. Language is the only barrier, since it is certainly easy to meet Brazilians. I've noticed that it is acceptable to ask anyone on the street for a cigarette. It...appears to have no relationship to age, sex or class. Last night an attractive and obviously respectable young woman, accompanied by her boyfriend, stopped me and bummed a cigarette. If I take a pack to the beach, it disappears within an hour, so that's 20 people I could have met.... Today P and I were at the beach, a guy came up for a cigarette, sat down and wanted to talk. He asked if I were an American and I said *sim*. He said something I didn't comprehend at all, so I didn't respond. He said, "well, obviously communication with you would be difficult" (I did understand that, though I can't remember any of the words now), and left. (p. 242)

Extract 2

Journal entry, Week 4

P and I started class yesterday. There are 11 in the class (of various nationalities). The teacher is young and very good. She introduced herself to us (in Portuguese); I am X, my name is X, I am your teacher, I am a teacher, I am a teacher of Portuguese. I'm also a teacher of English. I'm from [place], I'm single, I'm not married, I don't have children, I have a degree in applied linguistics etc. She went around the class, asking the same kinds of questions: What's your name? Where are you from? What kind of work do you do? Do you have any children?, etc. Most of the students could answer some of the questions, e.g. I know what my title is at the university. Everyone was rapidly picking up new things from the others' answers. For the rest of the class, we circulated, introducing

ourselves to each other and talking until we exhausted the possibilities. At the end of the class, X put the paradigm for SER on the board, plus a few vocabulary items. Great! This is better than *bom dia* and then silence....I'm sure I'll be asked all those questions thousands of times before I leave here. So I went out last night and talked to four people. It worked, and I'm invited to a party tomorrow night. Of course I quickly ran out of things to say and quickly stopped understanding what people said to me, but that just makes me eager to get back to class. (p.243)

Extract 3

Journal entry, Week 11

H and I ate dinner at Caneco 70. He complained non-stop about his job. I tried to say "you don't seem comfortable" with the job: *sinto que voce nao esta confortavel*, and his face showed complete non-comprehension. I grabbed my dictionary.

"Comfortable" is comfortavel, but it flashed through my mind that perhaps you can only say chairs are comfortable, not people. A few minutes later H said something with *nao deve*. I was taught DEVER as "have to" or "must," and I've been thinking that *nao deve* + Verb would mean "don't have to" and *deve nao* + Verb would mean "must not," but H's remark obviously meant "should not." So I learned something, but in general H is a terrible conversationalist for me. He doesn't understand things I say that everyone else understands. When I don't understand him, all he can ever do is repeat. (p. 246)

In the examples above, we may get general ideas what a diary looks like for the research purpose and how a diary can reveal insights into processes of L2 learning.

## *Advantages and limitations*

The use of diaries is more convenient and easier than think-aloud and interviewing. First of all, it does not take so much time on the part of

researcher as that required by think-aloud and interviewing since the latter methods have to be conducted on an individual basis and have to be transcribed afterwards. Secondly, it is less technically demanding than think-aloud and interviewing. Using think-aloud, the subjects must be cautiously selected and trained effectively before data-collection. In the case of interviewing, in addition to having a carefully-designed interview schedule, the quality of the data also very much depends on the interviewing skills of the researcher. For diaries, there is only one thing you need to pay great attention to, that is, the quality of instructions given to the subjects. However, the advantage can be easily turned into a disadvantage since the data with poor quality due to unclear instructions can be known only when the data have been collected, which is too late to remedy.

Althougb diaries sometimes can provide us with the data which would be difficult to gather in other ways, diary data face challenges and criticism. First of all, people doubt on their validity. To put it in simple words, people are not sure to what extent the recorded data in the diary can reflect the reality. Obviously, once the subjects describe their experiences and attitudes towards L2 learning activities after the events occur, they might distort them unintentionally. Furthermore, diaries usually document the things the subjects are conscious of. Finally, diaries are sometimes not clearly written or are difficult to understand.

### *Problems in using the technique of diaries*

Althougb the use of diaries is easy and convenient, it can be a total failure which goes unnoticed until the diaries are submitted to you. By contrast with think-aloud and interview where you may notice the problems in the process, problems occurring in diaries can only be realized after data-collection. Therefore, you must anticipate the problems the subjects may have and try to prevent them from happening by giving the subjects clear oral instructions as well as written ones or checking the diaries while in progress.

One problem our postgraduate students have is that in the subjects' diaries, much information is irrelevant and the relevant information is far less than needed. The cause for the problem is that the researcher did not explain to the subjects explicitly what they are expected to write. The second problem is that if instructions are biased, the subjects' diaries contain the information that reflects that bias. For example, in the students' diaries in one study, all of them discussed the reasons why they overused textual connectives by attributing this to the teacher's teaching. Once I read their diaries, my first question was "Did all the subjects involved here overuse the textual connectives?" When the researcher admitted that this was not the case, I asked the second question: "Why did they say so?" Later I found that one of the instructions given by the researcher was the following: Please try to explain in your diaries why you overuse textual connectives. Obviously, the instructions here were biased.

## SUMMARY

A typical case study involves a few subjects and aims to collect qualitative data. It may study a single case or multiple cases depending on the research purpose. In the selection of subjects, two questions need to be answered: (1) what attributes of subjects need to be neutralized and what attributes to be varied? (2) Are the subjects willing to cooperate with you? To collect the data in a case study, three techniques are discussed: (1) interview; (2) think-aloud and (3) diaries. An interview is an interaction through which the researcher asks tha interviewee a series of questions to find out some information about the interviewee. It can elicit information that cannot be obtained by other data-collection techniques, such as the interviewee's feelings, opinions in the past, at present and in the future. Its most obvious weakness is that self-reported data by the interviewee are only confined to those at the conscious level and can be distorted intentionally or unintentionally. There are three types of interview in terms of the degree of freedom on the part of researcher: (1)

unstructured interview; (2) semi-structured interview and (3) structured interview. For semi-structured and structured interviews you should prepare an interview guide or schedule which consists of a series of questions that must be truly open-ended. In order to conduct an interview effectively, you are advised to use a variety of strategies, such as using a tape recorder, probing, using presupposition questions and having a conversational style. Think-aloud is a technique to get inner thoughts in the subject's mind while s/he is performing a language activity. The use of this technique is time-consuming and skill-demanding. The subjects to be selected to do thinking-aloud need to be talkative, and be comfortable with this task and capable of completing it after the training. An effective training is very critical for the success of think-aloud. Diary is also an important tool in collecting data in a case study. It may record what has just happened in L2 learning and teaching or reflect on what happened long before. The use of diaries is more convenient and easier than think-aloud and interview, but the data reported may not truly reflect reality.

## DISCUSSION QUESTIONS

1. What is a case study? Can you find an example to illustrate your understanding?
2. What are the differences between the semi-structured interview and the unstructured interview?
3. What are the advantages and limitations of interviewing?
4. Conduct an interview while using a recorder. Play the tape in a small group and ask the other members to make critical comments on it.
5. How do you collect data by thinking-aloud? How can you make sure that the quality of data is high?
6. Can you explain under what conditions thinking-aloud can be effective?
7. What are the advantages and limitations of think-aloud?

8. When can you use a diary as a data-collecting technique? Please give some examples.
9. What are the advantages and disadvantages of a diary as a data-collecting technique?
10. What are the common problems in the use of a diary as research tool?

# 9. Basic statistics

Being lucky to live in a computer age, we don't need to analyze quantitative data manually with painstaking efforts. There are various statistical procedures available that can provide us with findings within seconds. However, to select an appropriate procedure to process the data effectively, we must have a good understanding of basic concepts and formulas used in statistics. This chapter just serves this purpose.

## A BRIEF DESCRIPTION OF STATISTICS

There are two branches in statistics, namely descriptive statistics and inferential statistics. In this section, I will first briefly describe what are descriptive statistics and inferential statistics respectively and then discuss two important terms: parameter and statistic.

### Definitions

If a researcher is interested only in describing a group from which the data are gathered, the statistics involved are called descriptive statistics. In many cases data in the field of applied linguistics are primarily descriptive in nature.

If his/her interest goes beyond describing the group from which the data are collected and s/he tries to draw conclusions about the population from which the group was selected, the statistics needed in this case are inferential statistics. The advantage of using inferential statistics is that they allow the researcher to make decisions about the population without studying the entire population. The use and importance of inferential statistics is ever growing.

## Parameter and statistic

A descriptive measure of the population is called a parameter. Parameters are usually denoted by Greek letters such as $\mu$, $\sigma$. Examples of parameters are population mean ($\mu$), population variance ($\sigma^2$), and population standard deviation ($\sigma$). A descriptive measure of a sample is called a statistic and is usually denoted by a Roman letter. Examples of statistics are sample mean (x), sample variance ($S^2$) and sample standard deviation (S).

Differentiation between "parameter" and "statistic" is important only in the use of inferential statistics. A statistician often wants to estimate the value of a parameter. However, the calculation of parameters is usually either impossible or not feasible because of the amount of time and money required to take a census. In such cases, the statistician can take a random sample of tbe population, calculate a statistic on the sample, and infer by estimation the value of the parameter.

Suppose you are an English teacher at Nanjing University and you want to know the relationship between the use of strategies and the English proficiency of all the Nanjing university students. There are more than 10,000 students on two campuses. Obviously, you cannot afford the money and time to take a census. What you may do is to survey 400 students as a sample by a questionnaire. Then you use descriptive statistics and inferential statistics to analyze the data.. For example, you use descriptive statistics to get to know which strategies are more frequently used by these 400 students. The results you have got are statistics on the sample. You may use inferential statistics to examine the relation between the use of strategies and English achievement, or to see whether the male students differ from the female students in their use of strategies based on the notion of probability.

Descriptive analysis or statistics, as its name suggests, simply describes the general pattern or tendency emerging from the data

collected, while inferential analysis or statistics, more complicated than descriptive, aims at predictions beyond the sample data.

## DESCRIPTIVE STATISTICS

The statistics used to summarize data are called descriptive statistics. Once the data have been coded and checked, descriptive statistics will be used to organize the data in terms of frequencies, central tendency and variability.

### Frequencies

The simplest way to organize the data is to describe their frequency distribution which can reduce and summarize data effectively and efficiently. Frequency distribution can be presented in a table or in graphic forms such as a bar graph, a polygon or a pie chart.

First of all, you will learn how to present frequency distribution in a table form. It is easy to construct frequency distributions if the total possible values are very limited. For example, if the students' responses to a five-point scale questionnaire items are no more than five: 1, 2, 3, 4, 5, frequency distribution can be obtained simply by tallying up all the responses, as shown in Table 9.1.

| Choices | Frequency | Percentage |
|---|---|---|
| 1 (Strongly disagree) | 30 | 30% |
| 2 (disagree) | 45 | 45% |
| 3 (Neither disagree nor agree) | 10 | 10% |
| 4 (Agree) | 5 | 5% |
| 5 (Strongly agree) | 10 | 10% |

Table 9.1: Frequency distribution

However, when the total possible values are various, such as people's age and students' test scores, we need to group the raw data into classes. In this case, the frequency distribution is concerned with classes rather than with each individual score. Then, how do we describe the frequency

distribution in terms of classes? There are no hard-and-fast rules to follow. Very often, the final shape and design of a frequency distribution vary from researcher to researcher even if the original data are identical. The following are the scores on the mid-term exam taken by a group of freshmen:

58, 65, 84, 70, 90, 75, 86, 76, 80, 82,
83, 84, 69, 84, 85, 86, 72, 89, 75, 92

Generally speaking, we may construct the frequency distribution of the above data in three steps:

(1) Determine the range of the raw data;
(2) Determine the number of classes;
(3) Determine the width of the class interval.

The range of the raw data is defined as the difference between the largest and smallest numbers. The range (92-58) here is 34. The number of classes should not be too many or too few. If there are too many classes, we might not be able to achieve generalization. If there are too few classes, we might not be able to see the important differences. One rule of thumb is that the number of classes selected is between 5 and 10. However, it is your personal decision and the final number of classes must serve your research purpose. An approximation of the size of the class width can be calculated by dividing the range by the number of classes. Usually we will choose a round number such as 5 or 10. If not possible, we might use 15, 20, 25, 50 and so on.

The frequency distribution must include all the data given. Therefore, the frequency distribution should start at a value equal to or lower than the lowest number of the ungrouped data and end at a value equal to or higher than the highest number. Class endpoints are selected so that no value of the data can fall into more than one class. We may use the word "under" to indicate the endpoint of each class.

| Class interval | Frequency |
|---|---|
| 50-under 60 | 1 |
| 60-under 70 | 2 |
| 70-under 80 | 5 |
| 80-under 90 | 10 |
| 90-under 100 | 2 |
| Total | 20 |

Table 9.2: Frequency distribution in terms of five classes.

Table 9.2 shows that the number of classes is 5 and the class width is 10. Once the data are presented in Table 9.2, you may offer a reasonable explanation without difficulty. You may say that the 12 students did quite well on the exam while 8 of them still need more effort. In general, the performance of the whole class was not unusual.

However, we may present the same set of data in an alternative way as shown in Table 9.3:

| Class interval | Frequency |
|---|---|
| 50-under 65 | 1 |
| 65-under 80 | 7 |
| 80-under 95 | 12 |

Table 9.3: Frequency distribution in terms of three classes

Frequencies sometimes do not make much sense. Percentage is more straightforward as shown in Table 9.4.

| Class interval | Frequency | Cumulative Frequencies | Percentage |
|---|---|---|---|
| 50-under 60 | 1 | 1 | 5% |
| 60-under 70 | 2 | 3 | 10% |
| 70-under 80 | 5 | 8 | 25% |
| 80-under 90 | 10 | 18 | 50% |
| 90-under 100 | 2 | 20 | 10% |
| Total | 20 | 20 | 100% |

Table 9.4: Frequency distribution with cumulative frequencies and percentages

If you want to know the relative standing of any particular score in a group of scores, you can show it by a percentile score. The following formula can be used to calculate the percentile scores.

$$\text{Percentile} = \frac{\text{Cumulative F}}{\text{N}} \times (100)$$

In the above formula, cumulative frequency (F) refers to the frequency of the score or the scores within the class plus the frequency of the score or the scores within the class just below. As shown in Table 9.4, the scores within the class 60-under 70 occur 2 times and thus their frequency is 2, while their cumulative frequency is 3 that is the result of 2 plus 1, which is the frequency of the scores within the class 50-under 60. The letter N represents the total number of scores.

A percentile can be defined as a number which represents the percent of scores that a particular raw score exceeds. For example, suppose one of your students received a score in the 67th percentile on TOEFL. It means that 67% of the students who took the test scored lower than that level. Or we say that his/her score is higher than 67% of the other people taking the test but lower than 33% of the people.

To make sure you have a correct understanding of the definition of percentile, Table 9.5 presents more percentiles and their interpretations.

| Percentile | Interpretations |
|---|---|
| 95th | The score concerned exceeds 95% of the scores. |
| 90th | The score concerned exceeds 90% of the scores |
| 75th | The score concerned exceeds 75% of the scores. |
| 60th | The score concerned exceeds 60% of the scores. |
| 50th | The score concerned exceeds 50% of the scores. |

Table 9.5: Percentiles and their interpretations

To visualize differences in frequency distribution, you can display frequency distribution in a graphic form. The common forms include a bar graph, a polygon or a pie chart. For example, the frequencies in Table 9.1 can be shown in figures (See Figure 9.1, Figure 9.2 and Figure 9.3).

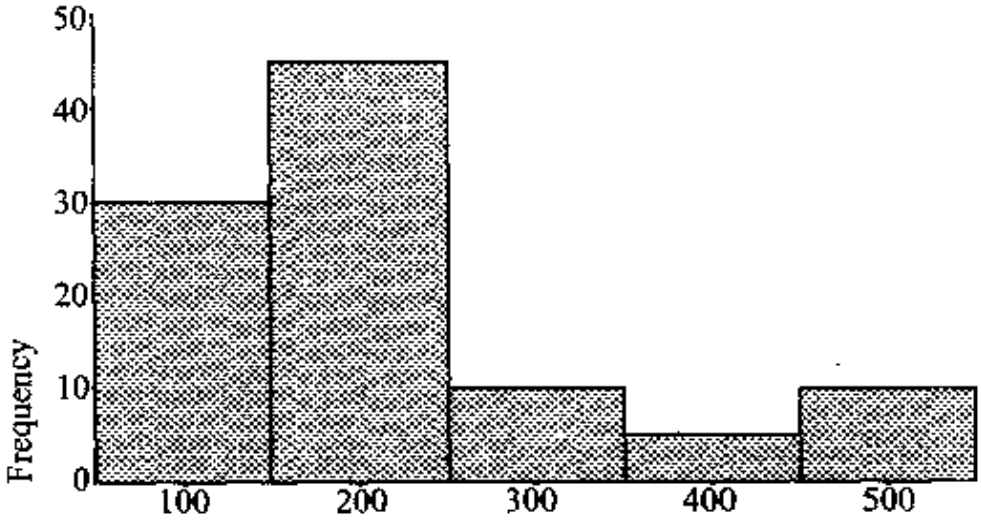



Figure 9.1: A bar graph

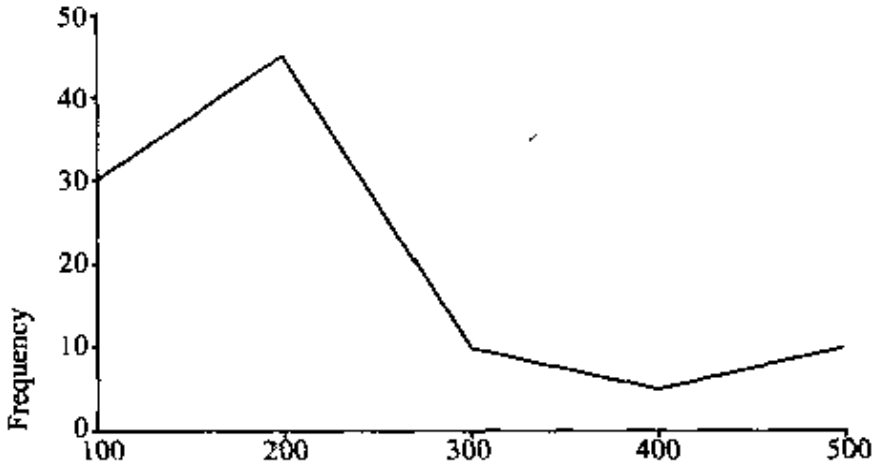



Figure 9.2: A polygon

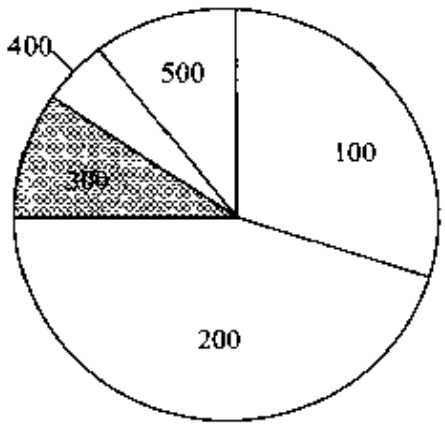



Figure 9.3: A pie chart

## Central tendency

With regard to central tendency, you may describe it in terms of mode, median and mean.

The mode is the most frequently occurring number in the data. For example in the following data set the mode is 84:

58, 60, 65, 70, 80, 84, 84, 84, 84, 90, 92

Sometimes you may have two modes if there are two values which occur equally frequently. For example, the following set of data has two modes, i.e. 65 and 84:

58, 60, 65, 65, 65, 65, 84, 84, 84, 84, 90

Using a mode to describe a central tendency is the easiest since it doesn't require any calculations. However, it has an obvious limitation, i.e. it is very sensitive to accidental scores. Let's look at the following example.

The mode of the scores: 58, 64, 65, 65, 65, 85, 85, 86, 86, 88, 90 is 65. However, if one of the students who scored 65 just by chance had scored

64 while one of the students who scored 86 had scored 85 instead, then the mode would change from 65 to 85. The difference would be 20. Therefore, we have to take such a limitation into serious consideration.

The median is the value which serves as a mid-point of a series of numbers. To find out the median score, you first arrange the given scores in rank order. If the number of scores is odd, the value at the center is the median. If the number of scores is even, you use the mid-point between the two middle scores as the median. The median score is often used as the measure of central tendency when the number of scores is small and/or when the data are obtained by ordinal scale.

The mean is the most frequently used indicator of central tendency because it is better than mode and median in the sense that it takes all the scores into account. The formula for calculating the mean is:

$$=\Sigma x/N$$

In the above formula, x-bar is the symbol for the mean of a sample; $\Sigma$ means to add up or to sum all the individual observations of x; N, to the total number of observation. In other words, to calculate the mean, you have to add up all the individual scores and then divide the sum by the total number of scores. "In any case, the mean is the best, most practical, most useful measure of central tendency" (Hatch & Farhady, 1982: 55). However, the mean is not appropriate with nominal and ordinal data.

## Variability

Once you have decided on your measure of central tendency and found tbe most typical value, there are some reservations to keep in mind about the typical score. Suppose we look at the mean scores of two different classes and they both turned out to have the same mean score. Does this imply that the two classes had the same performance? No, of course it doesn't. The variability among the values, i.e. how they spread out from the central point, may be quite different in the two groups.

Therefore, to be able to talk about data more accurately, we have to measure the degree of variability of the data apart from the measure of central tendency. Just as there are three ways of talking about the central tendency in the data, there are three major ways, too, to show how the data are spread out from the middle point: range, standard deviation, and variance.

### *Range*

The easiest way to talk about the spread of the distribution of scores is the range. Suppose in addition to the mean score of your class on the final exam, you also told the head of your department the range of the score. This would give him/her an idea of the spread of the scores. To calculate the range, you first arrange the values from the highest to the lowest and then subtract the lowest score from the highest score.

The weakness in using range as an indicator of variability is that it is easily affected by the extreme scores. For example, if one student in your class scored zero, then the range would be drastically increased just by that one particular score. Therefore, it is not a precise measure and is thus rarely used in a formal study.

### *Standard deviation*

The most frequently used measure of variability in a formal study is the standard deviation (SD). Before I explain what the standard deviation is, let me first give you an example. When the mid-term exam of English was over, by calculation, you told your students the mean score of the exam was 82. Naturally, as an individual student, s/he is concerned with the difference between his/her scores and the mean. Suppose Student A got 90 and thus her score was 8 points above the mean; Student B scored 65, and thus was 17 points below the mean. Then such a difference is the deviation from the mean. As researchers, our coucern goes beyond an individual deviation. What we want to know is the average deviation of all scores from the mean. The word "standard" is used simply because it

looks at the average variability of all the values around the mean. Then how do you calculate a standard deviation? You may simply follow the steps listed below:

1. Calculate the mean;
2. Subtract the mean from each score to get the individual deviation scores: x= X- (Little x stands for the deviation of the individual score from the mean);
3. Square each individual deviation and then add them up;
4. Divided by N-1: $\Sigma x^2/(N-1)$ (Mathematicians have determined that it is more accurate to divide hy N-1 with a small sample);
5. Take the square root of the result.

Consider, for example, the scores of 10 Ss on a dictation: 10, 9, 8, 8, 7, 7, 7, 6, 4, 4. The mean is 7. The following table shows you how to get the standard deviation of all the scores:

| X (Scores) | X (individual deviations) | $x^2$ |
|---|---|---|
| 10 | 3 | 9 |
| 9 | 2 | 4 |
| 9 | 2 | 4 |
| 8 | 1 | 1 |
| 7 | 0 | 0 |
| 7 | 0 | 0 |
| 6 | -1 | 1 |
| 6 | -1 | 1 |
| 5 | -2 | 4 |
| 3 | -4 | 16 |
| $\Sigma X=70$<br>N=10<br>Mean=7 | $\Sigma x=0$ | $\Sigma x^2=40$<br>$\Sigma x^2/N-1=4.44$<br>SD=2.11 |

If you are not good at calculation, you don't need to worry about it since with the help of the computer program, you can obtain the standard deviation of values without any difficulties.

What can the standard deviation tell us? We say that the standard deviation can tell us the average deviation of all the scores from the mean. The larger the standard deviation, the more deviant the scores away from the central point in the distribution; the smaller the standard deviation, the closer the scores to the central point. Suppose you were a new English teacher and the head of the department told you that you could select one class from the three of Band 3 to teach. Table 9 .6 presents the mean score and standard deviation of their scores on the final exam in the previous semester.

| Class Number | Means | Standard deviation |
|---|---|---|
| Class 1 | 83.4 | 5.2 |
| Class 2 | 82.3 | 5.0 |
| Class 3 | 84.1 | 8.4 |

Table 9.6: Descriptive statistics about the exam of the three classes

Based on the information provided in the above table, how can you make a better decision? If you simply look at their mean scores, Class 3 got the highest and Class 2, the lowest. However, the differences among their mean scores are rather small. Then you look at the standard deviations. You immediately find that the scores for Class 3 are much more widely spread than those for Class 1 and Class 2. What kind of class do you want to teach? Do you want a more homogeneous class or a more heterogeneous class? The larger the standard deviation, the more heterogeneous the class is; the smaller the standard deviation, the more homogeneous the class is. In this sense, the standard deviation gives you information which the mean score cannot. Therefore, in our data-

analysis, so we need to calculate both the mean score and the standard deviation.

### *Variance*

Variance is one of the three measures of variability. It is defined as "the sum of the squared deviation scores divided by *N*-1" (Hatch & Farhady, 1982: 60).

$$\text{Variance} = \Sigma(x - \quad)^2/N\text{-}1$$

You must have noticed that the standard deviation is the square root of variance. Thus, to obtain a standard deviation, variance is computed first.

## INFERENTIAL STATISTICS

Different from descriptive statistics which are only concerned with describing the sample data, inferential statistics go beyond the sample data to predict to what extent the sample can represent the population from which it is drawn. Before we discuss the details of inferential statistics, we must understand two important concepts: normal distribution and probability.

### Normal distribution

Before we come to a theoretical explanation of what is normal distribution, let us draw graphs to present the frequency distribution of two data sets listed below:

Data Set One:
80, 81, 81, 82, 82, 82, 83, 83, 84

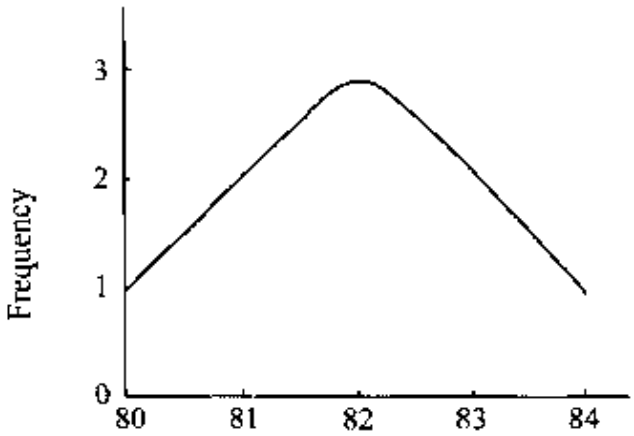



Figure 9.4: The polygon to represent Data Set 1

Data Set 2:

100, 99, 99, 99, 99, 98, 98, 98, 98,98, 98, 98, 98, 98, 98, 98, 98, 98, 98, 98, 98, 98, 97, 97, 97, 97, 96

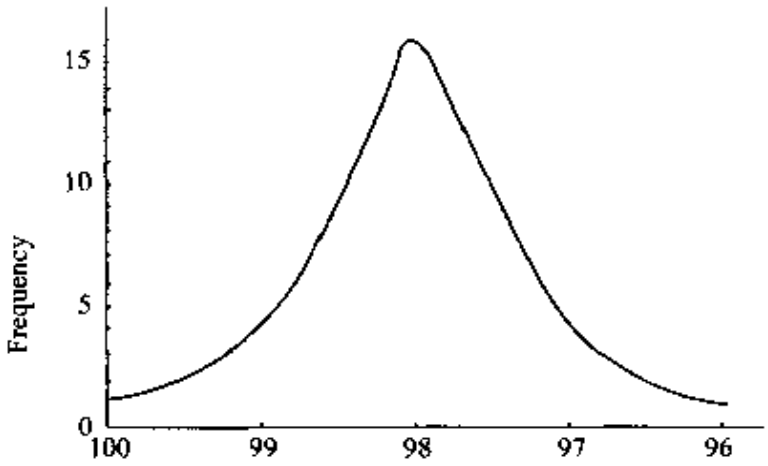



Figure 9.5: A polygon of Data Set 2

I think by drawing the polygons yourself, you must have got some idea of the shape of a normal distribution, that is symmetric. You must also have noticed that tbe two polygons are somewhat different: the first one has a much lower peak while the second one, a much higher peak. However, they are both normally distributed since they both possess the following three distinct properties:

1. The Mode, the median and the mean are the same.

2. The range is symmetrical. The lowest value and highest value have exactly the same distance from the mean.
3. The normal distribution does not have a zero score; the two tails never meet the X axis.

Normal distribution is a mathematical concept which is used to describe an ideal distribution. However, it rarely occurs in reality. If a sample is extremely large, the distribution you ohtain will be very close to the normal distribution.

## Probability

What is probability? It is difficult to define but it is easy to understand. Let's start with some examples. Every year before the Spring Festival, all the shopping centers try their best to promote their sales. One of the measures taken is to offer prizes to the winners. I remember once the first prize was a new flat. I could not resist such a temptation and like many other customers, went there to buy goods. Eventually, it turned out that there were altogether 250,000 customers. Then what is the probability that I could get the first prize? That is one out of 250,000. It is almost next to impossible for me to get the first prize. In an examination, very often there are multiple choice questions. If each question has four choices and there is only one correct answer, then what is the probability of guessing the right answer? 25%. If each question has five choices, the probability of guessing the right answer is 20%.

The formula for assigning probabilities is the following:

$P = n_e$ (number of desired events)/N(number of possible outcomes)

The desired outcome n can never be greater than N. The highest value of any probability is 1. If the probability of an outcome occurring is 1, it means that the event is 100% certain to occur. The smallest value of any probability is 0, meaning a certain event is 100% sure not to occur.

Therefore, the value of any probability is never negative and is never bigger than I. The value falls between 0 and 1. Probability value may be converted to percentages by multiplying by 100. For example, we may say that there is a 50% chance of guessing a correct answer in true or false questions. Or we say, the probability of guesing a correct answer in true-or-false questions is .50.

## Their practical uses

Normal distribution together with probability can help us decide the following two things:

1. What is the probability that a certain score can occur? What proportion of the values falls between a certain range?
2. Can a sample drawn from the population represent the population?

Since the normal distribution is symmetric, and the mean, median and mode are the same, half the scores fall above and half below the mean. Mathematicians further describe the probability of getting a score in Figure 9.6 (Nunan, 1994: 31):

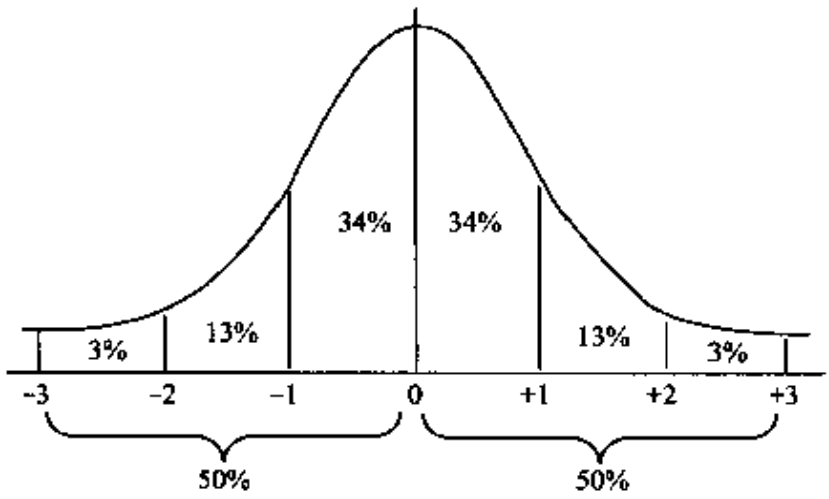



Figure 9.6: Proportions of standard deviations in a normal distribution

According to the above figure, we may say that the scores within the range of ±1 SD have the area of 68% (34%+34%). In other words, the probability of getting a score within the range of ±1SD is 68%. Similarly, we may say that the probability of getting a score within ±2SD is 95% and the probability of getting a score within ±3SD is about 99.75%.

With this knowledge about the probability and normal distribution, we may know whether the sample we studied can represent the population from which it is drawn. If the sample can represent the population, that means that there is no significant difference between the mean of the sample and the mean of the population. When can we say there is no difference between the mean of the sample and the mean of the population? Actually, we are testing a null hypothesis.

If your sample falls in the area of the 95% probability, then your sample is the same as the population. Thus the null hypothesis cannot be rejected.

If the sample falls in the area of the 5% probability, then your sample is different from the population. In other words, there are five chances in 100 that we are wrong. Or we say, we can be 95% confident that the difference between the sample and population is significant

The problem is that very often we cannot get the mean and standard deviation of the population. What we can do is to get as many sample means as we can and then use the sample mean to estimate the population mean.

## The process of hypothesis testing

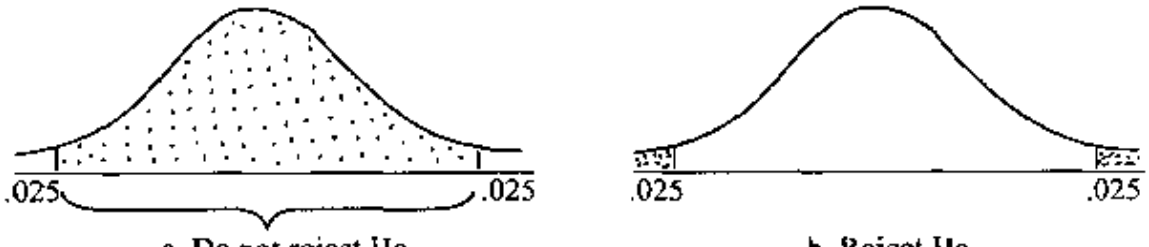
In research, you are usually anxious to discover the probability of your hypotheses about research outcomes. To do this, you first must state the hypotheses in a way that allows you to compare the sample data with that of the population from which the sample is drawn. To be specific, you need to set up a null hypothesis and an alternative hypothesis. The next decision you need to make is the significance level at which you are

allowed to reject the null hypothesis. Significance levels are usually .05, .01 or .001 in the field of applied linguistics.

Let's assume that you have stated a null hypothesis, and that you have selected the level of significance of .05 for rejecting the null hypothesis. If the results fall within the area of 95% as shown in Figure 9.10a, the null hypothesis cannot be rejected, for the scores are typical of those that would normally be found in such a distribution. On the other hand, if the sample data fall in the area shaded in Figure 9.10b you have to reject the null hypothesis since the scores are not those typically found in the distribution. If they fall in the lower left tail, the sample is worse than the population. If the data fall in the far right tail, the sample is much better than the population. In either case, you can reject the hypothesis that the sample is not different from the population from which it was drawn.

Similarly, if the chosen level of significance is .01, the null hypothesis cannot be rejected when the results fall within the area of 99%. If the selected level of significance is .001, the null hypothesis cannot be rejected when the results fall within the area of 99.9%

You will notice in Figure 9.10 that, although you have chosen .05 as a level of significance, you have to divide the .05 into two, giving an .025 area to each tail of the distribution. This is because the null hypothesis you formulated did not specify where the difference might be. That is to say, when there is no direction specified for possible differences in the hypothesis, you must consider both tails of the distribution. This is, therefore, called a two-tailed hypothesis.




a. Do not reject Ho b. Reject Ho

Figure 9.7:Rejecting or not rejecting $H_0$ (Hatch & Farhady, 1982: 87).

## SUMMARY

Descriptive statistics and inferential statistics are two branches in statistics. Descriptive statistics is concerned with the description of a group while inferential statistics aim at drawing inferences about the entire population from the group investigated. A parameter is used to describe a population and statistic to describe a group. This distinction is necessary only when the difference between a population and a group is essential. Parameters are denoted by Greek letters while statistics by Roman letters. Descriptive statistics are used to organize the data in terms of frequency, central tendency and variability. Two important concepts underlying inferential statistics are normal distribution and probability. By having these two concepts, we can test null hypotheses. If the probability of getting the result concerned is lower than .05, the null hypothesis can be rejected and the alternative one can be accepted accordingly. This probability level is also called the significance level and can be varied from .05 to .001 in the field of applied linguistics and many social sciences.

## DISCUSSION QUESTIONS

1. What are the differences between descriptive statistics and inferential statistics?
2. The following are scores on TEM-Band 4 obtained by 20 second-year English majors: 50, 60, 70, 75, 55, 60, 65, 70, 80, 85, 55, 60, 80, 80, 65, 45, 70, 65, 70, 75
   1) Divide the scores into three groups, and then present the results in a table and in a graphic form respectively.
   2) Divide the scores into four groups and then present the results in a table and in a graphic form respectively.
3. The following are scores on the composition of TEM-Band 8 obtained by 20 fourth-year English majors: 12, 14, 10, 12, 16, 18, 10, 12, 16, 18, 14, 12, 16, 10, 12, 14, 10, 18, 10, 12

3) Find out the mean, median and mode of the scores

4) Find out the range, the variance and the standard deviation of the scores

# 10. The analysis of quantitative data

As was discussed before, the quantitative data can be analyzed descriptively or inferentially depending on your research purposes. In this chapter, I will first give you a brief description of the statistical package (SPSS) and then focus on the analysis of questionnaire data from a survey study and an experimental study.

## OVERVIEW OF SPSS FOR WINDOWS

SPSS is a comprehensive and flexible statistical analysis and management system which is, however, easy to manipulate (Norusis, 1993). Most tasks can be accomplished simply by pointing and clicking the mouse. In this section you will have an overview of this statistical package.

### Three main windows

Once you start SPSS program, you may see three windows as shown in Figure 10.1 (SPSS Inc., 1993: 2):

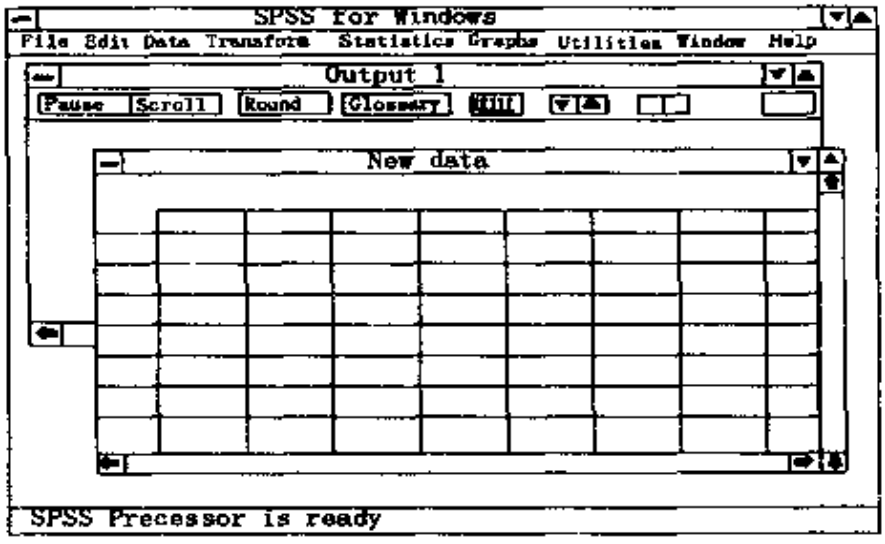



Figure 10.1: SPSS application, output, and Data Editor windows

The first window is the SPSS application window that contains the menu bar. You use it to open files, choose statistical procedures, and select the functions of the system. The second window is Output Window which displays results yielded by the statistical procedures such as frequency distributions, means and standard deviations. You can edit this output and save it as ".lst" file for later use. The third window is the Data Editor Window through which you can create a new file or read the existing file.

## Menus in the application window

SPSS for Windows is operated by simply using the menus shown in the application window. There are altogether nine menus in the main menu bar:

**File** The File menu is used to create a new SPSS file or read an existing file. It can also be used for saving an old file after editing or a new file. If you are familiar with the File menu in Microsoft Word, then there will be no problem for you to use the File menu in SPSS.

**Edit** The Editor menu is also like the one in Microsoft Word, which is for modifying or copying text from the output window.

**Data** The Data menu is for defining a variable, inserting a variable, merging files, spliting files, selecting cases for analysis, etc.

**Transform** The Transform menu is to make changes in the data file, such as computing a new variable based on the values of existing variables, recoding the variable into a new variable or recoding the same variable by changing the values systematically.

**Statistics** The Statistics menu is for you to choose the various statistical procedures for your own purpose. You may select correlation, compare means and linear regression.

**Graphs** The Graph menu is to produce various kinds of graphic representations.

**Utilities** The Utilities menu is to change font, display the information about the variables, or open an index of SPSS commands.

**Window** The Window menu is to arrange, select, and control the various SPSS windows. For example, if you want to change Data Editor Window into Output window, you may simply click the Window menu and then select the Output Window.

**Help** The Help window contains information on how to use the different menus properly.

## QUESTIONNAIRE DATA ANALYSIS

I will use my own study[12] as a live example to illustrate the details of the analysis of questionnaire data.

### Brief description of my own study

The study was conducted to find out how second-year English majors study English and to what extent their way of learning is related to their learning outcomes.

In September, 1997, 77 second-year English majors from three intact classes at Nanjing University were asked to answer the Language Learner Factors Questionnaire. The questionnaire was administered during class time and the response rate was 100 percent. Among the 77 subjects, 59 were females and 18 males. The ages ranged from 18 to 20 with the average being 19.4. They took a Band 4 English Proficiency Test for English Majors in May 1998.

The Language Learner Factors Questionnaire consists of three parts. Part A is divided into two subparts. The firsr one concerns personal details (that is, name, sex, date of birth, scores on the nation-wide English and Chinese matriculation tests, and an estimate of time spent in studying English outside class, which was used as the indicator of "effort" ). The other section of Part A establishes students' reasons for learning English. Students rated reasons listed on a five-point scale from "not important"

[12] The study described here is part of a big project.

(1) to "extremely important" (5). Part B consists of statements of beliefs about language learning. The students indicated their opinions in terms of a five-point scale from "strongly disagree"(1) to "strongly agree" (5). Part C contains statements concerning learning strategies. Some of them are management strategies, and the rest are language learning strategies. Students again responded on a five-point scale from "This statement is never or almost never true of me" (1) to "This statement is completely or almost completely true of me" (5).

| Variable Name | Brief description of variables | No of Items | Alpha |
|---|---|---|---|
| (1) English Achievement | Scores on Band 4 Test for English majors. | 1 | |
| (2) Sex | Biological sex | 1 | |
| (3) L1 Proficiency | Scores on the National Chinese Matriculation Test | 1 | |
| (4) L2 Proficiency | Score on the National English Matriculation Test | 1 | |
| (5) Effort | The self-estimated amount of time spent by learner outside class in studying English | 1 | |
| (6) Learning Purpose | Reasons for learning English | 3 | |
| (7) Effort belief | Views about the importance of effort in language learning | 5 | .51 |
| (8) Management Belief | Views on the importance of planning, setting goals, evaluating progress, etc. | 5 | .75 |
| (9) Traditional Belief | Opinions on the importance of repetition, memorization and intensive study of texts | 4 | .59 |
| (10) Nontraditional Belief | Opinions on the importance of extensive exposure to and communicative use of the target language | 4 | .64 |
| (11)Using-mother-tongue Belief | Opinions on the necessity of using the mother tongue in reading, listening, speaking and writing. | 4 | .69 |
| (12) Management Strategy | Actions in planning, goal setting, evaluating, etc. | 8 | .79 |
| (13) Traditional Strategy | Actions in traditional activities | 11 | .73 |
| (14) Non-traditional Strategy | Actions in non-traditional activities | 12 | .75 |
| (15) Using-mother-tongue Strategy | Actions in using the mother-tongue in L2 learning | 5 | .69 |

Table 10.1: Descriptions of 15 variables

There are altogether 15 variables listed in Figure 10.6. Among them, English Achievement is a dependent variable while others are independent variables. The first five variables only have a single indicator and thus they do not need a statistical check, whereas the remaining 11 are variables with multiple indicators. The alpha value indicates the internal consistency within the multiple items (Note: The meaning of the alpha value will be discussed in the next section "Categories confirmation")

## Setting up a data file

Once you have collected questionnaires, you are going to set up a data file by coding the data, numbering all the questionnaires, inputting the data and cleaning the data.

### *Coding the data*

Coding data is not a major stage of research, but it is indispensable. For instance, the fact that a subject is male or female cannot be automatically converted into the forms that can be computed by statistical devices. The solution is that the researcher has to select a numerical code to represent a subject's gender. You may choose 1 for a male subject while 2 for a female subject. The numerical code here is just another piece of datum which a statistical device can easily record, store, process, and retrieve. What number should be used in coding the data depends on the way a variable is measured. For nominal data, what number is used is arbitrary. But for ordinal and interval data, the numbers used in coding do have mathematical values and so you cannot assign them arbitrarily

For coding, you need a codebook. A good codebook can give you a lot of help. It tells you and other researchers what you have in your data — what variables you've investigated, what you have called these variables, and the meaning of different numbers in your data file. Without a clearly-defined codebook, other researchers cannot understand your data and even you cannot make sense of your own data several

months later because the human memory is not always reliable. What does a codebook look like? Table 10.2[13] displays some information in the codebook for my study mentioned above.

| Complete name of a variable | Coded variable name | Meaning of values |
|---|---|---|
| Sex | Sex | 1 = Male, 2 = Female |
| Learning Purpose | LP | 1 = Not important, 2 = quite important,<br>3 = important, 4 = very important,<br>5 = extremely important |
| Management Belief | MagBEL | 1 = strongly disagree,<br>2 = disagree, 3 = neutral,<br>4 = agree, 5 = strongly agree |
| Using-mother-tongue Belief | UMTBEL | 1 = strongly disagree, 2 = disagree,<br>3 = neutral, 4 = agree, 5 = strongly agree |
| Management Strategy | MagStr | 1 = Never, or almost never true of me,<br>2 = Usually not true of me,<br>3 = Somewhat true of me,<br>4 = Usually true of me,<br>5 = Completely true or almost completely true of me |

Table 10.2: An illustration of one page in the code-book

A code-book should contain the following three pieces of information:

1. A complete description of each variable;
2. A coded name for each variable, which should have no more than 8 characters[14];
3. The meaning of every possible value that each variable can take.

---

[13] This kind of table is often presented in the thesis as an appendix.

[14] The SPSS does not accept a variable with more than eight characters

### *Numbering all the questionnaires*

Before you input the data, you should carefully number all the questionnaires. That is to say, you have to assign a unique number to each questionnaire. For your convenience, you had better use sequential numbers, i.e. 1, 2, 3, 4, 5 and so on. When you input the data, this number will be entered in the first column as ID to represent each respondent (See Table 10.3). At the stage of cleaning data when you find a mistake, the number will help you to locate the original questionnaire.

### *Inputting the data*

Some use machines to read the data when the responses are recorded on a standardized answer sheet. However, in the Chinese context, almost all the postgraduate students still input the data manually. In the following part, you will learn how to input the data by using the Data Editor Window in SPSS.

Once you start an SPSS program, you see the Data Editor window on the screen. It is a spreadsheet on which rows are cases, and columns are questionnaire items. Each cell shows a single value of a variable for one case.

The data inputting should start with defining a variable. You click the Data menu where "Define variable" is on the top list. Then you open the Define Variable dialog box as shown in Figure 10.2 (SPSS Inc., 1993:57).

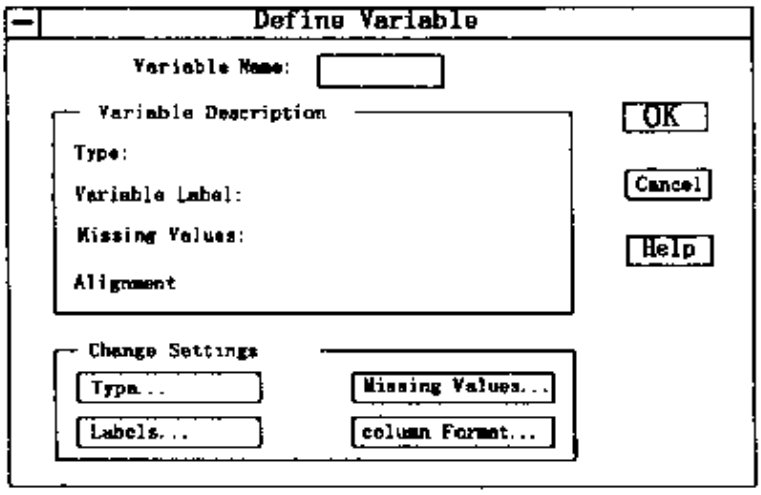



Figure 10.2: Define Variable dialog box

You should undertake one by one the following tasks which are specified in the Define Variable dialog box .

**Specify the name of each variable** The name must begin with a letter. The remaining specifications can be letters or numbers. The length of the name cannot exceed eight characters. All the names of variables should be mutually exclusive since duplication is not accepted. Finally, the name of a variable should be easy to recognize. As Table10.3 shows, ID stands for Identity number, P01 for the first questionnaire item in the section of Learning Purpose, B01 for the first questionnaire item in the section of Beliefs and S10 for the tenth questionnaire item in Strategy Section.

| ID | Sex | P01 | P02 | B01 | B02 | S01 | S02 | S03 | S43 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 4 | 5 | 2 | 4 | 3 | 4 | 4 | 5 |
| 2 | 2 | | | | | | | | |
| 3 | 2 | | | | | | | | |
| 4 | 1 | | | | | | | | |

Table 10.3: A spreadsheet of the data file

**Specify the variable type** By default, SPSS regards all new variables as numeric. Actually, our data are numeric and so we don't need to do anything about them.

**Describe labels** You double click Labels in the Define Variable Dialog Box where you give a full description to each variable and the meaning of the values. If you think your codebook is good enough for you to find out the information, then you can skip over the describing of labels.

**Specify missing values** When you come across a case where the respondent did not answer an item or gave an ambiguous answer, you have to treat them as missing values. Although some variable may not have missing values at all, you have to assume that there are missing values for all the variables and also decide what nutuber to use to

represent a missing value. Normally speaking, you can use "0" to represent a missing value. However, "0" may sometimes be meaningful as in the case of test scores where you may use "999" to represent a missing value.

Once you have defined all the variables, you can enter data case by case. You had better find someone to read out the numbers for you while you are keying them in. Furthermore, you had better have a short break once you feel tired. Very often you don't want to stop at this moment because you are eager to get the job done. However, by doing so, the efficiency of inputting is usually low and many errors occur at this time according to my own experience. Finally, I strongly suggest that you should make a copy of your data file before you start any analysis. The original data file should be kept in a safe place, separate from your working data file.

### *Cleaning the data*

By cleaning the data, we undertake two tasks: (1) checking whether there are errors in data entry and (2) dealing with missing responses. Errors must be corrected because incorrect data values distort the results of statistical analyses and contaminate all of the analyses. Some errors are easy to spot. For example, forgetting to claim a value as missing, using an invalid code, or entering the value 585 for age will be apparent from a frequency table. By a five-point Likert scale, the possible responses must be numbers 1 through 5. If the data file contained 34, 45, 12 as responses, they must be wrong. Sometimes, the number may not be totally impossible but rather unusual. For example, one student responded to the question "How many hours do you spend learning English outside class every week" by writing down 100. Although such a value is not totally impossible, it raises suspicion and should be examined to ensure that it is really correct. Other errors, such as entering an age of 62 instead of 26, or entering 5 instead of 3 may be difficult, if not impossible, to detect.

Missing responses refer to values of a variable that are unknown, either because respondents did not answer them at all or because they provided ambiguous answers. If the percentage of missing data is less than 20%, the following two options are available for the treatment.

Substitute a neutral value. Here a neutral value refers to the mean score of the item concerned. Such an substitution does not change the mean of the item and is helpful for further analysis. For example, you want to put together items 2, 12, 8, 10 and 23 to form a category. Unfortunately, two subjects did not answer Item 2 and thus they have a missing value for this item. In the process of forming a category, these two subjects will be automatically excluded by the SPSS Program from the sample for this category because of the missing value of one item. If we use a mean to replace the missing value of this item, the situation will be different. That is to say, these two subjects will be treated the same as other subjects.

Pairwise deletion. In pairwise deletion, cases involved in each calculation are only those who provided complete responses. In other words, those who did not respond to the item are excluded in analyzing that particular item. As a result, the cases involved in each calculation may be varied from item to item. This procedure is appropriate only when the sample size is relatively large.

## Categories confirmation

In the section on "Questionnaire design process", you are told there are two kinds of questionnaire: individual-item-based and category-based. In the latter case, you need to employ statistical procedures to confirm the categories which you justified conceptually. Specifically speaking, you need to make sure that all the items within the category are in the same direction and the category holds up by running the item-total correlation.

### *Recoding the responses if they are in the opposite direction*

By saying that items are in the opposite direction, I mean that these items are contradictory in meaning. Let's look the following two sentences.

(1) Accuracy is more important than fluency in L2 learning.

(2) Fluency is more important than accuracy in L2 learning.

In category confirmation, the first thing you have to do is to make sure that all the items within a category are not in the opposite direction. If they are, you have to make them in the same direction by transforming the values of some items. In my questionnaire I formed a category called Effort Belief which refers to the students' opinion about the important role of effort in L2 learning. The category includes the following five items and the students were asked to respond to each item on a five-point scale (1= strongly disagree; 2= disagree; 3=neutral; 4= agree; 5= strongly agree)

(1) Learning a foreign language requires painstaking effort.

(2) A person with a high language aptitude can grasp a foreign language without hard work.

(3) Learning English well is due to one's inborn ability with language.

(4) Without hard work, no one in China can learn a foreign language successfully.

(5) Effort is more impertant than inborn ability in learning a foreign language.

Let's examine the first three items

(1) Learning a foreign language requires painstaking effort.

(2) A person with a high language aptitude can grasp a foreign language without hard work.

(3) Learning English well is due to one's inborn ability with language.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Strongly Disagree | Disagree | Neutral | Agree | Strongly agree |

Logically speaking, if you strongly disagree upon the first item, you would strongly agree upon the second and third items. That is to say, a 1 on the first item is a 5 on the second and the third, and vice versa. In this sense, Item 1 and Items 2 and 3 are not in the same direction.

What should you do? Since the category which includes the five items is called Effort Belief, responses to Items 2 and 3 have to be recoded. 1 is changed into 5, 2 to 4, 5 to 1 and 4 to 2 so that all the five items are in the same direction in the sense they all emphasize the importance of Effort. Otherwise, they are in conflict since Items 1, 4, and 5 advocate the important role of effort while Items 2 and 3 emphasize the role of language aptitude.

How do you recode a variable? You just double click the Transform menu of SPSS where you can find "Recode". There are two options for you to choose: (1) Recode into the same variable and (2) Recode into a different variable. To be safe, you had better select the second option because if you make a mistake, you will lose your data.

## *Checking internal consistency by the item-total correlations*

Next we have to see whether the items within the category are internally consistent by the item-total correlation. What does the item-total correlation mean? Let's look at an example. The hypothesized category Management Belief contains five items and I tested these items on 77 second-year English majors.

1) Planning your study time is important for success.

2) Selecting appropriate learning strategies is important for success in learning a foreign language.
3) Setting your learning goals and evaluating your progress is important for success.
4) It is important for success that you frequently review your progress in learning English to find out your weak areas.
5) Reflecting upon whether your learning strategies are effective or not is important for success in learning a foreign language.

My data look like this:

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|---|---|---|---|---|---|
| Person 1 | x | x | x | x | x |
| Person 2 | x | x | x | x | x |
| Person 3 | x | x | x | x | x |
| Person 4 | x | x | x | x | x |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| Person 77 | x | x | x | x | x |

Table 10.4: A description of the data file

In Table 10.4, the x's are the scores for each person on each item. For five items, scored from 1 to 5, each person could get a score as low as 5 or as high as 25. In practice, of course, each respondent in a survey will get a total score somewhere in between. With 5 items, the total score gives you an idea of where each person stands on the category you're trying to measure. Then you run the item-total correlation. Here the total score does not refer to each individual's score but to the total scores of 77 respondents and similarly the score for each item is not one person's score, but 77 respondents'.

To run the item-total correlation, you may use Statistics Menu where you can fmd "Scale" on the list. Once you highlight "Scale", there are two choices for you to choose: (1) reliability analysis and (2) multidimensional scaling. You just click the first one. Table 10.11 presents the results yielded by the scale reliability analysis. Since the value of Alpha is as high as .7542, the proposed category "Management Belief" is confirmed. In other words, this category can be employed for further statistical analysis.

*Item-total Statistics*

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Alpha if item Deleted |
|---|---|---|---|---|
| Item 1 | 17.3247 | 3.5906 | .4758 | .7258 |
| Item 2 | 17.0000 | 3.4474 | .6147 | .6833 |
| Item 3 | 17.4156 | 3.3513 | .4343 | .7472 |
| Item 4 | 17.2468 | 3.1883 | .6588 | .6607 |
| Item 5 | 17.4545 | 3.3301 | .4658 | .7332 |

Reliability Coefficients

N of Cases = 77.0 N of Items = 5

Alpha = .7542

Table 10.5: The item-total correlation of Management Belief

All the proposed categories need to go through such statistical confirmation. Very often we are not lucky. Although conceptually the

items within the category hang together, they are not necessarily supported by empirical data. In a worse case, the whole category has to be thrown out. But more often than not we simply need to make some modifications such as deleting some items or rearranging some items.

Then how can you make modifications? In this case, you may look at the last column "Alpha if item deleted", which can give some information to you so that you can decide which item is better thrown out to make the category internally more consistent. In the case of Management Belief, no further modification is needed since deleting any item will reduce the degree of internal consistency.

Now let's look at another example shown in Table 10.6 in which four items form a category "Using-mother-tongue Belief". These four items cover the students' opinions about the use of L1 in L2 reading, listening, writing and speaking:

(1) To understand a text well, the best way is to translate it.

(2) The best way to memorize what you have heard in listening is to keep in your memory what you have heard in Chinese.

(3) To write well in English, the best way is to organize ideas in your mother tongue first.

(4) When speaking English, the best way is to think what you want to say in Chinese first.

According to the last column of Table 10.6 "Alpha if item deleted", if Item 4 is deleted, the alpha value will be increased to .6661. In contrast, if Item 3 is deleted, the alpha value will be dropped to .4988. To make a decision on deleting or keeping one item is not simply a matter of looking at the statistical figures. The alpha value is just one factor. We also need to consider the content of each individual item. In the case of Using-mother-tongue Belief, although the deletion of item 4 may increase the alpha value, this item describes the use of L1 in a particular area, i.e. the use of L1 in L2 speaking and thus it is better to keep it. Of course, if the

cost of keeping Item 4 is so high that the alpha value is too low for us to accept the whole category, then we have to remove this item without too much hesitation.

Item-total Statistics

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Alpha if Item Deleted |
|---|---|---|---|---|
| Item 1 | 12.7013 | 1.7912 | .3891 | .5902 |
| Item 2 | 12.7143 | 1.6015 | .4770 | .5278 |
| Item 3 | 12.5325 | 1.6733 | .5401 | .4988 |
| Item 4 | 13.2078 | 1.5615 | .3195 | .6661 |

Reliability Coefficients

N of Cases = 77.0 N of Items = 4

Alpha = .6386

Table 10.6: The item-total correlation of Mother-tongue Belief

In addition to the statistical procedure "Scale Reliability", we can also employ a more rigorous procedure, i.e. factor analysis to establish or confirm categories. Suppose your questionnaire contains 30 items and you intend to have five categories. If all these categories are discrete and undimensional, then there will be five distinct factors on which all the items within a category should load high respectively. However, the reality is more complicated than its conceptualization. Therefore, category developers may start with a large pool of potential category items and then ask different groups of people to respond to the items. Factor analysis is run repeatedly based on the responses given by different groups of subjects. Each time, only those items that load high on the

factor are selected. In other words, these categories must be tested repeatedly through factor analysis and in the process of testing, many inadequate items are screened out.

Although we know that forming categories through factor analysis is more rigorous than by checking internal consistency, some categories describing L2 learning process are not identified as distinct factors by factor analysis because these categories are conceptually distinct but empirically overlapping such as Beliefs and Strategies used in my study.

## Selecting appropriate statistics

There are a variety of statistics for you to choose from. Which one is the most appropriate to answer your research questions? You have to make a series of decisions in such a selection process. In the following section, the four frequently used procedures appearing in the Statistics menu of SPSS are introduced: Summarize, Compare means, Correlate and Regression. Before we go to the details, one thing I have to emphasize is that all the statistical procedures except Summarize, are operated under the assumption of normal distribution.

### *Summarize: Frequencies and Descriptives*

Once you open the Statistics menu, Summarize appears on the top of the list which includes a variety of options. Due to the limited space, only the first two options are explained here: Frequencies and Descriptives. A study for writing up an MA thesis or a doctoral dissertation usually goes beyond Frequencies and Descriptives which, however, are the first step of any complex analysis. Using Frequency and Descriptives, you lay the data out and get a "feel" for them.

For example, you may want to know how many students are females and how many are males in a study, what is their respective percentage? By operating Frequencies, you can get the answers to the above two questions. The result shown in the following tahle indicates that there are 18 males occupying 23.4% and 59 females, 76.6 %

When a questionnaire only contains a few items, you may end up with reporting frequencies and/or percentages only.

SEX

| Value Label | Value | Frequency | Percent | Valid Percent | Cum Percent |
|---|---|---|---|---|---|
| | 1.00 | 18 | 23.4 | 23.4 | 23.4 |
| | 2.00 | 59 | 76.6 | 76.6 | 100.0 |
| | Total | 77 | 100.0 | 100.0 | |

Valid cases 77 Missing cases 0

By Descriptives, we may know what is the average score and what is the standard deviation of each category The following results in Table 10.7 generated by Descriptives are the mean scores and standard deviations of Belief categories in my study. You may find the second-year English majors showed a negative attitude toward Using-mother-tongue Belief while having a positive attitude towards the remaining four beliefs. However, they held non-traditional beliefs more firmly than traditional beliefs since the mean score of the latter (3.96) is lower than the former (4.27).

| Variable Label | Mean | Std Dev | Valid N |
|---|---|---|---|
| UMTBEL(Using-mother-tongue Belief) | 2.13 | .56 | 77 |
| EFBEL(Effort Belief) | 3.87 | .48 | 77 |
| TRABEL (Traditional Belief) | 3.96 | .43 | 77 |
| NTRABEL(Non-traditional Belief) | 4.27 | .40 | 77 |
| MAGBEL (Management Belief) | 4.32 | .45 | 77 |

Table 10.7: Descriptive statistics of belief variables

(Note: The full name of a variable is not generated by Descriptives)

Furthermore, you may notice that the standard deviation of UMTBEL is the biggest and that of the NTRABEL, the smallest. It means that the students' views about UMTBEL are more diversified than their views about NTRABEL.

### *Correlation analysis*

Correlation analysis is used to test the significance of the relationship between two variables. By such an analysis, we want to know three things:

(1) How close is the relationship between the two variables concerned? In other words, to what extent can we predict the score of a dependent variable in our sample from the score of an independent variable?

(2) Is the relation due to chance, or does it exist in the overall population to which we intend to generalize? To put it another way, is this relationship significant?

(3) What is its direction? Is it positive or negative?

The concept of direction refers to whether a covariation is positive or negative. For example, the students' scores on English Matriculation Test are positively covaried with their scores on Band 4 Test for English Majors. It means that students who scored higher on English Matriculation Test tend to score higher on Band 4 Test for English Majors, whereas students who got lower scores on the first test tend to get lower scores on the second one. The use of the mother tongue in learning English is negatively covaried with the scores on Band 4 Test for English majors. It means that the more the students use L1 in learning L2, the more likely they are to obtain lower scores on Band 4 Test for English Majors, whereas the less the students use L1 in learning L2, the more likely they are to get higher scores on the test.

All in all, a positive direction refers to the case where two variables are varied in the same direction as shown in Figure 10.2(a) while a

negative direction, to the case where two variables are varied in the opposite directions as shown in Figure 10.3(b).

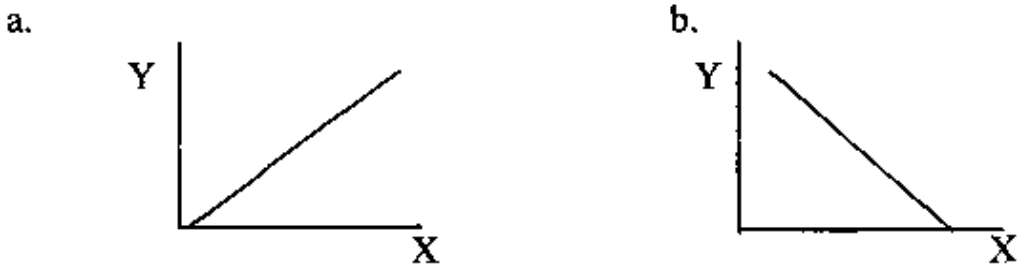



Figure 10.3: Two variables varied in the opposite directions

The extent to which two variables are correlated is indicated by the correlation coefficient which falls between 0 and ±1. If the correlation coefficient is 0, it means that the two variables concerned do not have any relation at all. When two variables are perfectly positively correlated, the coefficient is +1. The opposite case is –1. However, in most cases, the coefficient falls in between 0 and ±1. For example, the total amount of time used in learning an L2 tends to be positively correlated with L2 performance. However, there indeed are some people who work very hard but have very poor performance. Therefore, in most cases, the correlation is not perfect.

A correlation coefficient is said to be significant if it reaches the significant level of 0.05, 0.01.or 0.001. In the case of the .05 level of significance, our confidence level is 95%. In other words, there are five chances out of 100 where the result might be wrong. Similarly, the .01 level of significance is equivalent to the 99% confidence level or this means there is one chance in 100 where the result might be accidental. If the probability is bigger than .05, the correlation coefficient obtained is regarded as insignificant.

In terms of the levels of measurement what kind of variables can be tested for correlation? Generally, the correlation between two interval variables or between one interval variable and one ordinal variable, or between an interval one and a nominal one can be examined (Bernard, 1994).

To run a correlation analysis, you simply open the Statistics menu where by highlighting Correlate, you select Bivariate. The following dialogue box of Bivariate Correlations will be displayed on the screen.

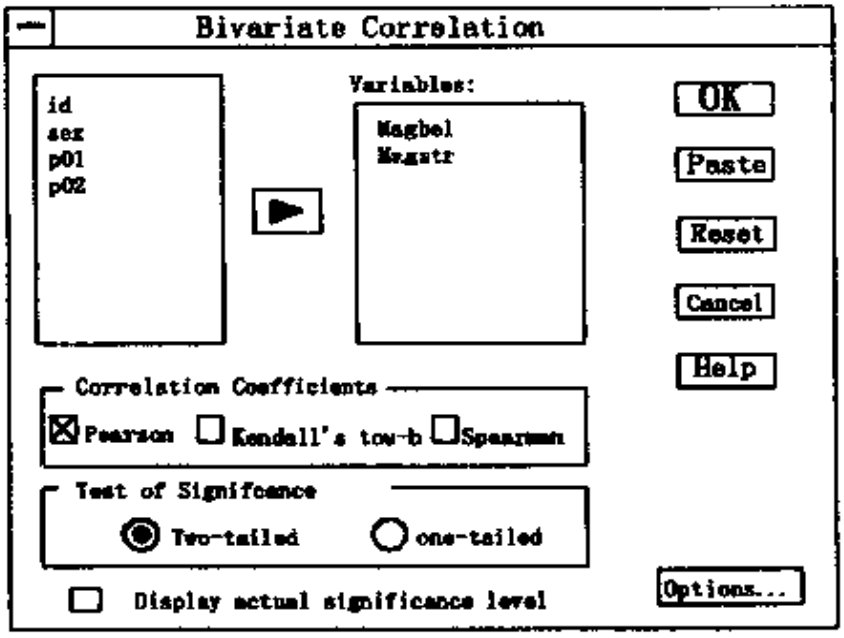



Figure 10.4: The dialogue box of Bivariate Correlations (SPSS Inc., 1993: 298)

You need to move the two variables which are supposed to be correlated from the left to the right under the heading of Variables. Table 10.8 shows the correlation coefficients between the students' beliefs and their use of corresponding strategies in my study. These four pairs of variables are treated as interval ones. According to the results, the four pairs are all positively correlated. However, the correlation between NTRABEL (Non-traditional Belief) and NTRASTR (Non-traditional Strategy) are not significant because its probability (.062) is greater than .05.

| Variable name | No. of cases | Correlation (*r*) | P |
|---|---|---|---|
| MAGBEL/MAGSTR | 77 | .3763 | .001 |
| TRABEL/TRASTR | 77 | .2498 | .028 |
| NTRABEL/NTRASTR | 76 | .2148 | .062 |
| UMTBEL/ UMTSTR | 77 | .5068 | .000 |

Table 10.8: The correlation between belief and strategy variables

When you read and interpret the results, you have to remember the following two rules:

(1) Not all significant findings at the 95% level of confidence are equally important;

(2) The significant, high correlation between two variables does not necessarily mean that there exists a causal relation between them.

### *Compare means: t-tests and One-way ANOVA*

Very often you need to compare two or more means in your study. The statistical procedures in SPSS for this purpose are called "Compare Means". Among them, Independent-samples *t*-test, Paired-sample *t*-test and One-way ANOVA (i.e. One-way analysis of variance) are more extensively used. In this part, I will first explain to you when Independent-samples *t*-test, Paired-sample *t*-test and One-way ANOVA are needed and how results yielded by these different analyses are interpreted. Secondly, I will show you how such analyses can be operated by SPSS.

Let's look at *t*-tests first. For example, in my study I intended to find out (1) whether the mean score of Traditional Strategy and that of Non-traditional Strategy are significantly different and (2) whether female students and male students show a significant difference in their use of Non-traditional Strategy. To address these two questions, we need the *t*-test.

We have two kinds of *t*-test: Independent-samples *t*-test and Paired-samples *t*-test. Independent-samples *t*-test, as the name suggests, involves two different groups of people, such as females and males, or higher

achievers and lower achievers. In contrast, Paired-samples *t*-test only involves one group of people but two different variables, such as Traditional Strategy and Non-traditional Strategy, Intrinsic Motivation and Extrinsic Motivation, etc.

To answer the question as to whether the mean score of Traditional Strategy and that of Non-traditional Strategy are significantly different, Paired-samples *t*-test should be used. Table 10.9 displays the results:

| Variable Name | Mean | SD | T-value | P |
|---|---|---|---|---|
| Trastr (Traditional Strategy) | 3.4749 | .398 | 4.21 | .000 |
| Ntrastr (Non-traditional Strategy) | 3.2982 | .421 | | |

Table 10.9: *t* -test of the use of Trastr and Ntrastr

The results say that our students used Traditional Strategy more frequently than Non-traditional Strategy. The difference is significant at the level of .000.

To find out whether female students and male students showed a significant difference in the use of Non-traditional Strategy, we should use Independent-samples *t*-test, Table 10.10 indicates that males and females did not show any significant difference in the use of Non-traditional Strategy although the mean score of the female group is slightly higher than that of the male group.

| Variable | Number of cases | Mean | SD | T-value | P |
|---|---|---|---|---|---|
| Males | 18 | 3.1713 | .458 | -1.48 | .144 |
| Females | 58 | 3.3376 | .405 | | |

Table 10.10: *t*-test of the use of Non-traditional Strategies by males and females

The difference between Paired-samples *t*-test and Independent-samples *t*-test lies in the fact that the Paired-samples *t*-test operates on the

same group of subjects with two different variables, while the Independent-samples *t*-test operates on two different groups of subjects with one variable.

If you have got more than two means to compare, you need to choose One-way ANOVA which stands for analysis of variance. Suppose we want to know whether students with varied scores on Test for Englisb Majors-Band 4 differ significantly in their use of Non-Traditional Strategy. What we need to do is first to divide the students into three groups according to their scores on Band 4 Test for English Majors: Top group, Middle group and Bottom group, and then compare their mean scores on the use of Non-Traditional Strategy.

The question is: Are the differences in these mean scores significant? Put it anotber way (the null hypothesis), despite apparent differences in the mean scores, are these three groups really from identical populations? Table 10.11 reports the results produced by One-way ANOVA. The information from the table which is important for us to decide whether the difference is significant are F Ratio and F Probability. Since the probahility (.0718) of F Ration (2.7372) is bigger than .05, the differences among the three groups are not significant.

Variable Non-Traditional Strategy
By Variable Band 4 scores

Analysis of Variance

| Source | D.F. | Sum of Squares | Mean Squares | F Ratio | F Prob. |
|---|---|---|---|---|---|
| Between Groups | 2 | 1.2501 | .6251 | 2.7372 | .0718 |
| Within Groups | 69 | 15.7564 | .2284 | | |
| Total | 71 | 17.0065 | | | |

Table 10.11: Differences in the use of Non-Traditional Strategy tested by ANOVA

Let's look at one more example which does show significant differences among groups. In Xu's study (2001) he compared the scores

on a composition written by the students across four grades (i.e. Freshmen, Sophomores, Juniors and Seniors) by One-Way ANOVA.

| GN | Mean | SD | GP1 | GP2 | GP3 | F | P |
|---|---|---|---|---|---|---|---|
| GP1 (50) | 3.17 | .56 | | | * | 28.39 | .000 |
| GP2 (50) | 3.32 | .44 | | | * | | |
| GP3 (50) | 3.65 | .53 | * | * | | | |
| GP4 (50) | 3.94 | .25 | * | * | * | | |

Table 10.12: Differences in composition scores of four groups tested by ANOVA

The probability (.000) of F-value (28.39) presented in Table 10.12 means that the overall differences in composition scores of the students across four university years are significant statistically. Furthermore, the results show that Freshmen and Sophomores did not show any significant differences in their scores but these two groups are significantly different from Juniors and Seniors respectively, and Juniors are also statistically different from Seniors.

How can we compare means by SPSS? First you should open the Statistics menu, Compare Means is the third item on the list. When your mouse highlights the third item, you can see a set of choices on your right side that include Independent-samples *t*-test, Paired-samples *t*-test and One-way ANOVA. Double clicking Independent-Samples T-Test on your right side, you will get the following dialog box shown in Figure 10.3. Under Test Variable(s), you have to select from the list on the left the variable, the means of which you intend to compare and under Grouping, you have to choose the variable that represents different groups. In the case of the example given in Table 10.10, Non-traditional Strategy should be moved from the left list to the right under Test and Sex to the right under Grouping.

Next you double click "Define Groups.." and get a dialogue box of Define groups as shown in Figure 10.5 in which you write 1 and 2 to represent females and males.

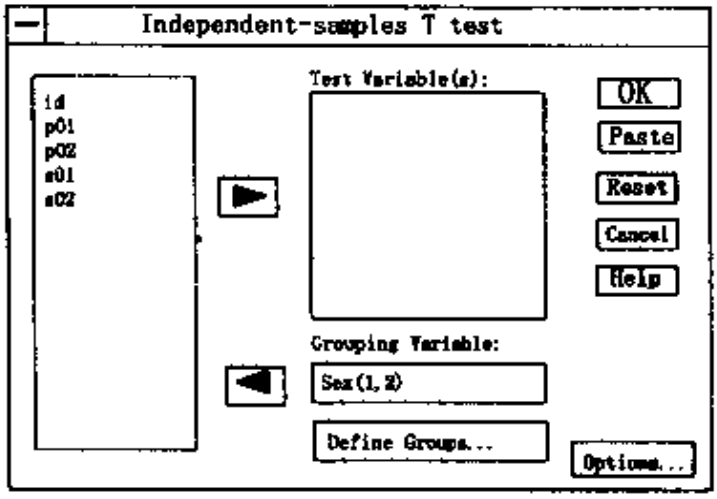



Figure 10.5: Independent-Samples T-Test dialog Box (Adapted from SPSS Inc.: 261)

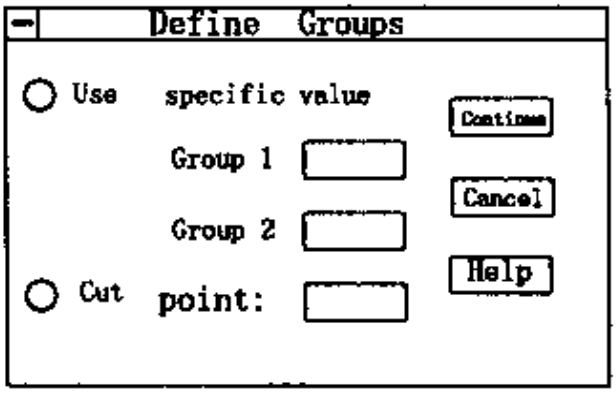
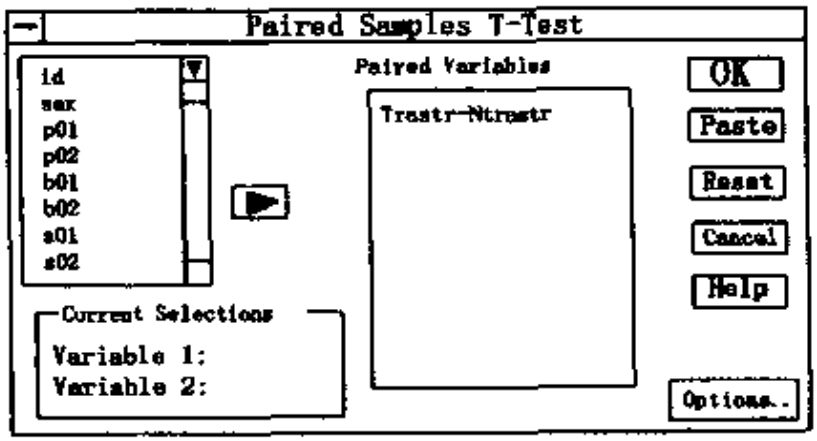



Figure 10.6: Define Groups dialogue box (adapted from SPSS, Inc.: 261)

To operate Paired-Sample T-test, you first highlight Compare Means on the Statistic Menu and you will see a list of choices on your right side. Double clicking Paired-Samples T-test on your right side, you will get the following dialog box presented in Figure 10.6. On the left side corner, under Current Selections there are two rows named Variable. You select the two variables to be compared by turns. In the case of the example in Table 10.9 the first variable is Traditional Strategy and the second one is

Non-traditional Strategy. Once the two variables have been selected, you move the pair to your right side under Paired Variables. By Clicking OK, the result will be obtained.

Figure 10.7: Paired-Samples T-Test dialogue box (Adapted from SPSS Inc.: 263)




To operate One-way ANOVA, you first highlight Compare Means on the Statistic Menu which will lead to a list of choices on your right side. One-way ANOVA is at the bottom of the list. By double clicking One-way ANOVA, the following dialog box will appear on your computer screen:

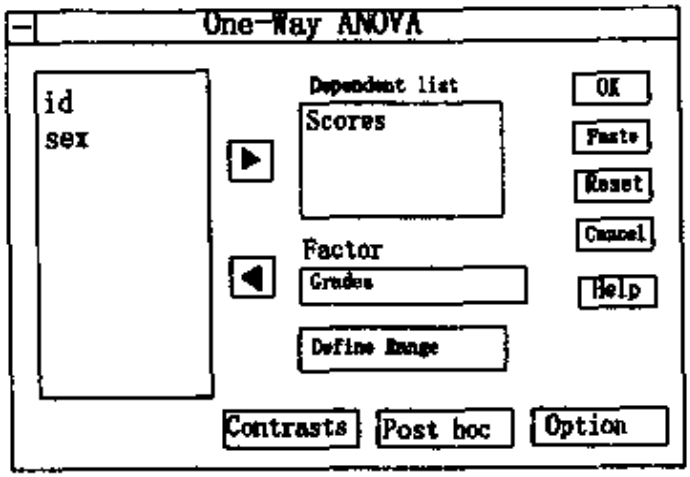



Figure 10.8: One-Way ANOVA dialog Box (Adapted from SPSS Inc.: 275)

To make an analysis by One-Way ANOVA, you need to go through the following three steps. First, you move the variable to be compared from the left to the right under the heading of Dependent List. Secondly, you move the variable representing different groups from the left to the right under the heading of Factor. In the case of the example shown in Table 10.12, the variable to be compared is Scores and the factor is Grades. Thirdly, you click "Define Range" to get a dialogue box entitled One-Way ANOVA Range as shown in Figure 10.9 where you fill in the numbers to indicate Group 1 and Group N. Finally, you click Post Hoc to get a dialogue box of One-Way ANOVA Post-Hoc Multiple Comparisons as shown in Figure 10.10 where at the bottom you tick Scheffe. Without the last step, you will not be able to get the detailed information of multiple comparisons.

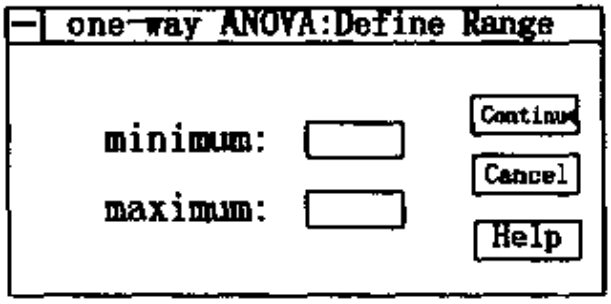



Figure 10.9: One-Way ANOVA Range dialogue box (Adapted from SPSS Inc.: 276)

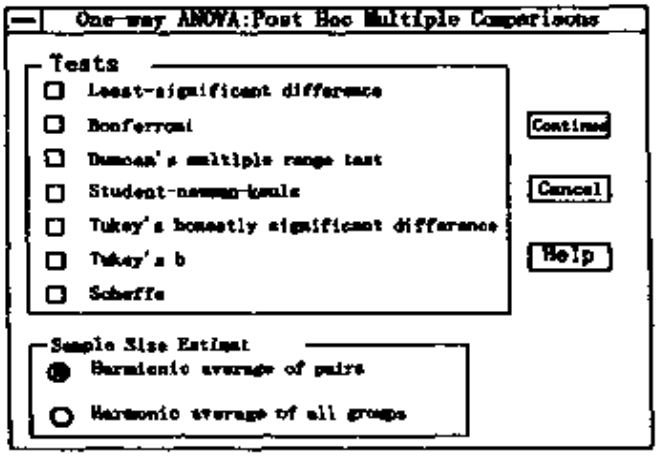



Figure 10.10: One-Way ANOVA Post-Hoc Multiple Comparisons (Adapted from SPSS Inc.: 278)

### *Multiple Regression analysis*

Multiple regression analysis tells us how much each of several independent variables contributes to predicting the score of a single dependent variable. Actually in our daily life, we often do such predicting without noticing the involvement of the multiple regression analysis. For example, we may estimate a person's weight according to his/her gender, age and height. Through a regression analysis, we may know the exact proportion of contribution each independent variable has made to predicting a person's weight.

In the field of applied linguistics, a question we frequently come across is to what extent we can predict students' L2 achievement through a series of independent variables such as L1 proficiency, Previous L2 proficiency, Motivation, Effort, Beliefs, Strategies, etc. In this case, you can use a regression analysis to answer the question. Table 10.13 presents the result yielded by the multiple regression analysis in my study of the relationsbip between modifiable learner factors and L2 achievement:

| Dependent Variable | Multiple R | $R^2$ | F | |
|---|---|---|---|---|
| Scores on Band 4 | .62 | .39 | 2.57** | |
| Independent Variable | | | Beta | Sig T |
| 1. L1 proficiency | | | .03 | .78 |
| 2. Previous L2 Prof. | | | .43 | .001 |
| 3. Deep motivation | | | .23 | .06 |
| 4. Surface motivation | | | .11 | .41 |
| 5. Effort | | | -.15 | .30 |
| 6. Management belief | | | -.14 | .32 |
| 7. Form-focused Belief | | | -.13 | .36 |
| 8. Meaning-focused Belief | | | .16 | .23 |
| 9. Mother-tongue Belief | | | -.01 | .95 |
| 10. Management Strategy | | | .02 | .88 |
| 11. Form-focused Strategy | | | -.07 | .59 |
| 12. Meaning-focused Strategy | | | .24 | .06 |
| 13. Mother-tongue Strategy | | | -.32 | .03 |

Table 10.13: The predicting power of learner factors on L2 learning

*Note*. ** < .001

The results in Table 10.13 indicate that 13 independent variables together can predict 39% variance in the scores on Band 4 at the significance level of .001. According to the absolute values of Beta, we may find that the most powerful predictor of the dependent variable is Previous L2 proficiency that has the highest Beta value, i.e. .43 at the significant level of .001. The second powerful predictor is Mother-tongue Strategy with its absolute Beta value being .32 at the significance level of .03. Deep motivation and Meaning-focused strategies are of equal predicting power since they have the same Beta value (.23) at the same significance level (.06). They are, however, less powerful than Mother-tongue Strategy and their power does not have statistical significance because their probability level is greater than .05. The remaining independent variables have very low absolute Beta values all of which do not reach the required level of significance. Consequently, they do not possess power at all. You might notice that some variables have positive Beta value and some, negative ones. The positive one means the predicting direction is positive while the negative one means the opposite direction. For example, the positive Beta value of Previous L2 proficiency means that the higher the level of previous L2 proficiency a student got, the higher the score on Band 4 s/he is likely to obtain; while the negative Beta value of Mother-tongue Strategy means that the more frequently this strategy is used by a student, the lower the score on Band 4 that s/he is likely to get.

To operate the regression analysis, you open the Statistic menu where by highlighting Regression, you select Linear. Having a click, you will see a Dialogue box on the screen (See Figure 10.11) which asks you to decide which is a dependent variable and which are independent variables. Once you have made a decision, you move them from the left to the right under the headings of Dependent and Independent(s) respectively. In the above example presented in Table 10.13, the dependent variable is Scores on Band 4 and independent variables are

altogether 13 such as L1 proficiency, Previous L2 proficiency, Deep motivation, Surface motivation, etc.

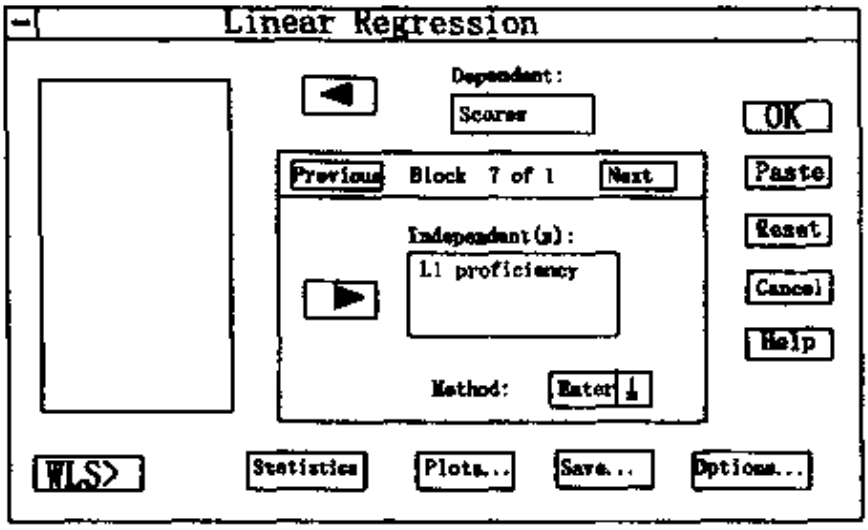



Figure 10.11: Linear Regression dialog box (Adapted from SPSS Inc.: 358)

## ANALYZING DATA FROM AN EXPERIMENT

To set up a data file for an experimental study is not totally different from a survey study. The procedures for setting a data file discussed in the previous part can be used to deal with the data from an experiment. Furthermore, the statistical procedures such as Summarize, Correlation analysis, Compare means and Multiple regression can be perfectly operated on the experimental data, too. The major difference is that the analysis of surveying data may end up with Summarize or Correlation analysis but the analysis of experimental data must involve Compare means. To avoid unnecessary repetition, I will only focus on some aspects that are particularly important for analyzing the data from an experiment.

## Problems in inputting the data

One common problem frequently encountered by novice researchers is in what format the data should be inputted. Suppose in your experimental study, you have 30 students as a control group and 30 students as an experimental group. How do you input the data? Do you place the data of the 30 students from the control group side by side with the data of the 30 students from the experimental group? Or do you put the data of these two groups sequentially (that is one group after another)? Very often a novice researcher is not clear which decision he /she should make. A wrong decision at the beginning would cost you enormous time to figure out why some statistical procedures cannot be implemented.

The simple rule is that if your experiment involves two or more different groups of people, you have to key in the data of one group after another in a sequence. Simply we may call such a format a sequential format. In addition to that, you have to specify whether a person belongs to a control group or to an experimental group. The data file looks like the one displayed in Table 10.14. Here the variable ID refers to the identification number you assigued to each student and the variable Group contains information concerning whether a person is in a control group or in an experimental group. Specifically, "1" represents an experimental group and "2" represents a control group. Such a sequential format of data is required by Independent-samples *t*-test and One-way ANOVA which, however, simply do not work on the data presented in an parallel format as shown in Table 10.15.

| ID | Group | Pretest | Posttest |
|---|---|---|---|
| 01 | 1 | 70 | 72 |
| 02 | 1 | 64 | 71 |
| 03 | 1 | 86 | 80 |
| 04 | 1 | 72 | 81 |
| ... | .... | .... | .... |
| 31 | 2 | 71 | 68 |
| 32 | 2 | 63 | 74 |
| 33 | 2 | 80 | 80 |
| 34 | 2 | 60 | 63 |
| ... | .... | .... | .... |
| 61 | 3 | 68 | 69 |
| 62 | 3 | 80 | 83 |
| 63 | 3 | 76 | 72 |

Table 10.14: A **correct** data-format for Independent-samples *t*-test and One-way ANOVA (A sequential data format)

| ID | Group | pretest | posttest | ID | Group | pretest | Postest |
|---|---|---|---|---|---|---|---|
| 01 | 1 | 70 | 72 | 01 | 2 | 71 | 68 |
| 02 | 1 | 64 | 71 | 02 | 2 | 63 | 74 |
| 03 | 1 | 86 | 80 | 03 | 2 | 80 | 80 |
| 04 | 1 | 72 | 81 | 04 | 2 | 60 | 63 |
| ... | .... | .... | .... | | .... | .... | .... |

Table 10.15: A **wrong** data format for Independent- samples *t*- test and One-way ANOVA

By contrast, Paired-sample *t*-tests can only function on the data presented in a parallel format in which two variables are placed side by side. As was mentioned before, Paired-samples *t*-test compares the means of two variables which are, however, related to one group of subjects. Let's look at Example Two again introduced at the beginning of Chapter 7. The researcher intended to find out whether negotiated interaction is better than modified instructions in improving L2 comprehension. The way to

measure students' comprehension is to see how many objects the subjects can correctly place on the board by following the teacher's instructions. In this case Paired-sample t-test is needed. The appropriate data format for such a *t*-test is like the one in Table 10.16

| ID | Modified instruction | Negotiated interaction |
|---|---|---|
| 01 | 13 | 16 |
| 02 | 9 | 15 |
| 03 | 7 | 10 |
| 04 | 11 | 12 |
| ... | .... | .... |

Table 10.16: A **correct** data-format for Paired-samples *t*-test (a parallel data format)

## "Compare means" in experimental studies

No experimental studies can avoid the use of "Compare means". This includes Independent-samples *t*-test, Paired-sample *t*-test and One-way ANOVA. How can you make sure the statistics you have chosen are appropriate? The following rules may help you make a better decision.

1. Compare the means of one variable exhibited in two groups ⇒ Independent-samples *t*-test that requires a sequential data format
2. Compare the means of one variable exhibited in three or more groups ⇒ One-way ANOVA that requires a sequential data format
3. Compare the means of two variables exhibited in one group only ⇒ Paired-samples *t*-test that requires a parallel data format

## SUMMARY

The Statistical Package for Social Sciences (SPSS) is a commonly used statistical package that consists of three main windows: Application, Ontput and Data Editor. The application window contains all the statistical procedures of analyzing data available in SPSS; the output

window displays results yielded by SPSS and the data editor window helps create a new file or read the existing file.

The analysis of quantitative data from a survey study starts with setting up a data file by coding the data, numbering all the questionnaires, inputting the data and cleaning the data. If the questionnaire is intended to be analyzed in terms of categories, the conceptual categories need to be checked for their internal consistency by running item-total correlations. What kind of statistics is selected to operate on the data depends on research questions. By running "Summarize", frequencies and descriptive statistics are obtained. Correlation analysis is helpful when the relation between two variables is the concern. The statistical operation of "Compare means" is needed when two or more means are to be compared. Specifically speaking, comparing the means of two variables related to one group requires Paired-samples *t*-test; comparing the means of two groups related to one variable needs Independent-samples *t*-test; comparing the means of more than two groups concerning the same variable needs ANOVA. Multiple regression is used to determine to what extent a set of independent variables can predict variations in a dependent variable. The procedures for analyzing the data from an experiment is more or less the same as those used in analyzing the quantitative data from a survey. The only difference is that no experimental study can avoid the use of "Compare means". The use of Independent-samples *t*-test and ANOVA requires a sequential data format, i.e. the data of the control group and of the experimental group(s) are presented in a sequential format, while Paired-samples *t*-test needs a parallel format, i.e. the two variables to be compared are placed column by column.

## DISCUSSION QUESTIONS

1. Design a mini-research project in which 30 subjects are asked to respond to a questionnaire in order to answer one research question.
   1) Set up a data file.

2) Check whether the conceptual categories you formed can meet statistical requirements.

2. What is a correlation analysis?
3. What is the difference between an independent sample *t*-test and a paired-sample *t*-test?
4. When do you use *t*-tests and when do you use one-way ANOVA?
5. Accomplish the following tasks by using the sample data provided by SPSS.
   1) Make a correlation analysis to find out the correlation coefficient between the initial salary and the current salary, and the correlation coefficient between the current salary and the education level.
   2) Compare males' and females' current salary to see whether there exists a significant difference between these two groups of employees.
   3) Compare the females' initial salary and the current salary to see whether the current salary is significantly different from the initial salary.
   4) Run one-way ANOVA to find out whether there exists a significant difference in the current salaries of the three groups of employees who differ in their education levels.
   5) Make a multiple regression analysis to find out the power of the variables concerned in predicting the employee's current salary.

# 11. The analysis of qualitative data

This chapter will describe how qualitative data are analyzed. It will first present choices you may have for analyzing qualitative data. Secondly, it will show you how to prepare the data for the qualitative analysis. Finally, it will discuss how the qualitative analysis can be undertaken successfully.

## CHOICES IN ANALYZING QUALITATIVE DATA

As was mentioned in Chapter 5, qualitative data can be analyzed qualitatively or quantitatively. By qualitative analysis, I mean that the results are presented in words while by quantitative analysis, the findings are presented in numbers. Figure 11.1 presents three common choices from which you can select for the analysis of qualitative data.

### Choice 1: qualitative analysis

By the first choice, the researcher tries to identify categories emerging from the data which may form further patterns or variations. For example, when the interview data concerning the use of vocabulary strategies are analyzed, the researcher reads the data again and again and eventually s/he has arrived at three general categories: (1) strategies used in remembering new words; (2) strategies used in understanding the meaning and (3) strategies used to overcome insufficient vocabulary. For each general category, s/he has also established some subcategories. These general categories together with their subcategories are the answer to the question: What strategies are used by English majors in vocabulary learning?

## Choice 2: quantitative analysis

By the second choice, the researcher obtains the results simply by counting something in qualitative data. For example, in the study conducted by Wen and Guo (1998) in which the subjects were asked to write a composition based on the given pictures while speaking out what was going on in their minds, we counted the number of Chinese words occurring in their verbal protocols and then calculated the percentages of L1 occurrence with reference to the total number of words produced by think-aloud. The result is the answer to the question as to how much L1 is used in the process of L2 composing.

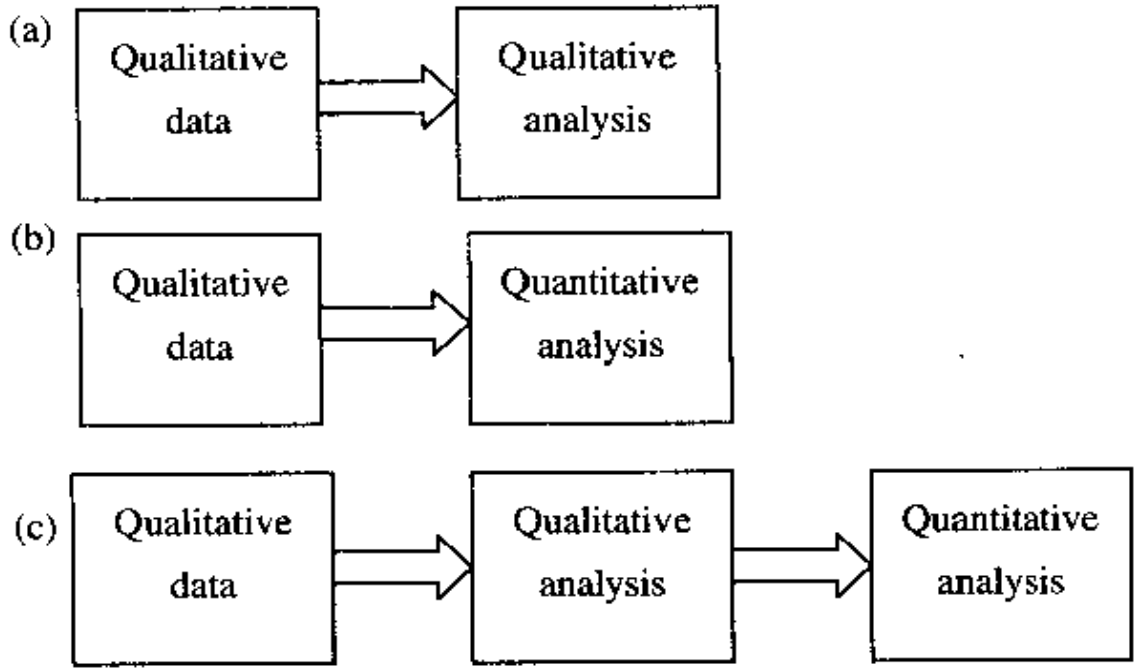



Figure 11.1: Choices in analyzing qualitative data

## Choice 3: Qualitative followed by quantitative

The third choice is more complicated than the first two and it is easy to mistake it for the second choice presented in Figure 11.1b, since by simply looking at the answer to the research question, the findings are quantitative. Nevertheless, the quantitative findings here are not directly

derived from the raw qualitative data. Let us look at Chen's study presented at the end of Chapter Five again.

The first research question in her study is "How frequently do Chinese children adopt strategies in remembering the spelling of new English words?" The findings in relation to this question are the following:

| Strategy categories | Frequencies | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9+ |
| Rote-learning | 28 | 5 | 16 | 16 | 18 | 8 | 5 | 1 | 2 | 1 |
| Form-association | 54 | 3 | 14 | 15 | 5 | 1 | 5 | | 1 | 2 |
| Familiar word | 5 | 1 | 8 | 8 | 11 | 13 | 23 | 6 | 9 | 16 |
| Phonetic method | 36 | 3 | 16 | 16 | 15 | 6 | 4 | 1 | 1 | 2 |
| Syllable | 5 | 1 | 8 | 13 | 15 | 12 | 17 | 9 | 11 | 9 |
| Word-building | 60 | 3 | 11 | 8 | 7 | 2 | 9 | | | |
| Pronunciation rule | 26 | 2 | 20 | 12 | 18 | 12 | 5 | | 2 | 3 |
| Visual frame | 55 | 7 | 22 | 8 | 5 | | 2 | 1 | | |
| Pinyin | 73 | 5 | 15 | 5 | 2 | | | | | |
| Arbitrary cutting | 44 | 7 | 21 | 8 | 4 | 10 | 3 | 3 | | |

Table 11.1: The frequencies of sharegies used in remembering the spelling of words identified from think-aloud data (Chen, 2001: 32)

It is true that the findings concerning frequencies are quantitative but the point is that the frequencies are not available in the raw data. In other words, the researcher was unable to obtain such frequencies directly from the think-aloud protocols. What she first did was to establish 10 categories of strategies by repeatedly reading the subjects' protocols.

This process is, however, less obvious to the reader. In fact, the process by which such a qualitative analysis is conducted must be detailed in the section of methodology.

The above three choices can be used independently for a study depending on the research questions. A clear understanding of your own choice(s) is essential for you to avoid confusions in analyzing the qualitative data and in describing how the data-analysis was made in your study.

## DATA PREPARATION

Like a quantitative analysis where the researcher needs to set up a data file by coding, inputting and cleaning the data, a qualitative analysis also needs a data file which can facilitate an effective and efficient analysis. It is obvious that setting up a qualitative data file is much more demanding and challenging than a quantitative one. Generally speaking, three tasks need to be accomplished: (1) Transcribing tape-recordings; (2) Segmenting the data in relation to a specific research question; (3) Displaying the segmented data in a table. For diaries, the first task is unnecessary while in the case of think-aloud protocols, it is not possible for you to perform the second and the third tasks since a protocol is a recording of a natural flow of thinking which is not responding to any specific question. The following section will discuss how each task is undertaken.

### Transcribing tape-recordings

In the case of interview and think-aloud, the participants' responses are normally recorded on tape. In order to have a data file, the researcher must transcribe the tape-recordings. I have to say, the task of transcribing is tedious, boring and time-consuming and requires painful patience on the part of the researcher. Yet it is indispensable and crucial for the data-analysis. Poorly transcribed protocols may reduce the validity and reliability of the study. Even worse, they can completely ruin a study.

Therefore, the researcher has to deal with this professionally and conscientiously.

The first and most important principle for transcribing is faithfulness. As a researcher, you bave to take down everything you can hear. If there is anything unclear, you had better ask the subject to clarify it by listening to the tape again rather than depending on your own speculation and imagination. The second principle is to standardize the way of transcribing. By standardization, I mean that you have to develop a set of codes that can indicate, for example, pause, stress, hesitation, etc. In the case of pause, you need to decide further whether the length of pause should be indicated, and how it is specified if such an indication is necessary. By employing this set of codes all the way through, the transcriptions are standardized and easy to process afterwards.

## Segmenting Data

In the case of interviews and diaries, you may segment the data with reference to each specific research question so that the irrelevant data can be put aside. Be careful that such an arrangement of data cannot be made in haste. The researcher has to read the raw data repeatedly and make sure the selected parts can well represent the interviewee's answer to the question concerned. If the decision is made in haste, you will most likely overlook some valuable information. Furthermore, whatever words or sentences appear in the segmented data file should have originated from the raw data.

Let's take Kwok's study (1999) for example. In her study, the subjects were the students who took a one-year full-time teacher's certificate course in the Faculty of Education in Hong Kong University. The project was set up to see what were the developmental patterns and variations in student-teachers through one year's full time course. The subjects were asked to write reflective notes at six points of the course: on the entry to the course, before the school experience, after the school

experience, before the Main school practice, after the school main practice, and at the end of the course.

The segmented data file is presented here concerning the research question "What is the difference in the perceptions of a good teacher between the experienced teachers and the novice teachers on the entry to the course?". Six student-teachers were selected for an in-depth analysis with three being experienced and three without any experience. Apart from the background information about the subjects[15], the phrases, sentences and quotes were all from the students' reflective notes on entry to the course, responding to the question: "What is your perception of a good teacher?"

Group 1: experienced teachers

Subject 1 (Philippino) Sybil Soler Maninang

Native language: Tagalon; Age: 26; Sex: F;

Marital status: unmarried; Teaching experiences: 4 years

* Learner-centered.

"Personally, I think this is a key factor in determining a teacher's effectivity. An effective teacher for me is one who can modify his lessons to suit the characteristics and needs of the students"

* Character molder.

"I believe that education must address itself primarily to character/personality formation. It must not get caught up in equipping the youth with technical knowledge to the neglect of human formation, which consequently, produce skilled robots and socially decrepit persons. With this, I do not undermine the importance of academic formation. I would just like to remind the teachers that the youth should be prepared primarily for day to day living, i.e. capable of handling the ups and downs of life."

---

[15] For the ethical reason, the names of the six students are not real ones.

Subject 2: Helen Wong

Native language: Chinese; Age: 30; Sex: F;

Marital status: Married; Teaching experiences:5 years

*Able to help students develop intellectually but also spiritually

*With good qualities: having a humble spirit, honesty and consistency without being hypocritical

Subject 3: Julie To

Native language: Chinese; Age: 42; Sex: F;

Marital status: Married; Teaching experiences: ? years

*Keen to learn (update his knowledge on the subject)

*Positive attitude towards his students

*Honest to his colleagues, students and himself

*Enthusiastic (enjoy teaching)

*Dedicated (have a strong sense of commitment to teaching)

Group 2: (no teaching experience)

Subject 4: Josephine Chui

Native language: Chinese; Age: 22; Sex: F;

Marital status: unmarried; Teaching experiences: 0

*Good personality: not a commander but friendly and helpful

*Have a good command of knowledge of the subject: a teacher is not a dictionary but understands the subject well and can teach them systematically; be well-prepared before she comes to the classroom

*Fair grading system

*Teach efficiently

Subject 5 Jocye Y. F. Tsao

Native language: Chinese; Age: 25; Sex: F;

Marital status: unmarried; Teaching experiences: 0

*Patience; *Love students

*Equipped with good explanation ability, fluent spoken English and good communication skills with adolescents to establish good relations with students

Subject 6: Ronica K. S. Wong

Native language: Chinese; Age: 22; Sex: F;

Marital status: Unmarried; Teaching experiences: 0

* Enthusiasm for teaching;

* Prepare before lessons

* Care for students without prejudice

* Continually learn and try to make improvement of himself

By having the above segmented data file concerning a specific research question, it is much easier and more convenient for you to make a further analysis.

## Displaying data

Data display is defined as a segmented data file with further reduction and synthesizing, which can be presented in a table or in a diagram, or a structured summary. By reduction and synthesizing, the details such as quotes and explanations are removed, and some loosely-structured ideas are summarized. As a result, a data display is shorter and more concise than the segmented one so that the researcher's attention resources can be used exclusively for thinking about the meanings of the most important data. However, such reduction and synthesizing also risk distorting the original data. To avoid this danger, you may take one of the two options. The first is to ask another researcher to check your work. If this is not possible, you may redo it a week later to see whether the result you get the

second time shows any difference. Remember for whatever reasons you are not allowed to distort the data.

| (Group 1: Experienced. Teachers)<br>**Sybil** | **Helen** | **Julie** |
|---|---|---|
| *Student-centered learning<br>*Character molder | *Help students develop academically but also spiritually<br>*Humble, honest, consistent | *Update the knowledge of the subject<br>*Positive attitude towards Students<br>*Honest, enthusiastic, dedicated |
| (Group 2: Novice teachers)<br>**Josephine** | **Joyce** | **Ronica** |
| *Not a commander but friendly and helpful<br>*Knowledgeable<br>*Teach systematically and efficiently<br>*Fair grading system | *Patience, love students<br>*Fluent spoken English<br>*Good communication skills with adolescents<br>*Good explanation ability | *Enthusiastic for teaching<br>*Care for students without prejudice<br>*Prepare before lessons<br>*Continually learn and improve oneself |

Table 11.2: Data display of perceptions of a good teacher

Now let's look at the examples presented in Table 11.2. In Sybil's response, by deleting two long quotes, her answer was shortened to two phrases: student-centered learning and character molder which are the subjects' own wording. In the case of Josephine, compared with the answer in the segmented file, except for the fourth point that remained as it was, the other three points all underwent great changes. The first point was shortened but kept in her own wording; the second point is summarized in a new term "knowledgeable" that was the researcher's wording; the third point is a result of combination of two ideas that were not together in the original data..

# QUALITATIVE ANALYSIS

As was mentioned at the beginning of this chapter, the qualitative data can be analyzed either quantitatively or qualitatively or in combination. Since the quantitative analysis has already been discussed in detail in Chapter Nine, I will not repeat it here. This section will be for qualitative analysis only.

## A framework for qualitative analysis

To show how qualitative analysis is carried out is difficult in a course unless you have patience and time to go through the laborious process of analyzing the data together with me. Now to illustrate it by writing is even more difficult since the process is very complicated and messy to some extent. To understand the way the data are analyzed needs a lot of illustrative examples which, however, cannot be condensed into one or two paragraphs. What can I do in this section? I will first try to introduce to you the framework for qualitative analysis proposed by Strauss and Corbin (1998) which is graphically described by the author as shown in Figure 11.2. Their framework is known as grounded theory which assumes that a theory is embedded in qualitative data and it can be developed through a systematic analysis. Secondly, I will use an example to illustrate how this framework can be operated. However, my introduction will be brief and not sufficient for you to analyze the qualitative data needed for your Master or Doctoral thesis. I strongly recommend that in addition to reading this section, you should read the book *Qualitative Data Analysis* written by M.B. Miles and A.M. Huberman in 1994 and the book *Basics of Qualitative Research* by A. Strauss and J. Corbin in 1998.

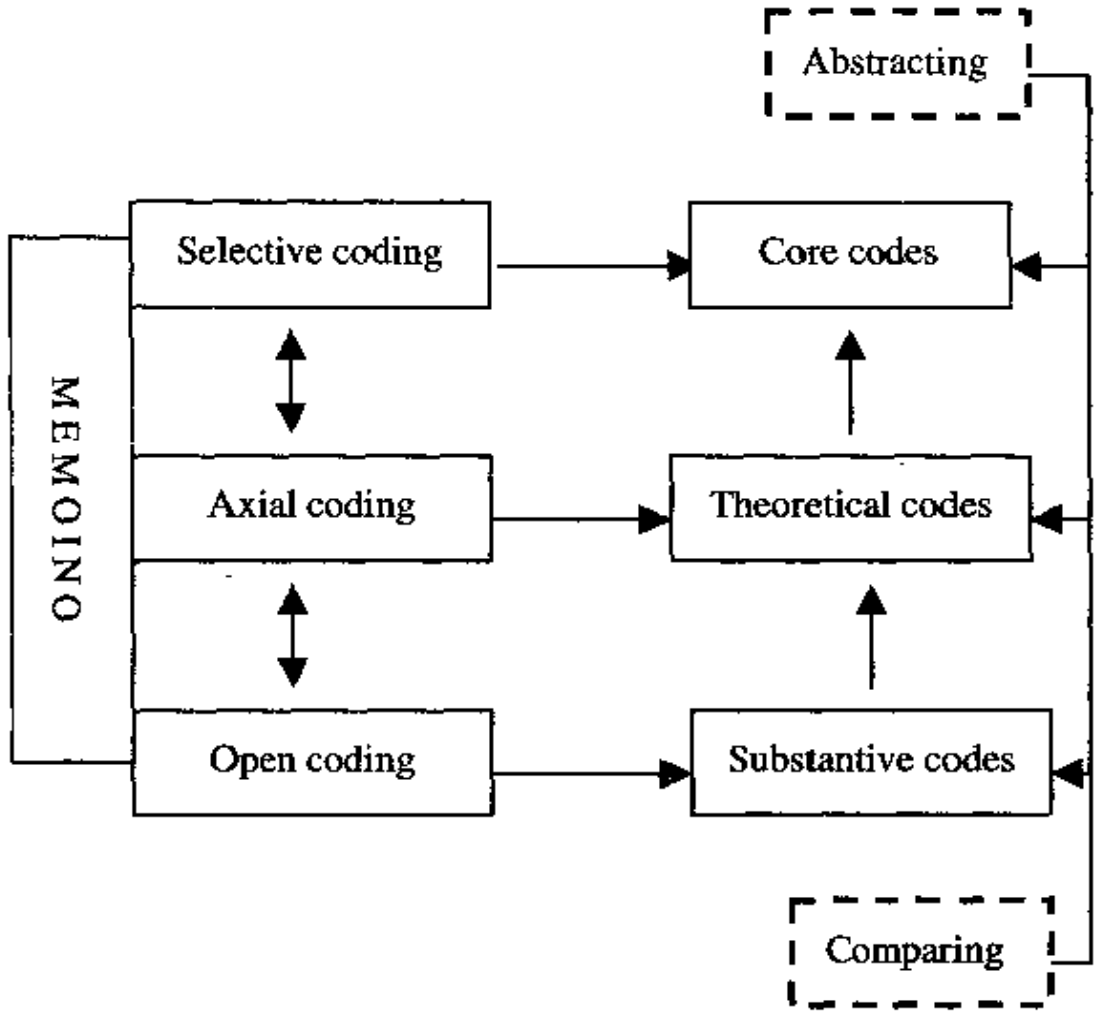



Figure 11.2: The framework for qualitative data analysis

According to Figure 11.2, you have to undertake two basic tasks, i.e. coding and memoing. The coding can be further divided into three kinds (i.e., open coding, axial coding and selective coding) that yield three types of codes accordingly. The resulting codes form a hierarchy in terms of abstraction, i.e., substantive codes from the opening coding are at the lowest level, core codes from the selective coding at the highest level and theoretical codes from the axial coding, in between. Memoing is to take notes while engaging in coding. Notes are multifaceted. They may be the researcher's analysis of, interpretations of, and comments on the data.

After all, they may record anything occurring in the researcher's mind while coding the data.

The primary mental operations involved in accomplishing the tasks are abstracting and comparing. Abstracting is to seek a more general term or a statement to describe and explain a phenomenon or less general concept. Comparing is to establish similarities and differences between objects or concepts.

## Primary mental operations

In this part, primary mental operations, i.e. abstracting and comparing, will be explained and two basic tasks, i.e. coding and memoing, will be described.

Abstracting and comparing are primary mental operations in the qualitative data analysis. That is to say, no matter what kind of coding you are doing, what your mind is actively engaged in is either abstracting or comparing or both in combination. The following sections will focus on these two mental activities one by one.

### *Abstracting*

Abstract is in contrast with "concrete". However, the degree of abstraction can be varied from less abstract to very abstract. The results of coding can be placed along the hierarchy of abstraction (See Figure 11.3). The lowest level of abstraction is concepts that stand for a set of discrete units which share a common property or display a similar attribute. The next level of abstraction is categories that subsume a group of concepts that show a common property. Hypotheses are more abstract than categories, and indicate the researcher's understanding of the relations between categories and their subcategories. The highest level of abstraction is theory in which all the hypotheses are logically related to each other within the same paradigm.

So far as a single project is concerned, it may not necessarily aim at constructing a theory. Which level of abstraction is the goal is decided by

your research questions. No matter what level of abstracting you want to reach, you are always expected to identify something that can be shared by the units you are working at. The shared element can be a property, or an attribute or a paradigm. For the lower level of abstracting, you look for a common property or a similar attribute possessed by the items for abstracting. For example, the word "liquid" can designate petroleum, corn oil or water. Although they differ in color, mass and function, each has the specific property of being able to flow. For the higher level of abstracting, you are interested in a paradigm by which the categories can be related to each other. For instance, liquid, air and solid are related to each other along the paradigm of being non-animate materials and thus they are called matter. Very often, the common property and paradigm are not evident particularly for novice researchers.

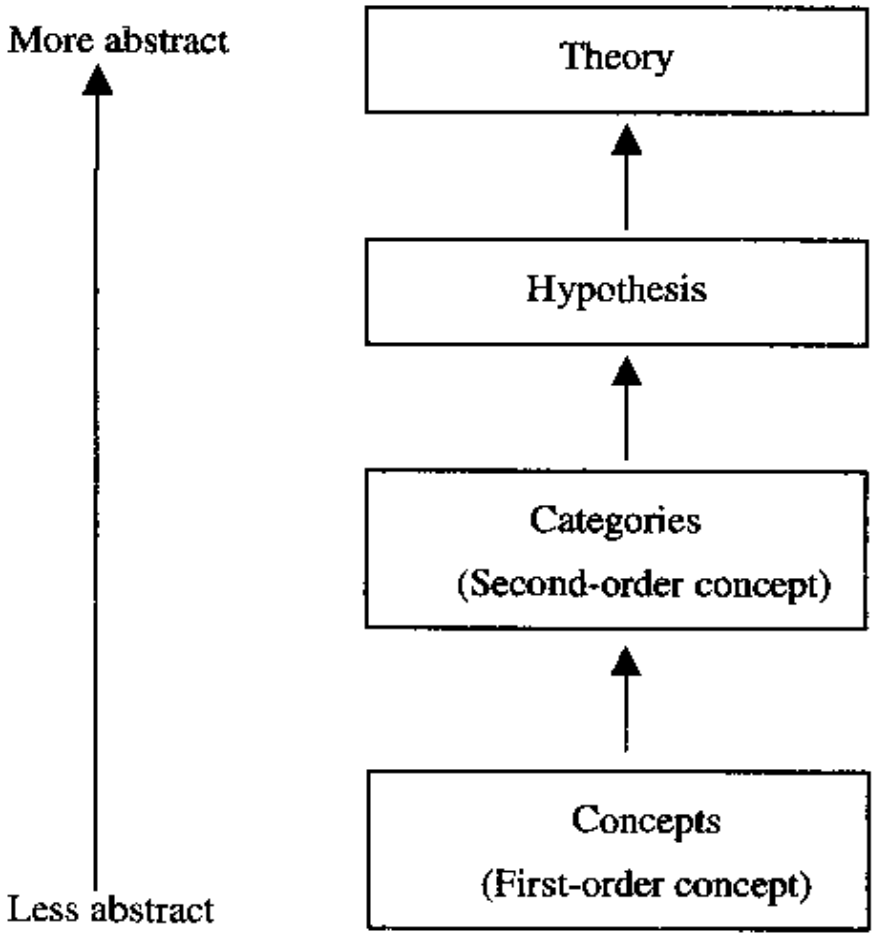



Figure 11.3: A hierarchy of abstraction

Although the results from coding are clear-cut in terms of their degree of abstraction, our minds do not work in the same manner. That is to say, the procedures by which our minds work are not parallel to the distinct levels of abstraction, moving from the lowest level of abstraction to the highest level. More often than not, our minds may work with concepts, categories, hypotheses and theory simultaneously and each may interact with the other. Sometimes we may work from theory to concepts and sometimes from concepts to theory. After all, our thinking routes are varied. However, for the sake of convenience of discussion, we break down our thinking processes analytically in this part, trying to say that there are three levels of coding which correspond to the three levels of abstraction: open coding leading to concepts and categories, axial coding to hypotheses and selective coding to theory. I hope this will not give you the wrong impression that qualitative data analysis consists of a set of procedures. Actually, the most challenging aspect of qualitative analysis is that there are no fixed procedures to follow.

### *Comparing*

Comparing is another essential mental activity in data analysis. It is intermingled with abstracting. Comparing may take place at different levels of abstraction. At the lower level of abstraction, you compare incident to incident, idea to idea, action to action in an attempt to give a name to similar incidents, or ideas or actions. Once concepts are identified, you compare concept to concept to look for similarities and differences to establish categories, and then compare category to category to see their relations. It is by comparison that a common property or attribute or paradigm can be identified. Comparing does not only take place within the study but also across studies.

## Two basic tasks

### *Coding*

Conceptually speaking, open coding, axial coding and selective coding are different in nature. Open coding aims at identifying substantive codes, axial coding at theoretical codes and selective coding at core codes. Although they are distinct operations, they are not necessarily undertaken sequentially. In many cases, they overlap and are operated synchronously.

**Open coding**

In open coding, the researcher should keep open all the theoretical possibilities in the data. That is to say, researchers should not think along a single track and confine themselves to one option only. It is better to do opening coding in a group of researchers rather than individually. If group analysis is not possible, you may put aside the provisional codes and several days later reexamine the same data with a fresh mind. By repeating this process, you turn the data upside down and inside out to try out on them all the theoretical possibilities.

In open coding, you first discover concepts in data and then classify the identified concepts into different categories. You examine every incident or act or idea very closely in the whole set of data. By comparing incident to incident, you look for similarities and differences. Based on similarities, concepts are identified and further classified into categories. The resulting concepts or categories are called substantive codes in the grounded theory.

It is important to have categories because they can help reduce the total number of units for analysis. Furthermore, categories become easier to remember and contemplate. Above all, they can be broken down into subcategories for further analysis. The name for a category is usually chosen depending on the researcher's perspective since the same phenomenon can be named differently. Above all, categories should be

logical and transparent. By saying logical, I mean that a name can be justified; by saying transparent, I mean that a name is clearly related to its referent.

Let's take for example a qualitative study on L1 involvement in L2 picture composition by thinking-aloud (Wen & Guo, 1998). One of the research questions is to identify the functions of L1 use. The results show that the primary functions of L1 use include: Transformation, Confirmation, Generating ideas, Retrieving L2 forms and Controlling the writing procedures. The name of these five functions are categories, some of which have subcategories. For instance, The category of Generating ideas contains the following subcategories: Reasoning, Judging, Associating, Questioning, Monitoring and Evaluating.

The whole process is not in a linear fashion: concept ⇒ category ⇒ subcategories. Actually, any concept you identify at the beginning is only provisional in nature and needs modifications. You should make sure the concept or category you generated can account for all the data. If anything conflicting occurs, you should revise the concept or category. The modification cannot be made at one go. By nature, category formation has to be progressive. However, to avoid too many subsequent modifications, you had better spend more time on the data and examine them closely before establishing any concept or category.

### *Axial coding*

While open coding breaks the data apart and keep them open to all theoretical possibilities, axial coding reassembles the categories and forms a whole picture again by looking for an axis that can connect categories and their subcategories. Once the connection has been set up, it can offer more precise and powerful explanations about phenomena. Although the link between categories is refined in Axial coding, an understanding of how categories are related to each other begins to develop in open coding. In this sense, open coding and axial coding are not clear cut. Let's look at the category of Generating ideas again. It is

not the case that we identified Generating ideas first and then searched its subcategories or the other way around. Actually, when we read the data, the category and the subcategories emerged almost simultaneously.

To find out an axis is the key to axial coding. But what is an axis? Simply speaking, it is a perspective or a paradigm used in the grounded theory from which you examine the data. It may refer to conditions under which a certain phenomenon occurs. These conditions concerning when, where, how and why form a contextual structure in which a phenomenon occurs. It may refer to a series of actions that evolve over time under specified conditions. The actions denote a process in which a phenomenon occurs. It may refer to consequences that are results of actions. Let's look at the category of Generating ideas again. The category has six subcategories: Reasoning, Judging, Associating, Questioning, Monitoring and Evaluating. Obviously, these six actions differ in nature psychologically. However, they are all actions in the process of generating ideas for writing an L2 composition. Therefore, the axis in this case is "function". In other words, these six activities all serve the same function, i.e. generating ideas.

Strauss (1987) points out that axial coding contains three major activities: (1) specifying the properties of a category and their dimensions; (2) identifying the possible conditions, actions and consequences associated with a phenomena; (3) looking for cues in the data that might be insightful for the relation among the categories.

One thing that Strauss and Corbin emphasize in their book is that to identify an axis or a paradigm is a means but not an end. The end is to seek explanations and obtain an understanding of phenomena rather than looking for conditions, actions and consequences.

### *Selective coding*

Selective coding, as its name suggests, selects one aspect as a core category or a central category which represents the main theme of the research. When the core category is chosen, the scope of analysis is

narrowed down and the focus comes to the surface. Around the core category, we can pull together all the relevant categories to develop a theory.

It is clear that a core category plays an essential role in constructing a theory. Then what is qualified as a core category and how is it identified? Strauss (1987: 36) listed a series of criteria for a qualified core category:

1. It must be central; that is, all other major categories can be related to it.
2. It must appear frequently in the data. This means that within all or almost all cases, there are indicators pointing to that concept.
3. The explanation that evolves by relating the categories is logical and consistent. There is no forcing of data.
4. The name or phrase used to describe the central category should be sufficiently abstract that it can be used to do research in other substantive areas, leading to the development of a more general theory.
5. As the concept is refined analytically through integration with other concepts, the theory grows in depth and explanatory power.
6. The concept is able to explain variations as well as the main point made by the data; that is, when conditions vary, the explanation still holds, although the way in which a phenomenon is expressed might look somewhat different. One also should be able to explain contradictory or alternative cases in terms of that central idea.

Knowing the criteria is not sufficient since knowing and doing are quite different in nature. It is not uncommon that beginning researchers who are familiar with the criteria still fail to detect the central category. They are confused by the flooding data and unable to look at the data from different perspectives. Strauss and Corbin (1998) give beginning researchers several valuable suggestions that can facilitate the selection of the central category and the integration of concepts. One suggestion is to write the storyline, that is, to write a few sentences to describe what the

research is about. If you find it difficult to start, you may go back to the raw data and reread several cases while keeping asking yourself the stimulating questions, such as "What is the main issue here?" and "What is most striking to me?" The second suggestion is to use diagrams. Diagrams are helpful tools to visualize the relations among the concepts. Once you start drawing a diagram, your thinking about the logic of relations must be clarified. Otherwise, the resulting diagrams are muddled and confused. The third suggestion is to review and sort through memos (how to write memos will be discussed in the next section). Memos have recorded all the ideas in your mind while you are coding. As the analysis proceeds, your memos contain more and more abstract ideas which are helpful for you to extract a central category.

### *Memoing*

Memoing is an essential task for qualitative data analysis. It records what you have obtained through abstracting and comparing, and how you went about them in the coding process. It forces you to be alert to the data and enables you to be clear about which direction you are going in. By having memos, you have products of analysis which are easily revised, supplemented and elaborated or negated in the later process. After all, memoing is a very flexible device, as Glaser's definition illustrates:

> A memo is the theorizing write-up of ideas about codes and their relationship as they strike the analyst while coding...it can be a sentence, a paragraph or a few pages... it exhausts the analyst's momentary ideation based on data with perhaps a little conceptual elaboration. (Glaser, 1978: 83-4)

Memoing should go hand in hand with coding. Whatever has resulted from coding should go into memos. They may be substantive notes, theoretical notes, methodological notes or personal notes. When they are substantive or theoretical notes, they should be primarily conceptual.

Furthermore, they should go beyond what is being analyzed. In other words, they move from the actual events to the abstract concepts, and from a lower level of abstraction to a higher level of abstraction.

One important thing you should remember is that memos evolve over time and memos are not differentiated from right to wrong or from good to bad. Actually, the content in memos is provisional and bound to change when the analysis proceeds. The notes in earlier memos might appear to be naive and even ridiculous. However, they become more and more sophisticated when you dig deeper into analysis. The category or pattern or relationship gradually emerges from the data. So long as you keep recording the ideas in the analysis and keep checking them against new data, you are moving towards more and more advanced levels of abstraction. Eventually, your memos serve as a basis on which a theory will be set up.

## An illustration

Let's closely examine the data presented in Table 11.2 and make a qualitative analysis. The question is to see whether there is any difference in the perceptions of a good teacher between experienced teachers and novice teachers when they were enrolled into the course.

Where should we start? Actually there is no standardized procedures to follow. What I am going to talk about is a simple description of the procedures by which I analyzed the data. As you know, our thinking usually does not happen in a linear fashion. It is very common that several ideas occur simultaneously in our minds. However, once they are to be described verbally, they have to be presented one after another as if they occurred in sequence. In this sense, the following description is not a true record of the process of my data analysis.

### *Detecting similarities in Group 1 and Group 2 respectively*

By reading the data presented in the first row, i.e. Group Gne, two terms caught my eyes, i.e. "Character molder" and "spiritually". I wrote in the memo: The experienced teachers seem to show great concern for the students' mental development. However, a similar term did not appear in the third subject Julie. By comparing Helen with Julie, I found that both of them mentioned some qualities required by a good teacher. This idea was also recorded in the memo. Now I was stuck and could not see anything shared by the three subjects. Instead of racking my brains thinking about the first group, I started to work at the second group. Two common features immediately emerged from the data after the first reading. The first common feature is that they were all concerned with their relations with the students; the second one is that they all emphasized effective teaching. Thus, I produced a table in which these two common features are presented together with evidence.

| Name | Relation with students | Effective teaching |
|---|---|---|
| Josephine | Not a commander but friendly and helpful; Fair grading system | Knowledgeable; Teaching Systematically and efficiently |
| Joyce | Love students; Good communication skills with adolescents | Fluent spoken English; Good explanation ability |
| Ronica | Care for students without Prejudice | Enthusiastic for teaching; Prepare before lessons; Continually learn and improve oneself |

Table 11.3: Two common features shared by three novice teachers

Once I identified these two concepts or substantive codes in Group Two, I went to check whether these two features were also available in Group One. The interesting thing is that the novice teacher's concerns did not gain any attention from the experienced teachers at all. It seems that these two groups approached the same question from totally different perspectives. What were their different perspectives? I tried to capture it by reading the data again and again. Suddenly, two terms struck me and gave me an understanding of the data, that is "fair-grading system" and "without prejudice". These two terms would be typically used by students when they talk about their teachers. Novice teachers without any teaching experiences naturally approached this question from the student's perspective. Thinking along this line, I realized the second feature "effective teaching" was either from the student's perspective or from the perspective of a subject teacher.

### *Identifying the differences between the two groups*

With these two categories in mind, I closely examined the data of Group One again. It seems to me that to Sybil and Helen, a good teacher was an educator who should help the student develop as a whole person, i.e. both intellectually and mentally. Julie's view seemed to be different. Her perspective is of a teacher in general and she was concerned with the general qualities of a good teacher. At this stage, I constructed a diagram in the memos. In the diagram, I tried to list all the different perspectives which are placed hierarchically. Furthermore, I tried to link the substantive codes with theoretical codes .

Once I finished drawing the diagram, the answer to the question "What is the difference in the perceptions of a good teacher between the experienced teachers and the novice teachers" was very clear as if they were self-evident.

The above illustration of qualitative analysis, although it is oversimplified, can well show you how comparing and abstracting

operate, how substantive and theoretical codes are obtained and in which way memos can be used to help you analyze the data qualitatively.

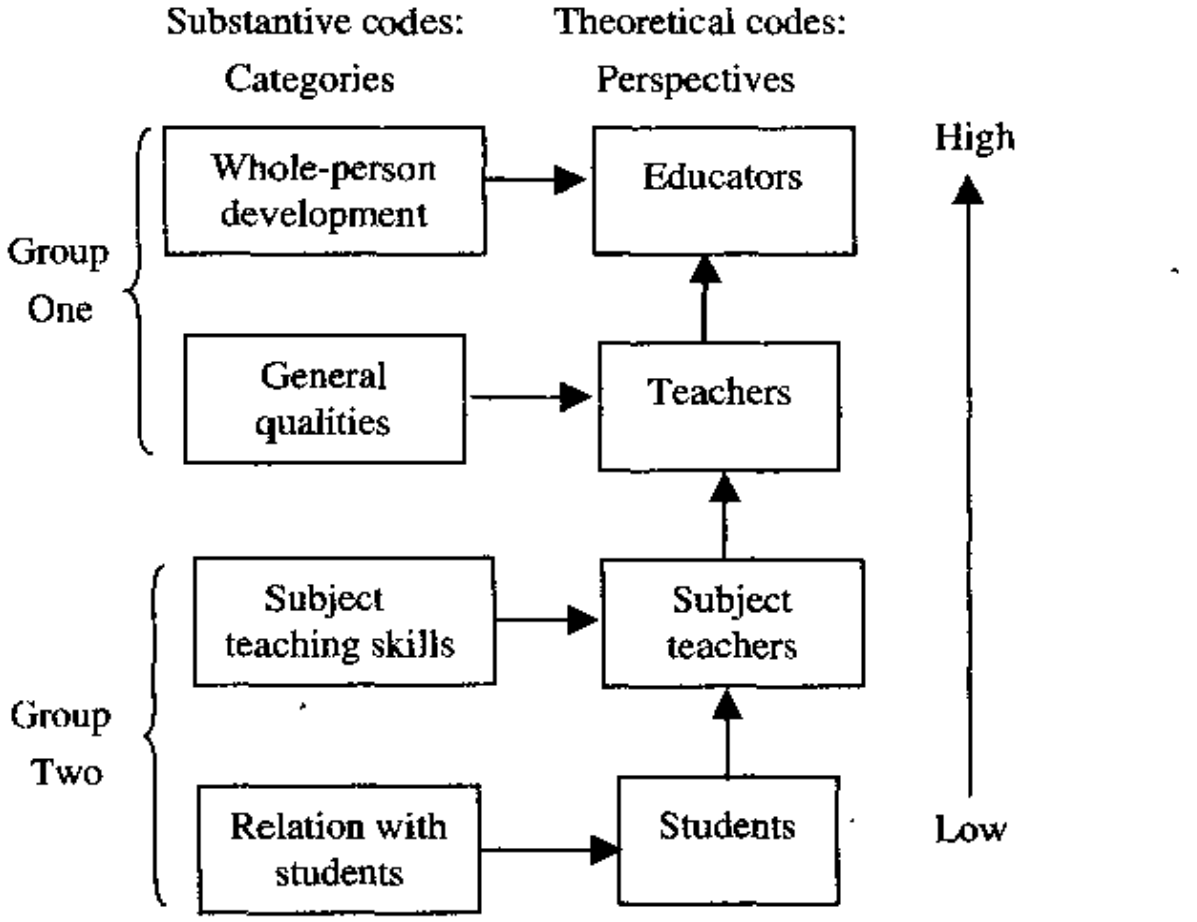



Figure 11.4: Differences in perceptions of a good teacher between Group One and Group Two

One thing that has to be emphasized here is that the answer to the same question might be varied if the researcher takes a different perspective. This is one of the reasons why qualitative analysis is fascinating. For example, the above question might have the following answer: The fundamental difference between the experienced teachers and the novice teachers in the perceptions of a good teacher is that the former are concerned with professional development while the latter, with the survival skills of a teacher.

## VALIDITY AND RELIABILITY

Qualitative analysis is often criticized as lacking reliability and validity. This could be a problem if no measures are taken to guard against flaws in the data analysis. As was said before, qualitative analysis has no fixed procedures available that can naturally lead to reliable results. Furthermore, substantive codes, theoretical codes and selective codes are not on the surface of the data and they can be identified only when the researcher has gained insights through repetitive reading of and continuous thinking about the data.

By saying the analysis is valid, I mean that the identified codes can explain the data. In other words, the codes should be fully supported by the data rather than something that the researcher has imposed on the data. By saying the analysis is reliable, I mean that two or more researchers can obtain similar codes if they are asked to work at the same set of data.

To make sure the qualitative analysis is valid and reliable, two researchers had better analyze the same set of data independently and then discuss together. If there are some controversies they cannot resolve, a third party needs to be invited to join the discussion.

## SUMMARY

Qualitative data can be analyzed either qualitatively or quantitatively or both. To prepare for analyzing interviewing and thinking-aloud data qualitatively, you have to transcribe the tapes faithfully with a set of standardized codes to indicate the pause, stress, etc. To clarify unclear parts of the tape it is better to ask for help from the participant rather than depending on the researcher's own speculation. Qualitative data are better to segment with reference to specific research questions and then display them through reduction and synthesizing. Segmenting combined with displaying can facilitate the qualitative analysis. Following the framework proposed by Strauss and Corbin, we are expected to carry out

two basic tasks: coding and memoing. By two primary mental operations: comparing and abstracting, the researcher starts with opening coding followed by axial coding and ends up with selective coding. These three kinds of coding lead to three kinds of codes: substantive codes, theoretical codes and core codes that can be placed on a bierarchy from the lower level of abstraction to the higber level of abstraction. Above all, you have to remember that these three kinds of coding are not ordered procedures and the process is intuitive and interactive. In memoing, you are supposed to write down whatever occurs in your mind in the process of comparing and abstracting. The ideas may be fragmentary or randoin at the beginning but can nevertheless show the development of your analysis and can also serve as a basis on which you can obtain insights. In order to improve validity and reliability of qualitative analysis, it is best to ask another researcher to analyze the same set of data. If this is not possible, you have to analyze the same set of data at least twice with soine time in between.

## DISCUSSION QUESTIONS

1. Go to the library of the School of Foreign Studies and find examples from students' theses to illustrate how qualitative data are analyzed quantitatively or qualitatively or in combination.
2. What do you need to consider, when you transcribe tape-recordings?
3. In the segmentation of data, what are you expected to do?
4. When you display the data, how do you deal with a segmented data file?
5. How do you differentiate three kinds of coding?
6. What is the function of memoing?
7. How can you inake sure the qualitative analysis of the data is valid and reliable?

# Part III

# Thesis Writing

Thesis/ dissertation writing is very critical for a graduate student. No one would deny its importance. However, More often than not, graduate students haven't received sufficient guidance in this regard. Some survive through desperate struggles; some fail after exerting painstaking efforts. In this part, I would like to share with you the lessons I have learned about thesis writing through trial and error. It consists of three chapters. Chapter 12 presents an overview of the structure of a thesis/dissertation. Chapter 13 describes how each part of the thesis/dissertation is written up. Chapter 14 discusses the problems in format and style. These three chapters will enable you to:

- Understand the structure of a thesis/dissertation.
- Know how to write up each part of the thesis.
- Deal with some typical technical problems in thesis writing.

# 12. An Overview of Thesis Writing

This chapter will start with definitions of a thesis/dissertation and then introduce to you the general structure of a thesis/dissertation. Finally, a brief description of each part of the thesis will be presented.

## DEFINITIONS

### *A thesis or a dissertation?*

There is no universal consensus on the definitions of a thesis and a dissertation. Cone and Forster (1993), by comparing the definitions of these two terms presented in Webster's New International Dictionary and Random House Dictionary of the English Language, conclude that the distinction between the two terms is not clear at all. However, they describe common practice in most American Universities, i.e. thesis is used to refer to the written work for a master degree and dissertation, for a doctoral degree. According to my own experience in the British education system, the term "thesis" can be used for both master and doctoral written work. In this book, I would like to take these two terms as synonyms and regard them as interchangeable.

### *What is a thesis/dissertation?*

A thesis/dissertation is an argument in which different parts are logically related and all center around research questions (Punch, 1998). It is neither a simple recording of what has been done in your research nor a description of what has been found in your study. Since it is an argument, in addition to describing, you need to explain why the study is conducted in the way you designed.

A research process is invariahly messy in one way or another. To make a messy process appear to be neat in a paper, you are expected to examine the whole process and report only the information relevant to answering your research questions. In other words, you have to ignore the irrelevant things you have done. To put it differently, a thesis is a

report of your study in which only important and relevant things are selected to present. However, selection does not mean hiding unexpected problems or results.

## THE STRUCTURE OF A THESIS

The following section will start with a general account of the structure of a thesis which is followed by a brief description of the three major components of a thesis, i.e. the beginning, the body and the ending.

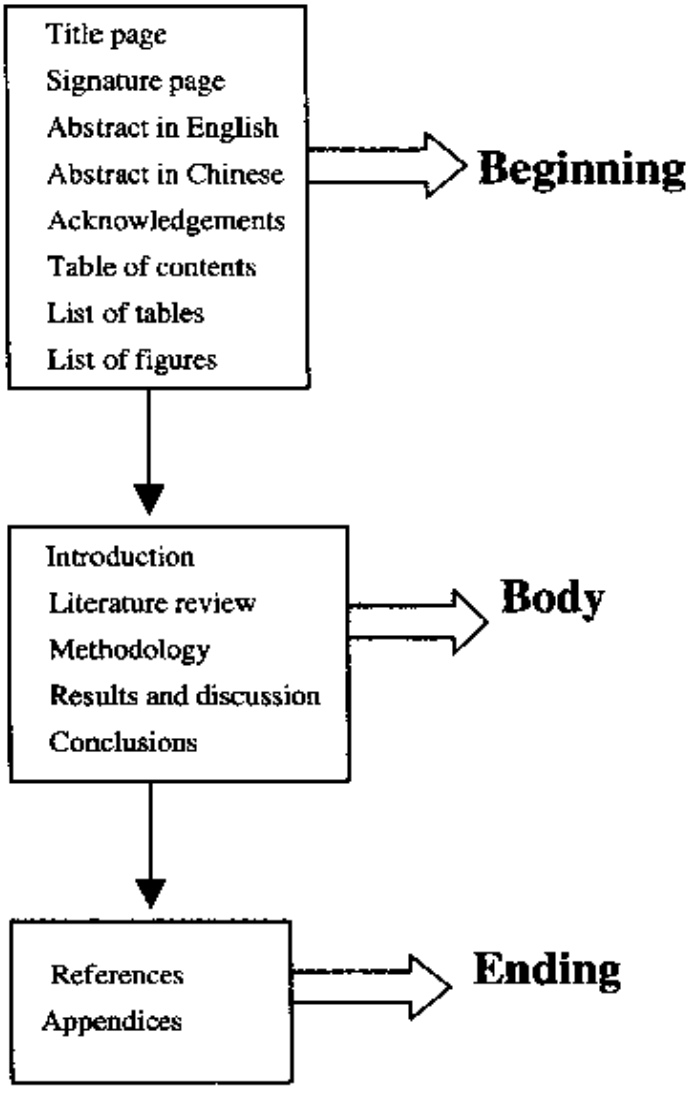



Figure 12.1: The structure of a thesis

## An Overall Picture

A general description of the structure of a thesis is given in Figure 12.1 which is the same as the actual sequence of the manuscript. We may divide the whole sequence into three parts: beginning, body and ending. The beginning part covers from "Title page" to "List of figures"; the body part contains five topics: Introduction, Literature review, Methodology, Results and discussion and Conclusions; the ending part contains "References" and "Appendices".

One thing that has to be emphasized is that the sequence presented here is not the actual sequence of writing your thesis/dissertation. A common practice is that the beginning part is written last, the ending part, second and the body, first. Nevertheless, for the convenience of discussing, the sections are organized according to the actual sequence presented in a thesis/dissertation.

## Beginning Part

The beginning part of a thesis includes the title page, the signature page, the abstract in English, the abstract in Chinese, acknowledgetuents, table of contents, list of tables and list of figures. The following section will briefly describe what should be covered in each part and what problems might arise in its writing.

### *Title page*

The first page of the thesis is a title page which should provide the following information: the title of the manuscript, whether it is a doctoral or a master's thesis, at which university the thesis was written, who wrote the thesis and when the thesis was finished. The format is varied from university to university. You had better go to the secretary of the graduate program to get the format first before you construct this page.

The title should be clear and succinct, and must describe the study. If there are only a few variables under investigation, the title may contain these variables. If many variable are under investigation, the title may

specify the type of independent and dependent variables being examined. If possible, the type of subject is better described in the title. The following are the examples of thesis titles[16]:

1. A contrastive study of English and Chinese compliment responses (Gong, 1998)
2. Advanced level English language learning in China: the relationship of modifiable learner variables to learning outcomes (Wen, 1993)
3. A process-oriented cognitive account of L1 influence on L2 writing in the Chinese context (Guo, 1997)
4. An investigation into the internal structure of EFL motivation at the tertiary level in China (Qin, 1998)
5. A study of second-year English Majors' pragmatic competence (Wu, 1998)
6. A study of the changes of tertiary English majors' beliefs about L2 learning (Su, 1996)
7. A study of university students' argumentative writing in English: rhetorical knowledge and discourse pattern (Wu, 1998)
8. Error treatment in EFL classrooms in universities in China (Hu, 1999)
9. Influence of different tasks, lengths of planning time and lengths of learning on L2 oral parformance of university students (Zhu, 1997)
10. L2 proficiency and comprehension strategy use of EFL learners in universities in China (Lu, 1997)
11. Risk-taking and English learning: a study of the risk-taking beliefs and behaviors of English majors in China (Wang, 1999)
12. The relationship of L2 learners' linguistic variables tu L2 writing ability of non-English majors in China (Ma, 1998)

---

[16] The titles are taken from either MA or Ph.D. theses but a few of them have been slightly modified in order to make them more explicit.

For some universities, the total number of words in a title has a limitation and the title of a thesis must be finalized three months before submitting the thesis. That is to say, once the title has been submitted to the higher degree committee, the writer cannot change the title.

### *Signature page*

The signature page is the second page of a thesis. Although this page is simple to construct, the information is very important. Without the supervisor's signature, thesis that has not legally been submitted.

### *Abstract*

The abstract is a summary of your study. It should be coherent by itself. It provides information concerning the following aspects: the purpose of the study, the research questions to be addressed, the subjects involved, the instruments used to collect the data, the procedures for collectiug and analyzing the data, the findings and the conclusions. If this is the first tiine for you to write an abstract, I strongly recommend that you read a few abstracts in a journal and analyze their structure before you write your own.

In the Chinese context, you are also expected to write an abstract in Chinese. No doubt, the content of the abstract in Chinese should be the same as that in English. However, a literal translation of the English version is definitely unsatisfactory. Some English majors seem to be weak in writing in Chinese. Very often the draft of an abstract in Chinese contains some sentences that do not sound like Chinese and some are even difficult to understand. The best way to ensure the readability of the abstract is to ask the students from other departments to read it and make comments on it before you submit it to your supervisor. One thing that I have to emphasize here is that the abstract in Chinese is normally written for the members on the high degree committee who are not good at reading in English. This abstract is therefore the only thing through which these members can evaluate the quality of your thesis/dissertation. In this sense, the quality of this abstract may even determine your success

or failure in obtaining a graduate degree. Therefore, I strongly suggest that you should spend as much time as you can on this task.

### *Acknowledgements page*

The acknowledgements page is the place to express the researcher's gratitude to those who have offered help in the process of research and thesis writing. The thanks are usually expressed to: (a) the supervisor; (b) the teachers and your classmates who once gave you suggestions or advice on your research; (c) persons who helped you collect data and do proof-reading; (d) family and friends who gave you either physical and/or mental support; (e) sources for financial support if there are any: (f) those who kindly permitted you to use their research instruments or other materials.

One common problem in this part is that the statements of thanks are very general and abstract. The information about the way in which you have been helped is not provided. It appears that the writer expresses his/her thanks simply for formality rather than from the bottom of the heart. Being over-modest could lead the reader to conclude that the writer might be an incompetent researcher. Being less-specific might leave a wrong impression on the reader that the writer is not sincere enough. The best strategy to express your thanks is to specify exactly why each person should be thanked. The specifications should say just what the person to be thanked has done in relation to your research and thesis writing. No more, no less.

### *Table of contents*

The table of contents lists the headings of the beginning, body and ending parts of a thesis. The beginning part is marked by Roman numbers while the remaining parts are sequenced by Arabic numbers. For the body of a thesis, the subheadings of each chapter are listed. They are indented to show their different levels. It is not common to put

all levels of subbeadings in the table of contents. Three levels are usually demonstrated. Table 12.1 is an example of part of a table of contents.

# TABLE OF CONTENTS



Table 12.1: An example of part of a table of contents (Wen, 1993: vii)

### *List of tables*

List of tables is a place where the tables spread out in a thesis are now put together while each table has a page number. Referring to the page number, the reader can quickly make access to the information needed.

### *List of figures*

List of figures provides information in the same way as a list of tables. It contains all the figures spreading out in a thesis with a page number being specified. The common problem I have observed is that some graduate students cannot make a clear distinction between a table and a figure. A figure may refer to a graph, a chart, a diagram, or a picture while a table must be either a square, a rectangle in which numbers and words are presented in rows and columns.

It is a good place to review these lists for their appropriateness of the titles and structures. Very often when the tables and figures are scattering on different pages, such problems are obscure. Only when the tables and figures are brought together, will the problems surface.

## Body Part

This section will first introduce to you two common approaches to the organizations of the body of a thesis: a one-level model and a two-level model and then briefly describe the content of five parts within the body of a thesis: (1) Introduction; (2) Literature review; (3) Methodology; (4) Results and discussion and (5) Conclusion.

### *One-level and two-level models*

Usually, the body of a thesis is divided into five chapters: (1) Introduction; (2) Literature review; (3) Methodology; (4) Results and discussion and (5) Conclusion. This is a one-level model. The alternative structure is a two-level model, where parts occur at the first level and chapters at the second level (See Table 12.2). Such a structure is particularly useful for a doctoral thesis. However, this book will only

focus on the one-level model. In the following sections, each chapter will be briefly described.

| |
|---|
| PART I BACKGROUND |
| Chapter One Introduction |
| Chapter Two Literature review |
| PART II METHODOLOGY |
| Chapter Four The quantitative design |
| Chapter Five The qualitative design |
| PART III Results and discussion |
| Chapter Six ... |
| Chapter Seven ... |
| PART IV Concluding |

Table 12.2: An example of using PART and CHAPTER together

### *Introduction*

The Introduction is an overview of the whole study. It is usually short without technical details and can be understood even by most casual readers. Another challenge of writing this chapter is to present sufficient information for the reader to get an overall picture of the study.

### *Literature review*

An introduction is followed by a literature review in which the relevant existing literature is reviewed to indicate what has already been done and what problems remain to be solved. Equipped with the knowledge of this chapter, the reader can visualize the research context where the link between the proposed study and the previous ones is displayed and its potential contribution to the existing literature is demonstrated.

### *Methodology*

The Methodology chapter describes the design of the proposed study. It includes general as well as specific research questions or hypotheses,

information about subjects, instruments, the procedures for data-collection and data-analysis. The information presented in this chapter should be so explicit and transparent that any other researcher can easily replicate your study if she/he wants to.

### *Results and discussion*

Chapter Four is entitled "Results and Discussion". The results are the answers to your research questions and tbe discussion explains the possible reasons for a specific finding, and the significance of the findings, the link between the present findings and the previous ones, etc. The discussion may go along with each finding presented or may be presented separately from the Results.

### *Conclusion*

The Conclusion chapter includes major findings, implications of the findings and recommendations for future research.

The format of the body of a thesis may vary from one university to another. You should check with your supervisor or the secretary of the graduate program in your department to get detailed information about the number of chapters and the required content within each chapter. You had better get the information before you start writing in order to avoid reformatting the whole thesis, which can be time-consuming and painful.

## Ending Part

The ending part contains two sections: References and Appendices. These are important materials for a thesis but it is not appropriate to integrate them into the body of a thesis because they disrupt the smooth flow of the text. The following section will describe what References and Appendices contain.

### *References*

You might have read hundreds of books and papers. Not all references are, however, required to be listed in the Reference section.

The references here only refer to the materials you cite in your thesis. Following the APA (American Psychology Association) reference format, in the section of Reference, all the materials cited in the manuscript should be listed in an alphabetical order by author. Each reference should contain the following information:

1. The author(s)
2. The date
3. The title
4. For a book, the publisher and the place of publication
5. For a journal article, the journal name, volume, issue number and pages

One of the common problems I have observed in graduate students' theses is that they do not differentiate a bibliography from a reference list. A reference list only contains cited works that you referred to in your thesis/dissertation. A bibliography may include any work that you have consulted during your research which may not be cited in your thesis/dissertation. In the APA style, your thesis/dissertation requires a reference list rather than a bibliography. Another frequent problem is that the reference list may include some works that were not cited in the text or some cited works are absent from the reference list, mainly due to the lack of timely revisions of the reference list along with repeated revisions of the text. How can you make sure the reference list can avoid these problems? With the help of a working bibliography, producing a reference list becomes simple as well as accurate. You go through the whole thesis or dissertation page by page and tick the references in the working bibliography that appear in your writing. Those that haven't been marked are deleted. The remaining ones all go to the section of references. Meanwhile you also add cited works to the reference list if they are not included in your working bibliography. One thing I have to remind you is that the final revision of the reference list should be done

when you are 100 per cent sure that no further revision will be made in the text. In the process of revision, you had better save the section of references as a new file while the previous one is still kept in your old file. The advantage of such a practice is that it is easy to retrieve references you have deleted before. Finally, you have to double check to make sure that references cited in the text are present in the reference list and each entry in the reference list is cited in the text. One common practice of the members on the oral defense committee is to point out that details of some works cited in your thesis are not available in the section of References.

If at the last stage, you cannot get the accurate spelling of the author's name or the exact year of puhlication, instead of turning back to piles of materials and searching for the information for hours upon hours, it is often hest to rewrite the paragraph to drop out the reference.

### *Appendices*

The Appendices section contains any materials that you think important but cannot be put in the body of the thesis. The following materials are typical examples of what is presented as appendices:

1. The complete questionnaire
2. The interview schedule
3. The diary instructions
4. The test papers
5. The materials used in the training section
6. The report on a pilot study
7. The tasks the subjects are expected to undertake such as reading, writing, listening, speaking, etc.
8. The detailed results of data-analysis

## SUMMARY

A thesis is an argument in which different parts are logically related and all center around research questions. It consists of three major parts:

beginning, body and ending. In the beginning part, it includes Title page, Signature page, Abstract in English, Abstract in Chinese, Acknowledgements, Table of contents, List of tables and List of figures. In the body part, it contains Introduction, Literature review, Methodology, Results and discussion, and Conclusions. The ending part has References and Appendices.

## DISCUSSION QUESTIONS

Go to the library of the School of Foreign Studies and get a copy of a Master or Doctoral thesis. Read the beginning part and the ending part of the thesis to answer the following questions:

1. What is the overall structure of the thesis?
2. Are there any problems in writing the abstracts in English and Chinese?
3. Do you think the "Acknowledgements" is appropriately written? Can you make any further improvement on it?
4. How is the "Table of contents" presented? Do you have any suggestions for a better arrangement?
5. How about "List of tables" and "List of figures"? Is there any mistake in them?
6. Please make some critical comments on References and Appendices presented at the end of the thesis.

# 13. Writing up a thesis/dissertation

This chapter focuses on the writing of the body of a thesis. Specifically, it describes how to write an introduction, how to write a literature review, how to describe methodology, how to present the results and discussion, and how to conclude the thesis. For each part, I will introduce to you major components, the procedures of writing, and common problems or difficulties you may experience in your writing.

## WRITING AN INTRODUCTION

Specifically, the chapter of Introduction usually includes three cotuponents: (1) a general description of your study; (2) the need for a study and (3) the overall structure of the thesis.

The simplest way to begin the first component is "The study was undertaken to find out.... The research questions to be addressed in the study are tbe following:..." Since this is an overview of the research problem, you only need to list general research questions. Details of research questions will be presented at the end of the Literature Review Chapter or at the beginning of the Methodology Chapter.

For tbe need for your study, you are required to explain why the study is worth undertaking. Usually, the reasons can be explained in terms of practical as well as theoretical importance. A study is said to have theoretical value when it can confirm or disconfirm a theory, or modify existing theories, or clarify a controversial issue, or develop a new model that is badly needed, or enrich our understanding of a phenomenon. A doctoral dissertation is expected to have theoretical significance in addition to practical value. Therefore, a study that can merely provide a practical solution to a problem is not suitable for a doctoral dissertation (Rudestam & Newton, 1992). The practical values in the field of applied linguistics, may be shown in the improving of L2 teaching and learning, in the compiling of bilingual dictionaries, or in translation, etc. To

justify the need for the study, you may refer to one or more previous studies that motivated you to choose this research topic. However, many references are not necessary since this is not the place to do a literature review.

When describing the overall organization of the thesis, the chapter headings should be covered. If "Part" is the first-order heading and "Chapter", the second-order, the description should cover both Parts and Chapters. In some cases, Introduction Chapter also includes the definitions of key terms, delimitations, and assumptions of the study (Newman, Benz, Weis & Mcneil, 1997; Rudestam & Newton, 1992). In my opinion, the conceptual definitions of key variables are better included in the Literature Review chapter since they must be developed based on others' definitions; the operational definitions are better placed in Methodology Chapter where the instrument is introduced.

Although the Introduction is the first chapter in the manuscript, it is written when the other chapters bave been finalized. The reason is obvious. The research questions which were developed in your proposal may have been modified in one way or another almost until you finish writing up the chapter of Results and Discussion.

## WRITING A LITERATURE REVIEW

A literature review demonstrates the writer's conceptual understanding of the research topic and his/her ability to argue for the need for the study. This is the best place to see whether the researcher has acquired the macro-level organizing ability to put the relevant materials together to develop an argument. This is the most difficult part to write and takes much more pain to finish than is usually expected. However, if you survive such a painful process and eventually produce a good literature review, I am sure, you would gain a profound insight into the essence of research which you would otherwise not be able to obtain. Unfortunately, many students and even a few supervisors do not pay enough attention to the writing of this part. Students tend to think that

this chapter is a sort of formality, or to make a thesis long enough to have sufficient pages as required. The students often write it at the last minute by patching together all kinds of definitions and studies at his/her disposal without any clear and logical link. Frankly speaking, quite a few literature reviews in master's theses, although they have been passed, are of poor quality. To learn how to write a literature review is an essential part of a graduate program. It demands time and effort from both students and teachers.

## Major Components

No research is carried out in a vacuum. It must be linked with the past and the future. A literature review provides such contextual information. It not only describes what has been done in the field and what problems remain to be solved, but also explains to what extent your study is different from the previous ones, what kind of contribution can be made to the existing knowledge, how the design of your study is growing out of the existing ones and what kind of theory your design is based on if there is any. By the time readers finish reading the last section, they will be naturally led to conclude that your proposed study is worth exploring and its design is the best available on the topic.

Specifically speaking, a literature review should contain the following components:

1. Conceptual definitions of key terms;
2. An examination of the research topic in light of the theoretical perspective;
3. A description of related empirical studies with evaluating comments;
4. A critical review of research designs including instruments related to your study;
5. A conceptual framework if there is any.

The above components commonly appear in a literature review. However, they are not necessarily in the order they are presented here. In which way your literature review is written and whether you need include

more components in your review depends on your purpose and your personal preference.

### *Define key terms conceptually*

As was mentioned in Chapter Three, all the variables involved in your study have to be defined conceptually and operationally. The Literature Review chapter is a place to state the conceptual definitions of the key terms in your study. The simplest way to fulfil the task is to list your conceptual definitions together with references. However, this is not the best. To make your definition more convincing, it is better to review critically the various definitions that have surfaced in the literature while indicating how your definition has been derived or why you preferred this definition to others.

One common problem in writing this section is that a dozen definitions are listed on one or two pages without any comments in between and at the end of the list, the researcher's own conceptual definition is given. Whenever I come across such a case, I ask my students, "How are these definitions related to your own definition?" Another problem is that some terms defined cannot be regarded as key terms. For example, to study the relationship between the use of communication strategies and L2 speaking ability, students tend to start with defining the term "communication" and "communicative competence" and then move to the term "communication strategies". You might ask: what should be defined? The simple solution is that all the variables that were measured in your study have to be defined conceptually.

### *Review from the theoretical perspective*

Research topics are very often derived from theories or linked with theoretical issues. For example, the study on the acquisition of morphemes was inspired by Krashen's natural order hypothesis. My own study on the use of L2 learning strategies was motivated by a new perception of the role of learners. It is common that a review starts with

explicating a theory or a controversial issue that is relevant to the proposed study.

### *Describe and evaluate relevant empirical studies*

The description and evaluation of relevant empirical studies is indispensable in a literature review. However, this kind of description does not necessarily cover every detail of a study. Furthermore, the evaluation need not be comprehensive since it is made not for the purpose of judging the quality of the study concerned but for justifying the need for your study. The most crucial thing here is that you need to classify the studies according to your needs. For example, some studies may be included for the purpose of illustrating the extensiveness of the existing research on this topic; some for the purpose of showing conflicting findings; some for the purpose of displaying methodological problems that will be avoided in your study; some for the purpose of demonstrating the small number of subjects studied in the study, etc. Once you are clear about your purposes, you can decide the emphasis and scope of your description and evaluation.

### *Examine research designs*

This component is needed when one of the differences between your study and previous ones lies in research design. The difference may be shown in the number or type of subjects, the instruments, and/or the procedures of data-collection and data-analysis, etc. For example, the previous studies on the topic only employ a quantitative method which is confined to correlation analysis or multiple regression analysis. Suppose your study combines both quantitative and qualitative designs and your quantitative analysis goes beyond multiple regression analysis. In this case, you need to review the designs used in the studies concerned while pointing out their weaknesses.

If the difference is displayed in instruments, your review must include the Instrument section where a brief description of the existing instruments is given together with critical comments. At the end of the

section, you need to say that the new instrument that is going to be used in your study will avoid the weaknesses mentioned before.

### *Describe the conceptual framework in your study*

This component is optional for a master's thesis but compulsory for a doctoral thesis. The conceptual framework is actually a list of assumptions that have been established based on your literature review. This is also the basis on whicb your bypotheses for testing are developed. For example, in Ma's study on the relationships of L2 learners' linguistic variables to L2 writing ability, he presents the conceptual framework[17] as follows (Ma, 1998:45):

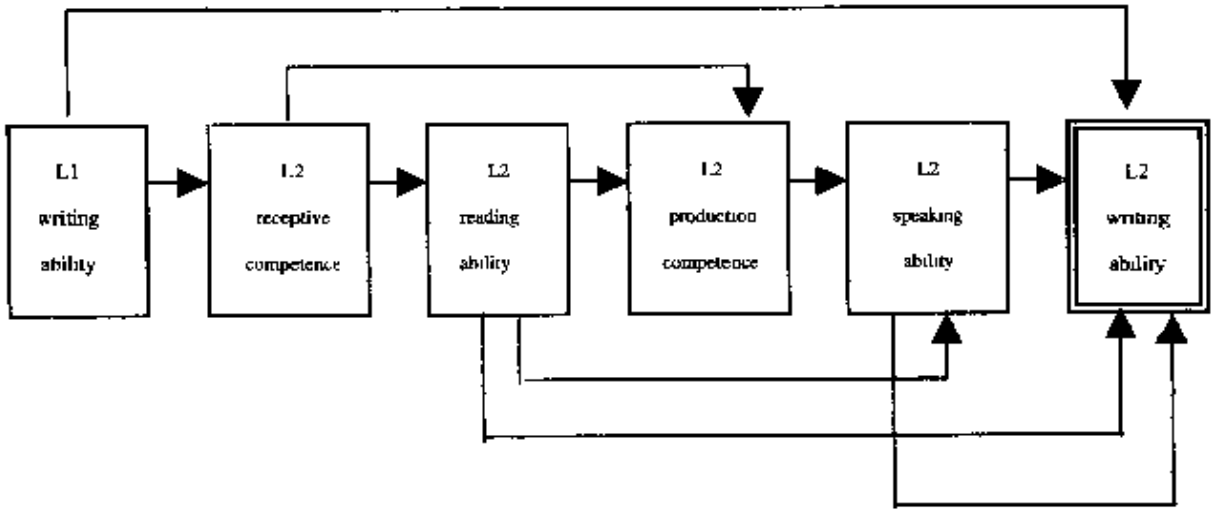



Figure 13.1: The specific conceptual model of L2 learners' linguistic variables affecting L2 writing ability

Ma explained in his thesis that "the model specifies the focus of the study and the causal relations between L1 writing ability and L2 writing ability,

---

17 Due to the limited space, a few changes are made in the framework, i.e. the names of categories are made by fewer letters but kept the same meaning.

between L1 writing ability and variables of L2 proficiency, and between variables of L2 proficiency and L2 writing ability."(p.45)

When your study is exploratory in nature, it may not be possible to present a conceptual framework in the chapter of Literature review. Nevertheless, it may develop out of the empirical findings yielded by your own study, which is most likely to be presented in the chapter of Results and discussion.

## Actual Writing

In this section, I will discuss with you how to write a literature review by describing the procedures, examining common difficulties and analyzing frequently-occurring problems at this stage.

### *Procedures*

A literature review usually contains an introduction and several subsections. An introduction is a sort of an advanced organizer in which the reader is informed of the scope of the review and its sequential arrangement. The subsections under specific headings should he clearly written and logically connected.

#### Construct an outline

Before you write a literature review, you should construct an outline in which subheadings are listed. The sequence of headings can show how the argument is developed. Two- or three-level headings, in my opinion, are sufficient for you to conceptualize the argument. The outline at this stage does not require the details since its major function is to help you to think of the literature review at the macro-level. In other words, it specifies the scope of your literature review.

It is very common that the outline needs to be revised several times. If you rush many things you write down will be proved to be irrelevant and have to be deleted later. One thing you have to remember is that even if the revised outline appears to be satisfactory and has been approved by your supervisor before, it will inevitably undergo changes along with the

progress of your research and writing. You should be prepared mentally for such changes.

**Draft the literature review**

Having produced a complete and well-organized outline, you can move to the next stage: writing the review section by section. You do not necessarily follow the order of sections strictly. There would be no problem at all if you choose to write first the section that you are most familiar with. Such a strategy can improve your writing efficiency since the section that is difficult today may become easy tomorrow.

**Revise the literature review**

The draft of a literature review inevitably contains a lot of holes that need further work. According to my own experiences, the draft of a review, no matter how good it apparently looks, needs revising at least three times. For the first time, you modify the macro-structure. The questions you have to ask yourself from time to time are: "Is this part relevant to my study?" and "How is this section related to the other sections of the literature review?". What you are concerned with is the organization of the whole review rather than small details such as grammatical mistakes or spelling errors or the improper structure of a sentence since these mistakes might not need correction at all if the whole section needs to be deleted later. The emphasis of the second revision is on the inter-paragraph level, i.e. the relations between paragraphs. The third revision is on the inter- and intra-sentential levels, i.e. the logical link between sentences and problems within a sentence. In other words, your attention should be paid to the flow from one sentence to another and linguistic errors in each sentence. Specifying a different focus for each revision can improve the efficiency of revision. Suppose we as supervisors insist the student should correct all the mistakes from the macro- to the micro-level. To meet this requirement we can imagine that how much time and effort has to be taken by a student. Actually, due to the limited attention of a student, s/he is most likely to correct some

mistakes at the lower level at the expense of the problems at the higher level. In many cases, the efforts have been wasted since the sentence, even the paragraph where mistakes occur may be deleted when considering the whole discourse. Therefore, I strongly recommend that each revision should have a focus. This principle can be applied to the revising of the other parts of a thesis as well.

## *Common difficulties*

The difficulties graduate students often encounter at this stage are various. Some relate to the content and others to the organization. In this section, I will discuss some common difficulties along with suggested solutions.

### Difficulties in content

One difficulty in content is that the students feel overwhelmed by the materials they read, they have too much to review. They tend to think that they should incorporate everything they read into the literature review. Actually, the literature review is not a place to display your knowledge. What is required here is to build up an argument in which all the information should be pertinent to your own study. To avoid including irrelevant materials and giving too much space to less important topics, you may list subheadings from the most relevant to the least relevant while specifying its approximate length. For example, Anne intended to study the relationship between motivation and L2 achievement of non-English majors in China. This is the old topic that has been researched for several decades. Faced with a mountain of literature that deals with this topic, she did not know what to do. After some thinking and discussion with her supervisor, she realized that she did not need to detail the empirical studies concerning this issue in the field of education, nor the studies on the motivation of the children and middle school students in L1 learning. The emphasis of her review should be on the L2 motivation in and outside the Chinese context with a

focus on adults while very briefly summarizing the studies on L2 motivation of children and adolescents.

By contrast to the previous case, another difficulty concerning content is that students may feel that they have no empirical studies to review although they have read a lot. The argument they often advance is that there exist almost no empirical studies similar to their study found in the literature. However, this is not a good reason at all to justify the exclusion of reviewing empirical studies. For example, there might not exist an empirical study on the use of L1 in the process of L2 writing. However, there are quite a few studies on the use of L1 in L2 reading, or on the use of L1 in L2 speaking, or on the use of L1 in an L2 writing product. If this is the case, instead of trying to review studies on the use of L1 in the process of L2 writing, you may examine the literature from a wider perspective and search for the most pertinent studies to evaluate. Remember you should use the zoom very flexibly depending on the richness of your literature. When the materials are abundant, you need to zoom in and have a look at the materials most closely related to the topic. You should zoom out if materials closely related to the topic are scarce.

**Difficulties in organizing**

One organization difficulty is often shown in the overall structure where the subheadings do not form a natural flow of thought. In other words, the logical link among the subheadings seems to be absent. The reader would feel that the thoughts are fragmentary and do not form a coherent picture. I would suggest that you use the following general principles to organize the subheadings:

(1) from the general to the specific;
(2) from the abstract to the concrete;
(3) from the theoretical to the empirical;
(4) from the "long shots" to the "close-ups";
(5) from others' studies to yours.

Metaphorically, Cone and Forster (1993) suggest the "funnel" approach by which the review "begins with the general context and becomes more and more specific, ultimately focusing on specific criticisms of existing studies and leading to the specific rationale for the study being proposed." (p.105).

Another difficulty in organizing the review arises when a series of empirical studies need to be described and commented on. The students often do not know how to present them. The following are some of the most common ways (Cone & Forster, 1993):

1. Presenting them according to variables. For example, my own study (1993) examines the effects of learner modifiable variables on L2 achievement. The learner inodifiable variables include motivation, management strategies and L2 learning strategies. Thus the previous empirical studies on motivation, management strategies and L2 learning strategies are reviewed one after another.
2. Presenting them according to research designs. Usually within the quantitative camp, the review of this kind moves from weaker ones to stronger ones. For example, survey studies are presented before experimental ones, correlation analysis before the multivariate analysis, cross-sectional ones before longitudinal ones. In regards to quantitative and the qualitative studies, the quantitative ones are usually described before the qualitative ones. By this method, you can start with a summary in which a large number of less important studies are reviewed by groups, and then detail a few more important studies which are most directly pertinent to your study.
3. Presenting thetu according to the types of subjects. The classifications may be first made between second language learners and foreign language learners. Within each group, further classifications can be made, either in terms of learning conditions such as learners in formal learning environments and those in informal environments; or in terms of age such as children, adolescents and adults; or in terms of education

level such as pre-school children, primary school pupils, secondary school students, university students, etc.

4. Presenting them according to the designs in data-collection. Data may be collected through interviews, observations, thinking-aloud, diaries and questionnaires. Suppose your proposed study plans to collect data by observation which has not been used in the previous studies. The reason you chose observation as a data-collection technique is that interviews, diaries and questionnaires can only obtain self-reported data which might be distorted in one way or another. In this case, to organize by data-collection designs can well serve the purpose.
5. Presenting them according to findings. By comparing the studies concerned, similarities and differences in the findings are identified. If the conflicts in findings are one of the reasons for your study, you may illustrate how these findings clash.
6. Presenting them according to theoretical assumptions. For example, L2 proficiency is perceived differently by people. The concept of L2 proficiency is therefore defined differently in the empirical studies concerned. If one of the reasons for conducting your own study is to challenge the widely-accepted view of a certain concept, this way of presenting the existing empirical studies is effective.

Obviously, the choice of a particular way of organizing empirical studies is not an accidental decision. It highlights the importance of a particular aspect of the literature that usually coincides with what you wish to improve through your study.

### *Frequently-occurring problems*

The problems I have often observed in graduate students' literature reviews can be summarized as: lacking direct relevance and having no proper headings and clear signposts. In the following section, I will discuss these two problems one by one while giving some suggestions to solve them.

**Without direct relevance**

A review that contains materials without direct relevance is more often than not excessively long. The students tend to equate the length of a review with its quality. When the supervisor asks them to delete some sections, they often feel reluctant or even pained to do so. I would say, relevance is the first and paramount criterion for judging the quality of a review. If a review contains a lot of materials without relevance, rewriting is the only choice. There is no other alternative.

"Relevance" cannot be described in terms of rules since it is meaningful only in the context. Therefore, it is not possible to follow a set of rules to meet the requirement. What you are advised to do is to ask yourself the question: "How is this relevant to my study?" whenever you write something down. If you can justify the need for writing each part of your review, you will avoid the problem of being irrelevant.

**Without proper headings**

Graduate students are often not good at using headings to give the reader signposts. They may write one section in a review running for several pages without any headings. In reading such a review, the reader can easily get lost or become absent-minded since human beings have limited attention span and limited memory capacity. As a writer, we should be friendly towards readers by striving to help them overcome their limitations. By having proper headings, the reader can anticipate what s/he is going to read and can easily capture the flow of the argument.

Of course too many headings are not good either since they may distract the reader's attention and interrupt the smooth development of ideas. What is the proper number of headings? It is very hard to give a clear-cut answer. What you can do is to test your own sustaining ability of reading and make a sensible decision.

For some students, the trouble is not in the number of headings but in their wording. For example, there are three subheadings in one section where the first heading begins with a verb; the second one, with a present

participle; and the third one, with an adjective. Obviously, these three headings at the same level are not in the same pattern in terms of wording. Very often in the process of writing, it is difficult for you to detect such problems since they are placed in different places. You may use the computer program to produce the subheadings at the different levels once you finish the writing of one chapter. When the subheadings are put together, it is easy for you to overcome this problem if it exists.

## DESCRIBING METHODOLOGY

The Methodology chapter presents a detailed description of the way your proposed study is conducted. The description should be explicit, transparent and sufficiently detailed. By reading this chapter, the reader can easily visualize your research process and replicate your study if s/he wishes to. If compared to a literature review, this chapter is fairly easy to write since everything in the chapter is what you have experienced. However, it is difficult to meet the requirements of explicitness, transparency and sufficiency in detail. In the following sections, I will first describe the major components of the Methodology chapter and then discuss how to meet the requirements in writing each component.

### Major Components

The major components in the chapter of methodology are displayed in Table 13.1. As was discussed before, a study may employ the design of a survey study, or an experimental study or a case study, or a combination. Regardless of the research design, the headings or components are the same for the description of each design but with different content. When a study etuploys two different types of design, a common practice is to describe the two designs separately. In other words, you may have two sections or two chapters: one section/chapter on the survey study and the other on a case study.

The following section will discuss the writing of each component while taking into account the differences in describing different types of design.

| |
|---|
| Introduction |
| General and specific research questions/hypotheses |
| Subjects |
| Instruments |
| Data-collection |
| Data-analysis |
| Limitations |

Table 13.1: Major components in the chapter of methodology

### *Introduction*

The chapter begins with an introductory paragraph in which the specification of the research design is followed hy an overall description of the organization of this section/chapter. For the research design, you only need to write one or two short sentences. Suppose in your study you asked 300 English majors from 5 universities to answer a questionnaire to find out how their learning strategies are related to their English test scores on English Proficiency Test for English majors-Band 4. It would be enough if you could write something like the following: This study employs a survey design to examine the relationship between the use of language learning strategies and L2 achievement.

The general organization is a sort of an advance organizer that prepares readers for what is to follow. You may describe the organization in terms of the first-level headings without the need to provid detailed content within each subheading.

### *General/specific research questions/hypotheses*

In this section, all the general questions are presented together with specific questions. If the researcher has formed hypotheses instead of questions, then she/he should list all the hypotheses. This section serves

two purposes. The first is to give the reader clear information about what questions the study is intended to address. The second is to provide a framework for the Results chapter. In other words, all the questions appearing in this section need to be answered in the Results chapter. However, graduate students often forget to refer to what has been written in this section when they write the Results chapter. As a result, the Results chapter does not cover all the questions or the order of questions presented is not the same as that in this section.

Different from the other sections in the Method chapter, all the questions/hypotheses are stated in the present tense rather than in the past tense. The wording in a question should be clear and precise, and each question should address one issue.

The questions/hypotheses in this section are often not the same as were developed at the beginning of the research. The reasons are various. First of all, data may be analyzed and reanalyzed many times before the researcher is satisfied. For example, your hypothesis may be that there is no relationship between risk-taking and L2 achievement. The result could not reject the null hypothesis. Then you chose age as a moderator variable to make a further analysis. Accordingly, you add one more question in this section, i.e. is the relationship between risk-taking and L2 achievement affected by Age? In another case, you may delete one or two research questions developed earlier simply because you realized there would be too much to report if all the questions are included.

### *Subjects*

Subjects only refer to the people who are studied and measured by your instruments and whose responses are used in answering the research questions, such as the students who responded to the questionnaire or who were interviewed or observed. Participants are often used in an experimental study, such as students in an experimental group and in a control group. The Subject section describes how many subjects bave participated, who the subjects are, and how they are selected. It is simple

and easy to record the number of subjects who participated in the study. However, when the total number of subjects is changed in the process, the issue becomes slightly complicated. For example, in a survey study where the relationship between the use of language learning strategies and L2 achievement is examined, some subjects responded to the questionnaire on the use of the language learning strategies but did not turn up in the English proficiency test; and a few subjects forgot to write down their identity number so that it was not possible to match their questionnaire responses with their test scores. In a longitudinal study, the subjects who appear the first time may be absent the second time due to various reasons. In this case, you have to report the change in the number of subjects and explain the reason for such a change.

Personal information about subjects include gender, age and education level. Sometimes, the family background, the type of middle school and the type of university where the subjects studied are also needed if these variables can produce effects on a dependent variable. For simplicity of presentation, the information concerned can be displayed in a table while giving a brief verbal account. The procedure for selecting the subjects must be detailed so that other people can easily replicate them. If the subjects form a random sample, you have to specify what kind of sampling technique is used: simple random sampling, or systematic random sampling or stratified random sampling. When your study uses a convenient sample, you ueed to explain how this convenient sample has been selected. For example, this convenient sample may be an intact class or an intact school which you have selected simply because the teacher who teaches this class is or was your classmate, or because you yourself teach this class. Sometimes, the subjects are chosen on a voluntary-basis. In this case, you describe them as volunteers. If these volunteers are given a small gift for their participation, this also needs to be mentioned.

If in the data analysis, the subjects are further classified into groups according to their test scores or their responses to certain variables, then

the information about the newly-classified groups needs to be provided in the Subject section and the criterion for such classification needs to be explained.

In writing this section, one mistake graduate students often make is that they include those who are not subjects. For example, in the survey study mentioned above, some may regard as subjects the raters who score the test papers and some may take the test designers as subjects. To make it simple, you may make a decision simply by asking a question: Do their responses provide answers to the research questions? If the answer is "Yes", they are then subjects. Otherwise, they are not. Another problem I have observed is that the section is not detailed enough. Some graduate students fail to describe how they obtained a convenience sample; some do not explain why the number of subjects has changed and some do not state how the volunteers were recruited.

### *Instruments*

The Instruments section describes research tools used to measure dependent and independent variables in your study such as a language test to measure subjects' L2 proficiency, a questionnaire to investigate subjects' attitudes towards L2 learning and an interview schedule to find out how they go about vocabulary learning. The description covers the content, the categories, the reliability and validity if the instrument has been used before. For the self-designed instrument, the development of the instrument and the results of a pilot study should be added. The operational definitions of independent and dependent variables should be listed here. The example of describing a questionnaire in the chapter on Methodology is presented as in Appendix Four.

### *Data-collection*

In the Data-collection section, information about gathering data is detailed. The section has to address the questions "When were the data collected? Who were responsible for data-collection? How are the data collected? And where?" The information concerning "When" should go

beyond the specific date of collecting data. The questions as to whether the data-collection was taken in a regular lesson or in a recess or at other time should also be answered. The person who is responsible for data-collection may be the researcher himself/herself or a helper, or both. If helpers are involved, you have to describe to what extent the helpers can do a quality job. Suppose you had given them training before the data-collection. Then how the training was carried out needs to be explained. The way the data are collected needs a detailed account. For example, the information about the mode of administration (by person or by mail) and the form of interview (individual or group interview; recorded or unrecorded) should be available. In regards to the environment, the information about its quietness and comfort should be provided.

If the subjects were measured on several independent and dependent variables, the description of each measurement should be given. For a longitudinal study, the sections need to specify whether the conditions for measurement at different points of time were the same or not.

In some cases, the training of subjects or interviewers/observers is needed. For example, to collect thinking-aloud data, a carefully-constructed training for the subjects must be conducted since thinking-aloud is not a normal behavior in daily life. When more than one interviewer is required for the study, the training of the interviewers is essential for standardizing input to the interviewee. The procedures and the result of training have to be recorded in the Data-collection section.

### *Data-analysis*

The Data-analysis section is concerned with the procedures for analyzing the data and its reliability. The section of a quantitative analysis may begin with a general description of what kind of statistical package was used and what statistical analyses have been operated in your study. The introductory paragraph is then followed by a description of procedures of data-analysis, i.e. what was done sequentially in the process. Let's take for example the analysis of questionnaire data. The

description should address the questions "How missing values were dealt with? Which items had been reversed values? what were the reliability indexes of intended categories and what categories had been deleted?" The resulting categories that enter the final analysis are better listed in a table together with the number of items and the reliability index.

For the qualitative data-analysis, the procedures should be specified and categories generating from the data need to be explained. Furthermore, details of steps taken to increase the reliability of the analysis of qualitative data need to be reported. Suppose the score on a composition is a dependent variable, you have to tell the reader who were scorers and what is the correlation coefficient between the two raters' scores.

### *Limitations*

Limitations concerns those conditions that the researcher is supposed to meet but has had difficulty in implementing. The researcher should honestly inform the reader about them in the Limitations section so that the reader will deal with the findings with caution. The limitations may be shown in a sample. For example, the sample size may be too small or the sample is not randomly selected, or the participants are volunteers. The limitations can also be displayed in the data collected, for example, where the rate of missing data is high. The limitations can also exist in tests; for example, we do not have the index of reliability and validity of our English proficiency test for English majors-Band 4 and Band 8 although they are the only available authoritative tests for English majors in China.

The limitations of a study can also be alternatively put in the last chapter, i.e. the Conclusion chapter. The advantage of such an arrangement is that you can discuss in detail how the limitations can influence the internal and external validity of the findings yielded from the study.

# REPORTING RESULTS AND DISCUSSION

The common practice is to present results together with the discussion. That is to say, in a thesis, results and discussion are integrated into a single chapter. The advantages of having a combined chapter are several. The obvious one is that it can avoid restating the results when you discuss them. The second advantage is that it is much easier for the reader to follow the development of the thesis.

The alternative practice is to have results and discussion separately presented, that is, the first half of the chapter for Results and the second half for Discussion. The people who are in favor of this alternative practice think two separate sections are neat and easy to manage. Personally I feel it more natural that the reporting of the results is immediately followed by discussion.

In this chapter, I will first discuss how the results are presented and then how the discussion is developed.

## Presenting Results

This section will begin with the question as to what is a result, then move to the question as to how results are organized. Finally, it describes how quantitative and qualitative results are presented.

### *What can be regarded as results?*

Results reported in this chapter are the answers to the research questions and the findings yielded from your study. However, not all the findings from the study are results that need to be reported since some findings may not directly address the research questions. Novice researchers often equate the findings with the results. Therefore, their report is dictated by the findings rather than by the research questions. To avoid this problem, the researcher should stick to the questions all the time and not be tempted to report the findings that appear to be interesting. In other words, the researcher needs to focus only on the findings that answer the research questions.

### *How are results organized?*

The best way to organize results is by research questions. To begin with, you had better restate the research questions in the introductory paragraph of the Results chapter since we can neither expect the reader to remember all the questions by one reading nor expect them to go back to the Method chapter from time to time to get the precise information about the questions. The subheadings in this chapter had better be presented as shortened references to the questions being answered. In other words, subheadings are used to describe the content of the question. As a result, by reading a subbeading, the reader can link specific results easily with the research question concerned.

One common problem in organizing the results by graduate students is that they like to organize the results by the type of statistical analysis such as T-test or ANOVA, or Multiple regression, etc. For example, they use as subheadings "The results of *t*-tests; The results of ANOVA". These subheadings appear to be neat and clear. However, statistical analyses are the tools for answering questions. The names of statistical analyses do not show how the results are tied to your research questions. The other problem I have found is that the order of the results is not the same as the order of the researcb questions/hypotheses stated in the Method cbapter.

Many studies have both quantitative and qualitative results. Then how are these two types of results presented? First of all, you should decide how these two types of results are linked to the research questions. Suppose the qualitative results are further illustrations of the quantitative results and both are related to the same research question. In this case, the qualitative results are better to be integrated into the quantitative part and the quantitative findings are reported before the qualitative ones. When these two types of results address separate research questions, reporting them together will be inappropriate.

### *How are quantitative results reported?*

Usually quantitative results are reported by tables/graphs coupled with a verbal account. These two different forms provide the reader with parallel information, numeric and narrative, that are complementary to each other.

Generally speaking, reporting quantitative results in a narrative form contains three basic statements (Rudestam & Newton, 1992). The first one is a signal statement indicating to the reader which table contains the results concerned. The second one, a technical description statement describes the findings in a technical way. The third one is a non-technical description statement. Accordingly, it presents the results by using non-technical terms. The following are typical examples of the three basic statements in a hypothetical study.

A signal statement:

The means and standard deviations of the use of different strategies by three groups of students are displayed in Table 1.

| Group Name | Management Strategy | | Mother-tongue strategy | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| High-achievers | 4.25 | .464 | 2.32 | .521 |
| Average-achievers | 3.52 | .521 | 3.05 | .563 |
| Low-achievers | 3.05 | .400 | 3.68 | .501 |

Table 1: the use of Management strategy and Mother-tongue strategies by three different groups

A technical description statement (A factual finding):

As shown in Table 1, the mean of the responses to Management Strategy by High-achievers is the highest (i.e. 4.25 on a five-point scale) among the three groups and the average mean of the responses to Mother-tongue Strategy is the lowest (i.e. 2.32 on a five-point scale); while the average mean of the responses to these

two strategies by Lower-achievers are just the opposite to High-achievers, i.e. the average mean of their responses to Management Strategy is the lowest among the three groups (i.e. 3.05) and to Mother-tongue Strategy is the highest (i.e. 3.68). The average-achievers are somewhat in between. Specifically, their mean of the Management strategy is lower than that of High-achievers but higher than that of lower achievers while their mean of the Mother-tongue strategy is higher than that of High-achievers but lower than that of Low-achievers.

A non-technical description statement:
It means that among the three groups, High-achievers used Management Strategy most frequently, Lower-achievers used it the least frequently and Average-achievers in between. On contrast, Lower-achievers used Mother-tongue Strategy the most frequently, High-achievers used it the least frequently and Average-achievers in between.

How can you make a tabular or graphic report effective and efficient? Two thing need to be considered. First of all, if a table can serve the purpose, don't use a graph since a graph occupies a lot of space in a thesis. Secondly, information is more easily processed when presented in columns than in rows. In other words, when the comparison of a set of numbers is the purpose, you had better place them in columns rather than in rows. More details on how to produce a table will be discussed under the heading of Tables in Chapter 14.

### *How are qualitative results reported?*

Qualitative results are primarily presented in a narrative form but occasionally supplemented by a figure or a table. Different from the narrative report of quantitative results which can be presented in three

types of statements, narrative reports of qualitative results are varied from one study to another.

For a single case study, you need to give a detailed account of the case centering around the research questions. Suppose a case study focuses on the developmental stages of the child's L2. You might then need to describe each of the distinct stages verbally coupled with illustrative examples. In a multiple case study, similarities and differences across the cases are the focus of your report. Meanwhile, you need to use a lot of direct quotes from interviews, think-aloud protocols, diaries, etc. or restate what bas been said or done by the subjects to concretize the similarities and differences identified.

Although the ways qualitative results are reported are diversified, they should share one thing in common, i.e. a description must be made through analysis and synthesis since no reader wants to read one quote after another from the interview scripts without knowing why the quotes are listed and how these quotes are related to the research questions. Sometimes, qualitative results can be described in a figure. For example, the different stages of the child's L2 development may be shown in a flow chart. In the study conducted by Wen and Guo[18] (1998), the dynamic functions of L1 use in producing sentences are diagrammatically presented.

## Discussing Results

Each result reported should be followed by a proper discussion. This is the place to illustrate the link between your study and the existing literature and to demonstrate your comprehensive understanding of the research topic under investigation. Hŏwever, quite a few students do not do justice to the Discussion section. As a result, the poor discussion relegates their theses to mediocre ones. One important reason is that they

[18] The figure is in the paper presented as an appendix.

feel that they have nothing to write in this section. Some of them even cannot differentiate reporting results from discussing the results. In the following section, I will first explain to you what is meant by discussing a result, then give you some suggestions on how to make a discussion comprehensive and appropriate.

## *What is meant by discussing a result?*

When you discuss a result, you should go beyond the result and express your own opinions about the result, i.e. your personal speculation on the reasons for a result, your judgement on its significance, implications, and the possible direction for future research, etc. The essential difference between reporting a result and discussing the result is that the former is fact-driven while the latter is opinion-driven. However, your opinions are not whatever you have thought about but they should be based on either existing theories or previous studies or common sense. Therefore, in discussion, references are indispensable. Usually the discussion going together with the report of a result cannot be too long since a lengthy discussion will separate the answers to the research questions and it will distract the reader's attention. The elaboration of a discussion, if it is necessary, can be made in the last chapter of Conclusion

## *How is a result discussed?*

### Explanations for a finding

Once a specific result is read, one question that the reader will immediately ask is "Why there is such a result?" Therefore, discussing the result usually begins with speculating about the explanation for the result. The speculations are not wild guesses. They should be logically sound. That is to say, they can be supported by the existing theory or empirical findings from other studies. The weakest explanation would follow from our intuitive feelings. Furthermore, the speculations should take into account rival explanations or the limitations in your study. The

following is the example of discussing a result in my own study (Wen, 1993). The specific result discussed here is that there is a strong correlation between gender and L2 achievement. In other words, female students tend to achieve higher scores than male students on English Proficiency Test for English Majors-Band 4.

> The reasons for such a strong correlation can be various. One interpretation for such a gender difference is that females have talents for language learning while males have talents for mathematics and sciences. Such a view has been reported by quite a few researchers such as Allen and Valette (1977), Farhady (1982), Larsen-Freeman and Long (1991), Maccoby and Jacklin (1974), and Nyikos (1990). The alternative interpretation is that the male students who major in English are not the best students in the whole group since according to the Chinese tradition, the best male students usually go into the science stream. Following this argument, we will challenge the explanation that females have greater talents for language learning (pp. 143-144).

**Evaluations of the significance of a result**

Evaluating the significance of a result also falls into the scope of discussing. The significance may be theoretical or practical. By evaluating its theoretical significance, you try to link the result to the previous findings or existing theories to see whether your result is the same as the previous ones or in conflict and to see whether your result is in support of the existing theoretical assumptions or against them. Remember this is a place where you frequently refer back to the literature review. If your result echoes previous studies, it means that the reliability of previous findings is increased. Otherwise, you need to offer the reader various possible explanations. If the conflict is not caused by methodological problems, it will add complexities to our understanding of a phenomenon and enrich the existing theory.

Very often, a result has not only theoretical significance but also practical significance. By pointing out its practical significance, you try to specify how L2 teaching and learning can benefit from the result. In other words, you try to give practical suggestions based on the result for improving L2 teaching and learning. One thing you have to remember is that your suggestions must naturally grow out of the result. If the suggestions, although they are conducive to L2 learning, are not related to the result, they have to be deleted.

**Making recommendations for future research**

The discussing of a result may move further to make suggestions for future research. Suppose your result is not consistent with the previous findings. Your speculated reasons are: (1) the sample size of your study is not big enough; (2) your subjects were adult learners but the previous studies all involved non-adult learners. Therefore you suggest the study needs to be replicated with a bigger sample size in future. Once you embed the recommendations in the context of your own study, you can avoid a common problem occurring in graduate students' theses, i.e. the recommendations made are not developed out of their own findings.

## *Problems in discussing a result*

The worst case is that the discussion component is simply absent in the Results and Discussion Chapter where only a set of tables and figures are presented together with a few lines describing the results. Remember it is the discussion part where the extensiveness of your background knowledge and the depth of your understanding are demonstrated. Without this part, the reader would think your background knowledge is limited and your understanding of the issue is shallow. A more common problem is that the explanation for the result is simple and sometimes even illogical. To overcome the problem, it is best to discuss the results with your classmates and supervisor. Through the discussion, your understanding of the results can be deepened. Furthermore, you should

refer back to the previous studies mentioned in the literature review to see whether your findings are consistent with theirs or not.

## WRITING THE CONCLUSION CHAPTER

Human beings have limited capacity to remember things. Therefore, it is not uncommon that the reader has forgotten some important things of a thesis when s/he finishes reading the Results and Discussion chapter. The last chapter of Conclusion can help readers refresh their memory and review what is essential in the thesis.

If you think you have already said everything in the previous chapters and the last chapter does not add any weight to the thesis, you would be proved to be totally wrong since the last chapter will certainly leave a deep impression on the reader if s/he cannot remember what has been read before. In this sense, the last chapter weighs more in determining the quality of your thesis. This is particularly true for the external examiner and the members of the oral defense committee when they are drowned in reading a dozen theses in the season of oral defense. Therefore, I strongly advise that you write this last chapter with conscientious efforts.

The Conclusion chapter usually consists of three sections: (1) major findings; (2) implications; (3) recommendations for future research. I will discuss with you how to write each part more effectively.

### Major Findings

The major findings are an integrative summary of the most important results. Therefore, you cannot repeat verbatim what has been said in the Results chapter. To organize this section, there are two alternative ways. The first is to present them by research questions. That is, to begin with, you briefly restate each question in order to save the reader's trouble of turning back to the earlier section. In most cases, you do not need to go beyond the general questions. The results are then summarized in relation to the questions. The second alternative is to list the most important findings without referring to individual questions. The common practice

is to sequence the major findings that mirror the order of research questions.

In either case, you should not use jargon and statistics. What is expected here is a non-technical verbal account of major findings, which is usually short and brief. In addition to the major findings related to the research questions, you may also report unexpected results, i.e. the results yielded by analyses that you did not plan to make originally but later proved to be interesting.

## Implications

In the previous chapter when each result was discussed, you have already drawn implications for theory and for practice. However, the implications were discussed there with specific results. As a reader, what s/he has seen is only trees scattered around rather than a forest. Now it is the time for you to synthesize the implications to present the reader with a whole picture. You need to take a top-down approach, i.e. you overview the results and think of their implications along three perspectives; theory, methodology and practice.

### *Theoretical implications*

In the Literature Review chapter, you examined different theoretical assumptions about your topic. Most likely, the assumptions examined are various and some of them are in conflict. Are your results in support of one or more theories? Why or why not? Remember your interest is not in implications drawn from one specific result which are presented in the chapter of Results and Discussion. Rather, you talk about a bunch of results that are related to a similar theoretical issue. If your results are not predicted by the existing theories, you have to explain why it happened in this way.

The following is an excerpt discussing the theoretical significance of the results in my study. Due to the limited space, the quoted section only

mentions one theoretical implication. Actually, from one study you may draw several theoretical implications.

In Section 2.3.4, Chapter Two, the three controversies or continua (i.e. L1-L2 connection, the explicit-implicit option and the code-communication dilemma) proposed by Stern were introduced. Two types of learning approaches (i.e. the Traditional and the Non-traditional) are assumed to be placed at the extreme ends of each continuum. The popular view in the field of L2 teaching and research is that these two approaches cannot be reconciled and the Traditional Approach produces poor learning outcomes but the Non-traditional Approach leads to successful learning.

The findings from this study seem to provide a completely different view about these controversies from the learner's perspective. Successful learners do not perceive the ends of each continuum are in conflict and cannot be reconciled; rather, they regard them as complementary to each other and are active to combine them in carrying out learning activities When learners are only in favor of either the Traditional or the Non-traditional Approach, they obtain similar results. However, they cannot make outstanding achievement as those who actively use both approaches.

One theoretical implication can be drawn from the above findings, i.e. prejudice against the Traditional Approach should be eradicated and too much enthusiasm about the Non-traditional Approach should be weakened. Those people like Krashen and his followers need to reexamine their theories and modify their views that only implicit learning and implicit knowledge can lead to proficiency of L2 learning but explicit learning and explicit linguistic knowledge are of very limited function. However, it is uncertain whether the aforesaid findings from this study are universally applicable. If empirical findings from future studies on L2 learning in different contexts are found to be similar to those obtained from this study, their theories should include the views that both approaches have an equal role to play in L2 learning and the best way to learn an L2 is to combine two approaches flexibly in accordance

to the different nature of tasks. If not, their theories have to take the factors of culture and/or learning conditions into consideration (Wen, 1993: 249-250).

## *Methodological implications*

The development of research methodology is also very important for research advances. Do your results make any contribution to this area? If one of the reasons for you to carry out the study is to employ a new method, your results will definitely have methodological implications. For example, I have drawn two methodological implications from my study. First of all, I used path analysis to examine the relations of a set of learner variables to L2 learning outcomes. In the previous studies, the researcher usually used multiple regression to solve a similar problem. The finding from my study revealed that path analysis is much more powerful than the multiple regression since the latter can only show the relative importance of each variable in predicting L2 learning outcomes but the former is able to go beyond that, displaying the relations among the independent variables. Secondly, in my study I used the qualitative data to illuminate and supplement the quantitative data. It turned out that the qualitative data were more revealing than the quantitative data. The implication from this finding is that the combination of quantitative and qualitative designs is much more effective than a single method.

## *Practical implications*

Research in the applied linguistic field may have practical implications. Furthermore, you may go beyond the description of the potential use and discuss further how these suggestions can be implemented. For example, the following is the first paragraph under the heading "Practical implications" in my thesis (Wen, 1993: 250):

**12.3 Practical implications**

Practical implications drawn from the findings from this study will be discussed in terms of two areas. The first is how teachers can make use of the findings concerning the overall relations between modifiable learner

variables and English achievement to promote L2 learning efficiency and independent learning. The second is how teachers can make use of the findings in relation to the effects of learning approach on English achievement to avoid their own bias towards learners' learning approaches and at the same time to make learners discard their biases through daily teaching.

## Recommendations for Future Research

In this section, you need to make suggestions for future research on the topic under investigation. The section of Recommendations should be discussed at a macro-level as well as a micro-level . In other words, you should talk about the general directions as well as suggestions on some specific research topics. The following are the first few paragraphs under the heading "Suggested research topics" (Wen, 1993: 258):

The suggested research topics are divided into two sets: one set is at a macro-level and the other at a micro-level. The topics at the macro-level address the relationship of a whole set of modifiable learner variables to English achievement. The topics at the micro-level concern individual modifiable learner variables, or relationships between two types of variables.

1) Macro-level

First of all, the same study can be replicated with more English majors. The findings can present us with a complete picture of advanced level English learning in China. The same study can also be conducted on middle school students and university students who major in subjects other than English. The results can be used to delineate the whole picture of the relation of the modifiable learner variables to English achievement of Chinese EFL learners. By carrying out a series of such studies, hopefully, a standardized questionnaire on learner factors can be developed and a causal model of factors affecting the English achievement of Chinese EFL learners can be established. If possible, such a study can be undertaken

cross-culturally. The findings obtained from different cultural contexts would be insightful for constructing better L2 learning theories.

Secondly, ...

Thirdly,...

Fourthly,...

Fifthly,

To sum up, future research on the relationship between modifiable learner variables and English achievement can be conducted on a large sample cross-sectionally or longitudinally, and cross-culturally if possible...

2) Micro-level

... There are three areas (i.e. vocabulary learning, the use of mother tongue and tolerating ambiguity), according to the findings from this study, which most need to be further investigated. For vocabulary learning, future studies can be carried out in middle school students , non-English majors and English majors. Such studies may focus on the differences between successful and unsuccessful learners in their development of strategies....

The suggested topics must be developed out of your own study. They should be genuine topics that need to be investigated. Some graduate students, however, write this part without sufficient thought and topics proposed are not significant or do not derive from their research.

## SUMMARY

The Introduction in a thesis presents the reader with an overall picture of the study, typically including three components: (1) a general description of the study; (2) the need for the study and (3) the overall structure of the thesis. The Literature review is the place where your study is situated in its context. It may cover five topics as follows: (1) conceptual definitions of key terms; (2) an examination of the research topic with reference to a specific theoretical perspective; (3) a description

of related empirical studies with evaluating comments; (4) a critical review of research methods including instruments related to your study if the methodological issue is critical for your study; (5) a conceptual framework if there is any. The description of Methodology is a detailed description of the way your study is undertaken. Specifically, this part should answer the questions "What are the research questions?", "Who are the subjects and how were they selected?", "What are the instruments and how were they developed?", "How were the data collected and how were the data analyzed?" and "What are the limitations?" The description should be explicit, transparent and sufficiently detailed so that other researchers can easily replicate the study if they want to. The part of Results and discussion should be organized in terms of research questions. Only the findings that answer the research questions are reported. The discussion of a result is expected to offer reasonable explanations of the result, evaluate its significance and make recommendations for future research. The Concluding part is to summarize the major findings, discuss the theoretical and practical implications and propose suggestions for future research. Different from the part of Results and discussion, your discussion in the Concluding part need not follow specific research questions. Instead, it centers around general issues that the study attempted to resolve.

## DISCUSSION QUESTIONS

Get a copy of a thesis from the library of the School of Foreign Studies to answer the following questions:

1. What do you think of the Introduction? Is it properly written?
2. How is the Literature Review organized? Can you make suggestions for further improvement?
3. Does the Methodology contain sufficiently detailed information?
4. How are the results reported? Were all the research questions answered?
5. Are the headings appropriate? If not, can you revise them?

6. Is the discussion sufficient and appropriate? If not, can you elaborate on it?
7. What do you think of the last chapter? Can you make some critical comments on it?

# 14. Writing Style

Writing a thesis is not like writing a poem, a novel or an essay. In creative writing, you exert efforts in order to be different from the others both in content and in form. However, a thesis has a required writing style. Although the studies reported in theses are varied from one to another, the style remains the same. As a graduate student, you have to be familiar with such a writing style and make sure that your thesis can meet the required standard. This chapter begins with the features of academic writing style, followed by various problematic aspects in writing a thesis.

## ACADEMIC WRITING STYLE

A thesis should be written so that the information is presented as briefly as possible and as clearly as possible. Its writing style differs from creative writing in two major aspects: the organization and the language (Newman et al., 1997).

### Organization

As was mentioned at the beginning of Chapter 12 (An overview of thesis writing), a thesis includes a beginning, a body and an ending. The components in each part are somewhat standardized. For example, the body of the thesis must have one chapter on literature review, one chapter on metbodology, one chapter on results and discussion and one chapter on conclusions. As an individual writer, you have no freedom to drop out any component or to add any new component that has never occurred in other theses.

For the structure of each chapter, you have to provide the reader with an introductory paragraph under each major heading. In such an introduction, you need to produce an advanced organizer to tell the reader what is to follow. At the end of each chapter, you summarize what was

said in the whole chapter. In between, you often provide transitional sentences between paragraphs or sections by which you provide closure to what has been said previously and what is going to be said in the next paragraph or section. Subheadings are frequently used to alert the reader when new topics are introduced and to remind the reader of the important things when s/he finishes reading one section. Tables and graphs are often used for the clarity of presenting information.

You should make your purpose in writing explicit in the thesis so that the reader is not kept in suspense at any stage or won't have any question like "why is this part written here?". After all, as a writer, you should keep the reader in your mind all the time.

## Language

In a thesis, the language used should be formal and succinct. The passive voice is more often used than the active one. The first person hardly appears in a thesis. Flowery words and words indicating emotions such as fantastic, fascinating, terrific, wonderful, unfortunately, amazingly etc. should be avoided.

Tense is another typical feature of the language used in a thesis. Normally speaking, you move back and forth between the past and the present tenses. The improper use of tenses may violate scientific ethics (Day, 1989). According to the scientific ethics, once a paper is formally published, it has become established knowledge while your present study before its publication is not regarded as accepted knowledge. Accordingly, you use the present tense to quote others' published studies to show your respect while using the past tense to describe your own study to display your modesty. The following are the situations for the present tense and the past tense:

1) Present tense:

a) Describe needs for the study.

b) Describe the finding in the previous studies presented in a journal.

c) Discuss the existing theories and well-known principles.

d) Describe the results in a table or a figure.

e) Discuss the results in your own study.

f) State the conclusions of your own study.

g) Put forward suggestions for future research.

2) Past tense:

a) Describe the methodology of your present study

b) Report the results yielded by your present study.

c) Describe the procedures in which a previous study was carried out.

## APA WRITING FORMAT

Theses in the field of applied linguistics require that the manuscript be typed in the format described in a manual published by the American Psychological Association (APA). Such a format is called APA style in short. In this section, I will introduce to you some rules that you need to follow in writing your thesis.

### Tables

As was mentioned earlier, academic writing often uses tables to present information together with a verbal account. A well-designed table can be more effective and economical than a verbal account alone since a table can enable the reader to perceive patterns and relationships in the data that are not readily seen in the verbal description. A table is typically used to display quantitative findings. Occasionally it does present qualitative results, too. In the manual published by APA, there are a set of rules regarding constructing a table that consists of five parts: number, title, headings, body and notes. The rules concerning each part will be briefly described below.

#### *Tables or figures*

It is a common problem that graduate students do not make a distinction between tables and figures, and thus in their thesis there is confusion. Tables are made of rows and columns containing numbers

while figures refer to any type of illustrations other than tables, such as charts, graphs, drawings and photographs.

### *Numbering tables and figures*

All the tables and figures in a thesis must be numbered. According to the APA style, the tables are numbered sequentially in the order they occur in the manuscript and the figures are numbered as a separate sequence.

Suppose you have altogether 30 tables in your thesis. You might like to start with Table 1 and end with Table 30. However, this way of numbering tables, in my opinion, is suitable for a paper but not convenient for a thesis since in a thesis, dropping out or adding one table is not uncommon. If the tables are numbered from the beginning to the end, that means you have to renumber the tables all the way through whenever there is a change. Therefore, I suggest that tables be numbered within chapters. For example, tables in Chapter One are named as Table 1.1, Table 1.2 and so on, and tables in Chapter Two, as Table 2.1, Table 2.2 and so on. Doing this can make revision much easier. Furthermore, tables must be sequenced by Arabic numbers. Therefore, you are not allowed to use Table A and Table B, nor Table I and Table II.

Like the tables, the figures are also numbered within chapters and sequenced by Arabic numbers. They form a separate sequence as Figure 1.1, Figure 1.2 in Chapter One and Figure 2.1, Figure 2.2 in Chapter Two.

### *Table titles*

Each of the tables must have a title. The title should be short but explicit and self-explanatory. It should contain major information about the tahle. Titles such as "The results of t tests" or "The results of Multiple regression" are not good ones since such titles do not contain any specific information about the results. They had better be changed into: "*t*-tests: Differences in test scores on CET-Band 4 between the control and experimental groups" and "Multiple regression: Power of variables in predicting the students' scores on CET-Band 4".

### *Table headings*

Headings are used to organize a table and inform the reader what has been presented in a row or a column. If the headings are abbreviated, the meaning of the abbreviations should be obvious. Otherwise, you need to explain the meaning of each abbreviation in a note.

### *Table body*

The body of a table refers to the data which are in most cases numbers. For the sake of readability, round numbers are always preferred if possible. Which should be put in a row and whicb should be put in a column depend on several factors. One is the limitation of space. Usually you have more space for rows than columns since the width of a paper is shorter than its length. The other is the way your eyes perceive numbers. It is easy to perceive numbers in a column than in a row particularly when you want to compare numbers.

### *Table notes*

Table notes can be classified into three kinds: general notes, specific notes and probability notes. General notes provide information about the table as a wbole, for example, they explain the meaning of abbreviated terms used in a table. According to the APA format, a general note is shown by the word "*Note*." The word is italicized and ended by a period. Specific notes for stating the content of a specific cell that are denoted by a superscript lowercase letter. If several items need to be explained by a specific note, you may mark them from the upper left of a table and move from left to right across rows. Probability notes display the significance level of a result. Within a table this is marked by one or two or three asterisks. "*" means .05; "**" means .01; "***" means .0001.

If a table requires all the three kinds of notes, you start with the general notes, then specific notes and finally probability notes.

## *Examples*

Table 14.1 and Table 14.2 are examples of presenting quantitative and qualitative results respectively.

### Presenting quantitative results

| Type of learner | L1 Prof.[a] Mean | S.D. | L2 Prof. [b] Mean | S.D. |
|---|---|---|---|---|
| Grp 1 (Trad) | 73.10 | 5.63 | 90.37 | 4.12 |
| Grp 2 (Non-trad) | 73.29 | 5.38 | 91.08 | 3.51 |
| Grp 3 (HB) | 74.02 | 5.51 | 91.62 | 3.40 |
| Grp 4 (LB) | 73.10 | 5.04 | 91.00 | 3.74 |

Table 14.1: ANOVA: No differences found amongst the four groups of learners in L1 Prof. and L2 Prof.(Wen, 1993: 185)

*Note*. Trad = traditional learners; Non-trad = non-traditional learners; HB = High-balanced learners; LB = Low-balanced learners. } General Note

a: L1 proficiency is represented by scores on Matriculation Chinese Test.
b: L2 proficiency is represented by scores on Matriculation English Test. } Specific Note

### Presenting qualitative results

| What | Why | How | When |
|---|---|---|---|
| Listening to magazine show on the VOA | ● Practice listening<br>● Increasing background knowledge<br>● Practice writing | ● Take down notes<br>● Write down what has been retained in the memory in my own words | Three times a week |
| Memorizing new words (new words encountered in a week) | Increase vocabulary | Use cards | Once a week |
| Fast reading (simplified stories) | Increasing reading speed | Read as fast as possible | Every night before going to sleep |

14.2: The self-study plan made by a good learner based on her interview data (Wen, 1993:194)

## Numbers

How is a number dealt with in a thesis? The APA manual has a series of rules specifying the ways a number is expressed. Now we single out several common areas of concern for discussion.

### *Figures or words?*

Is a number expressed as a figure or as a word? A general rule is that numbers are expressed as words when numbers are smaller than 10; numbers are expressed as figures when they are 10 or above. However. there are some exceptions. For example, when numbers below 10 are used together with numbers 10 and above, they are expressed as figures. For example, it is correct to say "5 (Note: five is not correct.) out of the 36 second-year students" and "they are ranked 3rd (Note: third is not correct.)and 15th respectively".

When numbers below 10 are used to represent time, ages, scores and points on scales, they are also expressed as figures as exceptions. For example, it is perfectly acceptable to say "They finished the test within 2 (Note: two is not correct) hours" and "Her score on the composition is 5 (Note: five is not correct)."

### *Decimal fractions*

When a number cannot be bigger than 1 such as correlation coefficient and probability values, you do not need to put a zero before a decimal fraction, such as .40 or .05. However, when a number can be greater than 1, a zero is needed before a decimal fraction, such as 0.40 and 0.05.

### *Plurals of numbers*

The plurals of numbers are expressed by simply adding "s" alone without an apostrophe. For example, we should say "in the 1990s" rather than "in the 1990's" or "between the 80s and 90s" rather than "between the 80's and 90's".

## Spelling

The publication manual of the APA deals with several important issues concerning spelling, such as preferred spelling, hyphenation and capitalization. In the following part, these issues will be discussed one by one.

### *American spelling or British spelling?*

According to the APA style, you need to follow the standard set up by Webster's New Collegiate Dictionary. Evidently this is American spelling. If you prefer to use British spelling in your thesis, you are definitely allowed to do so. The only inconvenience is that the computer spelling check system normally follows the American standard. Whichever you choose, you must be consistent.

### *Hyphenation*

One difficulty in spelling is how to use a hyphen when compound words are involved. In most cases, you can find a solution in a dictionary. However, quite a few compounds may not be available in a dictionary. If this is the case, you can follow general principles provided by the Publication Manual of the APA. The first principle is that if the meaning of a compound adjective is clear and cannot cause any misunderstanding, you should not use a hyphen, e.g. modifiable learner variables, L2 linguistic competence. The second principle is that if you invent a new compound that is used as an adjective before a noun, you had better use a hyphen to avoid a potential misunderstanding. For example, if "different word lists" refer to the lists with different words rather than the word lists that are different from other word lists, a hyphen placed between "different" and "word" can help express the intended meaning explicitly. The third principle is that a hyphen is often used when a compound adjective occurs before a noun. If a compound adjective is after a noun, a hyphen is usually not needed. The following are examples:

learner-centered classroom but not the classroom is learner centered

t-test results but not results yielded by t tests (APA, 1992: 57)

The fourth principle is that "if two or more compound modifiers have a common base, this base is sometimes omitted in all except the last modifier, but the hyphens are retained" (APA, 1992:57). For example,

long- and short-term training, five- and ten-year plan.

## *Capitalization*

A complete sentence should begin with a capital letter. This is a rule every second language learner is familiar with. However, in the writing of a thesis, you may come across some cases where you will be uncertain about the rules of capitalization. In the following part, I will introduce to you some important rules concerning capitalization stated in the APA Publication Manual.

(1) When a complete sentence is placed after a colon, the first letter of the first word should be capitalized. For example,

> We can draw an inference from the findings: The Traditional and Non-traditional approaches are equally effective but they are less effective than a combination of these two approaches.

(2) In titles of books and articles, the first letter of content words such as nouns, verbs, pronouns, adjectives, adverbs and prepositions that consist of four letters or more are capitalized while the function words such as conjunctions, articles and prepositions with less than four letters are not. If a title contains a compound word with a hyphen, both words need to be capitalized. However, this rule does not apply to the titles of books and articles in a reference list.

(3) When a noun is followed by a numeral or a letter that has a specific place in a sequence, the first letter of the noun is capitalized, for

example, on Day 1 of Semester 1, Group A and Group B did, in Table 2 and Figure 7.

## Quotations

Quotations are used only when they are more powerful and more effective than restating the material in your own words. They are either positive or negative statements. That is to say, they are in support of your argument or opposite to your own views. When a quotation contains fewer than 40 words, you do not need to put it as a separate paragraph. But when it has 40 words or more, it should be a block quotation (See Quotation 3). The block quotation begins as a separate paragraph where each line is indented five spaces from the left margin and each subsequent line is flush with the paragraph indent. The block quotation does not need any quotation marks.

No matter whether a quotation is inserted in a text or a separate paragraph, the author, date and page number should all be specified with accuracy (See Quotation 2). If you want to produce any changes in the original material, you have to follow the rules specified in the APA Publication Manual.

According to the rules in the APA Publication Manual, some changes from the source are permitted without any explanation, such as changing the first letter of the first word in a quotation and changing the punctuation mark at the end of a sentence in order to avoid syntactic errors. For other changes, you bave to provide a sort of explanation in one way or another. In the case of omitting materials, you are required to use three ellipsis points (...) (See Quotation 3) within a sentence and four ellipsis points (....) between two sentences to indicate that you have omitted materials from the original source. When you insert some materials in the original source, you should use brackets ([ ]) (See Quotation 3) rather than parentheses to enclose the inserted materials which may be additions or explanations. If you want to emphasize some part of the original materials, you can underline the emphasized part and

italicize it. Immediately after it, you insert brackets in which the words italics added are placed (See Quotation 3).

### *Examples*

Quotation 1:

> Conceptually, Gardner (1985) sees motivation as "the combination of effort plus desire to achieve the goal of learning the language plus favourable attitudes toward learning the language" (p.10).

Quotation 2:

> The following section will discuss empirical studies on learning strategies in terms of their findings and methodological problems. If we "review the whole of the learner-strategies research, we have to say that the area is at an embryonic stage. Conflicting results and methodologies proliferate. There are few hard findings." (Skehan, 1989, p. 98)

Quotation 3: A block quotation

> Rubin (1987) more explicitly states the importance of studying students' beliefs in her review of the research on learner strategies:
>
> > ...to better understand how learner strategies come to be used, it is essential that we account for a learner's knowledge about language and his/her beliefs about the language learning process [that is, his/her views about how learning can be successful.] because his knowledge can *form the basis for selecting and activating one strategy over another* (italics added) (p. 19)

## Citations

Citations refer to the surname of an author, the year of publication, and page references if specific words or arguments are drawn from an author in parentheses. Such information can help readers to locate the source of information in the reference list at the end of a thesis. The APA publication manual describes how to make a citation in different situations. The following section will introduce to you how you cite one

work by a single author, two authors or more than two authors; and how you cite two or more works by different authors.

### *Citing one work*

The simplest case is that you cite one work by a single author. You need to put in parentheses the surname and the year of publication with a comma in between (See Example One). If the surname of an author has already appeared in the text, you just put the year of publication in parentheses immediately after the surname (See Example Two).

Example One

It has been argued that teachers' role is to provide the students with optimal conditions which can facilitate learning so that students can achieve similar successful results (Bloom, 1976).

Example Two

Gagne (1977) also noticed that adult learners were less affected by external instruction events.

If the cited work was written by two authors, you always cite both authors (See Example Three). However, when more authors but fewer than six authors are involved, you are only required to cite all the authors the first time the reference appears in text (See Example Four). In subsequent citations, you simply cite the surname of the first author followed by "et al."(See Example Five).

Example Three

The disadvantage of the multiple regression analysis is that it cannot show the complex interrelations between independent variables (Bryman & Cramer, 1990).

Example Four

Studies of the good language learner (for example, Naiman, Frohlich, Stern, and Todesco, 1978) have tried to identify the strategies which successful learners use (first citation) (Ellis, 1994: 37)

Example Five

Naiman et al. (1978) found a similar relationship. although in this case 'effort' on the part of the learners was also associated with instrumental motivation (subsequent citations) (Ellis, 1994: 512)

When a work has six or more authors, you only need to cite the surname of the first author followed by "et al." for the first and subsequent citations. If one work was written by two or more authors who have the same surname, you must put in parentheses the authors' initials in all text citations.

### *Citing two or more works in the same parentheses*

If two or more works written by the same author are cited in the same parentheses, you are required to give the surname once, then the years of publication from the past to the present. For example,

Empirical studies on student learning carried out since the 70's have found that students' learning outcomes to a great extent depend on their choice of learning approaches (Biggs, 1979, 1987)

The major controversies on L2 learning, as Stern (1973, 1983, 1992) says, center around three key issues.

If two or more works written by different authors are cited in the same parentheses, you need to arrange the citations in alphabetical order by the first author's surname. For example,

One major finding from the earlier studies (Barley, 1969, 1970; Clement et al., 1978; Gardner & Lambert, 1972; Gliksman, 1976; Spolsky, 1969) was

that learners with an integrative motivation tended to obtain better achievement than those with an instrumental motivation.

## APA Reference Style

The basic rule of ordering references in a reference list is to arrange them in alphabetical order. The elements of a reference include the information about the author, the year of publication, the title of the cited materials and the place of publication. A comma is used to separate parts of elements and a period to finish an element. All of you are familiar with the general structure of a reference. Therefore, the following section will only focus on potential difficulties you might come across in producing a reference list.

### *A reference to a journal*

When you write a reference to a journal, the following things need to be paid attention to:

(1) Capitalize the first letter of the title and the subtitle of an article. Don't underline it or add double quotation marks to it.

(2) Capitalize the first letter of major words in the title of a journal and underline a full title.

(3) Use Arabic numbers to specify the volume number and underline it. If there is an issue number, place it in parentheses immediately after the volume number.

(4) Use Arabic number to indicate the inclusive page numbers which immediately follow the volume number, or the issue number if there is any. Remember "pp."is only needed before the page numbers in references to newspapers or magazines but not to a journal.

The following are the examples:

Block, E.L. (1986). The comprehension strategies of second language readers. TESOL Quarterly, 20(3), 463-94.

Boyle, R.P. (1979). Path analysis and ordinal data. American Journal of Sociology, 75, 461-480.

Bracht, R.M. & Glass, G.V. (1968). The external validity of experiments. American Educational Research Journal, 5(4), 437-74.

## *A reference to a book*

A reference to a book is not the same as a reference to a journal although they share many common elements. The following are the differences:

(1) In a reference to an edited book, the editors' names are placed in the author's position followed by parentheses with the abbreviation "Ed", or "Eds" when the book is edited by more than one person.

(2) The first word in a title or subtitle needs to be capitalized and the whole title should be underlined.

(3) Information about the number of edition (e.g. 2nd ed.) is put in parentheses immediately after the title

(4) The place where the publisher is located is indicated before the name of the publisher.

Now look at some examples.

Oxford, R. L. (1990). Language learning strategies: What every teacher should know. New York: Newbury House Publisher.

Marton, F., Hounsell, D. & Entwistle, N. (Eds.). (1984). The experience of learning. Edinburgh: Scottish Academic Press.

Brown, H. D. (1987). Principles of language learning and teaching (2nd ed). New Jersey: Prentice-Hall, Inc.

## *A reference to an article or chapter in an edited book*

Sometimes a reference is not a paper in a journal or an entire book but is an article or a chapter in an edited book. For example, in my study I

referred to an article written by Palmberg entitled "How much English vocabulary do Swedish-speaking primary school pupils know before starting to learn English at school" that is in a book entitled "Foreign language learning and bilingualism" edited by Ringbom in 1985. How is this reference made? The following is this reference:

> Palmberg, R. (1985). How much English vocabulary do Swedish-speaking primary school pupils know before starting to learn English at school. In H. Ringbom (Ed.). Foreign language learning and bilingualism (pp. 89-97). Abo: Research Institute of the Abo Academi Foundation.

When you produce the above reference, you had better pay special attention to the following two things:

1. When an editor's name is not in the author position at the very beginning, the surname is not placed before the initials.
2. Inclusive page numbers are given in parentheses immediately after the book title.

Now look at some more examples:

> Schmeck, R. R. (1983). Learning styles of college students. In R. R. F. Dillon & R. Schmeck (Eds.). Individual differences in cognition (Vol.1, pp. 233-279). New York: Academic Press.

> Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.). Input and second language acquisition (pp. 235-253). Rowley, M.A.: Newbury House.

## Reference Styles in influential international journals

After a close look at three influential journals, i.e. TESOL Quarterly, Studies of Second Language Acquisition and Applied Linguistics. I have noticed that the first two journals adopt the APA style except for two differences. The first difference is that in a reference to a journal, the title of the journal and the number of volume are not underlined as in the APA manual. Instead, they are italicized. The second one is that in a reference to a book, the title of the book is italicized rather than underlined. However, the reference style in Applied Linguistics shows more differences from the APA style. In addition to the two differences mentioned above, they also have the following differences:

1. Except for the first author, the surname is not placed before the initials and furthermore, between the first author and the remaining one(s), "and" is used instead of "&".
2. The date of publication is not in parentheses.
3. The title of a journal article is put within single quotation marks.
4. No comma is between the information about the volume and the name of the journal and a colon is used between the volume number and page number.
5. In a reference to an edited book, when the word editor(s) is abbreviated and put in parentheses, the first letter is not capitalized and a colon is used instead of a comma after the parentheses.

The following are examples from these three types of journals:

1. TESOL Quarterly

Beck, I., & McKeown, M. (1985). Teaching vocabulary: Making the instruction fit the goal. *Educational Perspectives, 23*, 11-15.

Dubin, F., & Olshtain, E. (1993). Predicting word meanings from contextual clues: Evidence from L1 readers. In T, Huckin, M. Haynes, & J. Coady (Eds.). *Second language reading and vocabulary learning* (pp. 181-202). Norwood, NJ: Ablex.

[The above examples are taken from TESOL Quarterly 31(1), Spring 1997 (pp. 140-141)]

2. Studies of Second Language Acquisition

Berg, T. (1986). The problem of language control: Editing, monitoring, and feedback. *Psychological Research, 48*, 133-144.

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken. J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Eds.). *Linguistic disorders and pathologies* (pp. 1-15). Berlin: de Gruyter.

[The above examples are taken from Studies of Second Language Acquisition, 22, 2000 (p. 163)]

3. Applied Linguistics 21/1: 47-77:

Blerwisch, M. And R. Schreuder. 1992. 'From concepts to lexical items.' *Cognition* 42: 23-60.

Ellis, R. 1994. *The study of second language acquisition.* Oxford: Oxford University Press.

Kroll, J. F. And J. Curley. 1988. 'Lexical memory I novice bilinguals: The role of concepts in retrieving second language words' in M. Gruneberg, P. Morris, and R. Sykes (eds.): *Practical Aspects of Memory, Vol. 2.* London: John Wiley and Sons.

[The above examples are taken from Applied Linguistics 21/1: 47-77, 1999.]

It is clear that different journals prefer their own reference style. Therefore, we must follow the style that a specific journal requires when we write a paper. Otherwise, the paper will be unacceptable. For your thesis writing, I suggest you follow the APA style consistently all the way through rather than a mixed one.

## SUMMARY

A thesis should be written in a clear and explicit style that is different from creative writing. It should follow a relatively standardized structure and use formal and succinct language. Theses in the field of applied linguistics are usually written according to the publication manual of the American Psychological Association (APA). In this manual, there are a list of rules that tell you how to use a tense, how to produce a table, how to deal with numbers, capitalization, and spelling, how to cite a work in a thesis, how to quote something from a work, and how to produce a reference list. To present a thesis in a professional manner, you must consistently follow one widely-used style rather than a mixed one.

## DISCUSSION QUESTIONS

Get a copy of Master's thesis from the library of the School of Foreign studies and work with a group to answer the following questions:

1. Did the thesis have a standardized structure? If not, what are the problems?
2. What do you think of the language used in the thesis? Do you think the language is formal? If not, can you give some examples?
3. Are there any problems in the use of tense? Please list them and discuss them.
4. Do you think the tables in the thesis were properly produced?
5. Did the author use hyphens correctly? If not, can you give some suggestions?
6. What do you think of the citations in the thesis? If there are any mistakes, please correct them.
7. Are the quotations properly presented? If not, please revise them.
8. What are the problems in the reference list? Can you give some suggestions for further revision according to the APA style?

# References


Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 11, 256-274.

Allen, E.D. & Valette, R.M. (1977). Classroom techniques: Foreign languages and English as a second language. New York: Harcourt Brace.

Bachman, L. F. (1997). Fundamental considerations in language testing(4th ed). Oxford: Oxford University Press.

Bailey, K. M. (1983). Competitiveness and anxiety in adult second language learning: looking at and through the diary studies. In H. Seliger and M. Long (Eds.), Classroom oriented research in second language acquisition. Rowley, Mass.: Newbury House.

Bejarano, Y. (1987). A cooperative small-group methodology in the language classroom. *TESOL Quarterly*, 21, 483-504.

Bernard, H. R. (1994). Research methods in anthropology: Qualitative and quantitative approaches (Second edition). California: SAGE Publications.

Bialystok, E. (1978). A theoretical model of second language learning. *Language learning*, 28(1), 69-83.

Black, K. (1992). Business statistics: An introductory course. New York: West Publishing Company.

Blaxter, L., Hughes, C. & Tight, M. (1996). How to research. Buckingham: Open University Press.

Block, E.L. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20(3), 463-94.

Boyle, R.P. (1970). Path analysis and ordinal data. *American Journal of Sociology*,75, 461-480.

Bracht, R.M. & Glass, G.V. (1968). The external validity of experiments. *American Educational Research Journal*, 5(4), 437-74.

Brown, J. D. (1988). Understanding research in second language learning. Cambridge: Cambridge University Press.

Chen, J. (1996). A study of discomfort in EFL classroom. Nanjing University: Unpublished BA thesis.

Chen, H. (2000). A study of vocabulary learning strategies used by Chinese children. Nanjing University: unpublished MA thesis.

Cohen, A. (1987). Using verbal reports in research on language learning. In C. Faerch & G. Kasper (Eds.), Introspection in second language research (pp. 82-95). Clevedon, England: Multilingual Matters.

Cohen, L. & Manion, L. (1991). Research methods in education (3rd edition). London: Routledge.

Collins Cobuild English Language Dictionary. (1987). England: HarperCollins Publishers.

Cone, J. D. & Foster, S. L. (1993). Dissertations and theses from start to finish(2nd ed). Washington, DC: American Psychological Association.

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics*, 5(4), 1961-70.

Day, R. A. (1989). How to write and publish a scientific paper. Cambridge: Cambridge University Press.

Dong, Y. (1999). ESL Learning On the Internet — A Case Study on a Joint Online Course. Nanjing University: unpublished MA thesis.

Ellis, R. (1994). The study of second language acquisition. Oxford: Oxford University Press.

Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL Quarterly*, 16, 43-59.

Flower, L., Stein, V.E., Ackerman, J. M, Kantz, M.H., McCormick, K.M., & Peck, N.D. (1990). Reading to write: Exploring a cognitive and social process. New York: Oxford University Press.

Gardner, R. C. (1985). Social psychology and second language learning: The orle of attitude and motivation. London: Edward Arnold.

Gardner, R.C. & Lambert, W. E. (1972). Attitudes and motivation in second-language learning. Rowley, Massachusetts: Newbury House Publishers.

Glaser, B. (1978). Theoretical sensitivity. Mill Valley, CA: Sociology Press.

Gong, X. (1998). A contrastive study of English and Chinese compliment responses. PLA Foreign Languages University: Unpublished MA thesis.

Goodman, Y. M. (1989). Retrospective miscue analysis: History, procedures and prospects. Occasional Paper No. 19. Tucson: Arizona Center for Research and Development, Program in Language and Literacy (pp. 45-68). Cambridge: Cambridge University Press.

Gu, Y. Q. (1997). A study of vocahulary learning strategies by Chinese EFL learners. Hong Kong University: Unpublished Ph. D. thesis.

Guo, C. J. (1997). A proces-oriented cognitive account of L1 influence on L2 writing. Nanjing University: Unpublished MA thesis.

Hatch, E. & Farhady, H. (1982). Researcb design and statistics for applied linguistics. Rowley: Newbury House Publishers, INC.

Hosenfeld, C. (1984). Case studies of nimth grade readers. In J. C. Alderson & A. H. Urquhart (Eds.), Reading in a foreign language (pp. 231-249). London: Longman Group U.K.

Huang, X. H. (1984). An investigation of learning strategies in oral communication that Chinese EFL learners in China employ. Chinese University in Hong Kong: Unpublished M.A. thesis.

Hudelson, S. (1989). A tale of two children: Individual differences in ESL children's writing. In D. M. Johnson & D. H. Roen (Eds.), Richness in writing: Empowering ESL students (pp. 84-99). White Plains, NY: Longman.

Hu, Y. Z. (1999). Error treatment in EFL classroom. Nanjing University: Unpublished MA thesis.

Hymes, D. (1971). On communicative competence. Philadelphia, P.A.: University of Pennsylvania Press.

Johnson, D. M. (1992). Approaches to research in second language learning. London: Longman.

Keeves, J. P. & Sowden, S. (1992). Analyzing qualitative data. In J. P. Keeves (Ed.). Methodlogy and measurement in international educational surveys. The International Association for Evaluation of Educational Achievement (IEA).

Krashen, S. (1985). The input hypothesis. Hong Kong: Longman Group Limited.

Kwo, O..W. Y. (1999). Reflective classroom practice: Case studies of Hong Kong student teachers. In A. Y. Chen & Maanen, J. V. (Eds). The reflective spin: Case studies of teachers in higher education transforming action. Singapore: World Scientific.

Labovitz, S. (1971). The assignment of numbers to rank order categories. *American Sociological Review*, 35, 515-524.

Lado, R. (1961). Language testing. New York: McGraw Hill.

Larsen-Freeman, D. & Long, M. (1991). An introduction to second language acquisition research. New York: Longman.

Lauer, J. M. & Asher, J. W. (1988). Composition research: Empirical designs. Oxford: Oxford University Press.

Longman Dictionary of English Language and Culture. (1992). England: Longman Group UK Limited.

Liu, Z. C. (1999). Learners' attitude towards errors and error correction. Nanjing University: Unpublished MA thesis.

Lu, L. P. (1997). L2 Proficiency and comprehension strategy use. Nanjing University: Unpublished MA thesis.

Ma, G. H. (1998). The relationship of L2 learners' linguistic variables to L2 writing ability of tertiary-level Non-English Majors in China. Nanjing University: Unpublished Ph.D. thesis.

Maccoby, E. E. & Jacklin, C. N. (1974). The psychology of sex differences. Standford: Standford University Press.

Malhotra, N. K. (1993). Marketing research: An applied orientation. New Jersey: Prentice-Hall. Inc.

Markee, N. P. (1994). Toward an ethnomethodological respecification of second-language acqusition studies. In E. E. Tarone et al (Eds). Research methodology in second-language acquisition. New Jersey: Lawrence Erlbaum Associates Publishers.

Marton, F. (1981). Phenomenography-describing conceptions of the world around us. *Instructional Science*, 10, 177-200.

Marton, F. & Saljo, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, 46, 4-11.

Maxim, P. S. (1999). Quantitative research methods in the social sciences. Oxford: Oxford University Press.

Miles, M. & Huberman, A. (1994). Qualitative data analysis (2nd ed). Thousand Oaks (CA), Sage.

Norusis, M.J. (1993). SPSS for windows: Base system user's guide. Chicago: SPSS Inc.

Nunan, D. (1994). Research methods in language learning (3rd edition). Cambridge: Cambridge University Press.

Nyikos. M. (1990). Sex-related differences in adult language learning: Socialization and memory factors. *The Modern Language Journal*, 74, 273-87.n

Oxford, R. & Crookall, D. (1989). Research on language learning strategies: Methods, findings and instructional issues. *The Modern Language Journal*, 73(4), 404-419.

Patton, M.Q. (1980). Qualitative evaluation methods. Beverly Hills, CA: Sage.

Pica, T, Young, R., & Doughty, C. 1987. The impact of interaction on comprehension. *TESOL Quarterly*, 21(4), 737-758.

Patton, M. Q. (1990). Qualitative evaluation and research methods. Newbury Park, CA: Sage.

Punch, K. F. (1998). Introduction to social research: Quantitative & qualitative approaches. London: SAGE.

Publication Manual of the American Psychological Association. (1992).

Qin, X. Q. (1998). An investigation into the internal structure of EFL motivation at the tertiary level in China. Nanjing University: Unpublished Ph.D. thesis.

Raimes. A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, 19(2), 229-258.

Rivers, W. M. (1983). Communicating naturally in a second language: Theory and practice in language teaching. New York: Cambridge University Press.

Rosenberg, M. (1979). Conceiving the self. New York: Basic Books.

Rudestam, K. E. & Newton, R.R. (1992). Surviving your dissertation: A comprebensive guide to content and process. California: SAGE Publications.

Schmidt, R. & Frota, S. (1985). Developing basic conversational abiltiy in a second language: a case study of an adult learner of Portuguese. In R. Day (Ed.), Talking to learn. Rowley, Mass.: Newbury House.

Seliger, H. W. & Shohamy, E. (1990). Second language research methods. Oxford: Oxford University Press.

Skehan, P. (1989). Individual differences in second language learning. London: Edward Arnold.

Smith. M. J. (1998). Social science in question. London: SAGE.

Stern, H. H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review*, 31, 304-18.

Stern, H.H. (1983). Fundamental concepts of language teaching. Oxford: Oxford University Press.

Stern, H. H. 1992. Issues and options in language teaching. Oxford: Oxford University Press.

Strauss, A. (1987). Qualitative analysis for social scientists. New York: Cambridge University Press.

Strauss, A. & Corbin, J. (1998). Basics of qualitative research. Newbury Park, CA: Sage Publications.

Su, X. J. (1996). A study of the changes of tertiary English majors' beliefs. Nanjing University: Unpublished MA thesis.

Tashakkori, A. & Teddlie, C. (1998). Mixed methodology. London: SAGE.

Theodorson, G. A. & Theodorson, A.G. (1969). A modern dictionary of sociology. New York: Thomas Y. Crowell.

Stake, R. (1988). Case study methods in educational research: Seeking sweet water. In R. M. Jaeger (Ed.), Complementary methods for research in education. Washington: American Educational Research Association.

Strauss, A. & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage.

Vygotsky, L.S. (1986). Thought and language. Massachusetts: The MIT Press.

Wang, L. (1999). Risk-taking and English learning: a study of the risk-taking beliefs and behaviors of English majors in China. Nanjing University: Unpublished MA thesis.

Wang, W. Y. (2000). L1 thinking and L2 writing: A study of the process of L2 by Chinese university students. Nanjing University: Unpublished Ph.D. thesis.

Wen, Q. F. (1993). Advanced level English language learning in China: The relationship of modifiable learner variables to leaning outcomes. Hong Kong University: Unpublished Ph. D. thesis.

Wen, Q. F. (1999). Testing and teaching spoken English. Shanghai: Shanghai Foreign Language Education Press.

Wen, Q. F., Wu, C. X., & So, L. (1999). An evaluation of English Majors' spoken English-Band 4. *Foreign Language Teaching and Research*, 1, 29-34.

Wen, Q. F. & Guo, C. J. 1998. The relationship between thinking in L1 and L2 writing ability: A study of the process of English picture composition. *Linguistics and Applied linguistics*, 4, 27-48.

Wen, Q. F. & Wang, H.X. (1996). The relation between learner factors and the scores on CET-Band 4. *Foreign Language Teaching and Research*, 4, 33-39.

Wen, Q. F. & Wu, X. S. (1998). On teaching spoken English in a language laboratory. *Foreign Language World*, 1, 30-33.

Wu, C. X. (1998). A study of second-year English Majors' pragmatic competence. Nanjing University: Unpublished MA thesis.

Wu, J. (1998). A study of university students' argumentative writing in English: rhetorical knowledge and discourse pattern. Nanjing University: Unpublished MA thesis.

Xu, H. M. (2001). A study of the use of meta-discourse markers by Chinese EFL learners in L2 writing: a developmental perspective. Nanjing University: Unpublished Ph.D. thesis.

Xu, Y. C. (1998). Cohesion in EFL writing. Nanjing University: Unpublished MA thesis.

Ye, X. L. (1994). The role of schema and language proficiency in EFL reading. Nanjing University: Unpublished MA thesis.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, 17(1), 165-187.

Zamel, V. 1987. Recent research on writing pedagogy. *TESOL Quarterly*, 21(4), 697-715.

Zhu, L. Z. (1998). Influence of different tasks, lengths of planning time and lengths of learning on second language oral performance. Nanjing University: Unpublished MA thesis.

# Appendixes

## APPENDIX ONE: A SUREVEY STUDY

(Note: The following is a longitudinal survey in which the same group of students were asked to respond to the same questionnaire three times in an attempt to detect developmental patterns in motivation, beliefs and strategies.)

### Developmental Patterns in Motivation, Beliefs and Strategies of English Learners in China

Qiufang Wen Nanjing University

Abstract:

The paper reports a study on developmental patterns in the modifiable learner variables (i.e. motivation, beliefs and strategies) and in their relations based on the logitudinal questionnaire data. The subjects involved in the study were 72 English majors who were admitted into Nanjing University in 1996.

Key words:

English language learners, modifiable variables, motivation, beliefs, strategies

# 英语学习者动机、观念、策略的变化规律与特点

南京大学 文秋芳

提要： 本文运用定量研究的方法，三次跟踪调查了南京大学1996年入学的72名英语专业学生可控因素（动机、观念和策略）及其关系的变化情况。

关键词：英语学习者、可控因素变化、动机、观念、策略

纵观动机、观念和策略的研究历史，尽管研究的广度和深度在不断发展，但几乎所有的研究都是静态研究（例如:Cohen，1998；Ellis，1994；Gardner & Lambert,1972；Huang & Van Naerssen，1985；Oxford，1990；Wen，1993；王文宇，1998；文秋芳，1995；文秋芳和王海啸，1996a，1996b），从动态角度研究的则寥寥无几。笔者曾于1996年就三个语言学习策略的变化进行过尝试性的研究，但研究没有涉及其他的学习者可控因素，更没有涉及各可控因素关系之间的变化。与以前的研究不同，本文所要报告的九五社科项目“英语学习者动机、观念、策略的变化规律”是从动态的角度，考察几乎所有学习者可控因素的变化情况[19]。具体要回答的问题包括：

1. 通过大学两年的学习，大学英语专业的学生在可控因素（动机、观念和策略）上会发生那些变化？变化的特点和趋势是什么？
2. 通过大学两年的学习，各类可控因素之间的关系是否发生了变化？如有变化，变化的特点和趋势是什么？

## 一、 可控因素的定义及关系

### *1．动机、观念和策略的定义*

可控因素分为动机、观念和策略（Wen，1993；文秋芳和王海啸，1996）。动机分为表层动机和深层动机（Biggs，1979)。表层动机通常与个

[19] 由于版面的限制，这里只能登载该项目的部分研究结果。有兴趣者可阅读该项目的结项报告（见文秋芳等，2000）。

人的前途直接相关，动力来自外部；深层动机一般不与学习者个人的前途和经济利益发生直接的联系，学习动力来自对英语语言或文化本身的兴趣。

观念指学生在学习英语过程中通过自身的体验或别人的影响所形成的一种看法体系。英语学习者的观念大致分为两类：一类为管理观念，另一类为语言学习观念。管理观念指学生对确定目标、制定计划、选择策略、调控策略等一系列管理活动重要性的认识。语言学习观念是指学生对如何才能掌握好语言知识、语言技能和交际能力的主张。语言学习观念[20]包括：形式操练观念(即认为掌握语言知识对学好外语非常重要)；功能操练观念(即认为在交际中运用语言知识对学好外语非常重要)；依赖母语观念（即认为通过翻译来学习第二语言是捷径）。

策略指学习者为有效学习所采取的措施，这种策略可以是外部活动，也可能是内部活动。策略分为管理策略与语言学习策略。前者与语言学习过程的管理有关，后者与语言学习材料发生直接的关系。管理策略涉及目标的制定、策略的选择、时间的安排、策略有效性的评估和调整，以及情感的控制与调整。与语言学习观念相对应，语言学习策略包括：形式操练策略、功能操练策略和依赖母语策略。

### 2. *动机、观念和策略之间的关系*

在学习者调控系统内部，动机影响观念，观念影响策略的选择。在观念与策略之间，管理观念影响管理策略的使用和选择；语言学习观念影响语言策略的使用和选择。管理观念又影响语言学习观念；管理策略又影响语言学习策略的成效。从这个意义上说，与管理学习过程有关的观念和策略在高一个层面上，与学习语言知识和技能本身直接有关的观念和策略在低一个层面上。

## 二、研究设计

### *1. 研究对象*

参加该项研究的学生共有72名。他们于1996年9月考进南京大学外国语学院的英语专业，其中女生57名，男生15名，平均年龄为18.7岁。

---

[20] 语言学习观念与语言学习策略分类的理论依据是Stern （1983）提出的三大理论争端，详细内容请阅读“Fundamental concepts of language teaching”。

### 2. 研究工具

“学习情况调查问卷”是笔者根据撰写博士论文时所设计的学习者因素问卷(Wen，1993)修改而成，并将其翻译成中文。问卷由四部分组成：(1)个人简况；(2)英语学习动机；(3)管理/语言学习观念；(4)管理/语言学习策略。除第一部分外，其余都是选择题。与学习动机相关的题目有六个，其中三个属于深层动机，三个属于表层动机，学生需要从五个选项中挑选一个答案（即从“这个原因对我不重要”到“这个原因对我极其重要”）。有关观念的问卷题共有22个，涉及管理观念、形式操练观念、功能操练观念和依赖母语观念。学生可挑选的答案也有5个，即从“我坚决不同意这个看法”到“我非常同意这个看法”。有关策略的问卷题共有35道题，涉及管理策略、形式操练策略、功能管理策略和依赖母语策略。每个问卷题后仍旧提供五个选项，即从“这个句子完全或几乎完全不适合我的情况”到“这个句子完全或几乎完全适合我的情况”。

### 3. 数据收集和分析

笔者请一、二年级担任综合英语课的老师和三年级担任阅读课的老师分别在1996年、1997年和1998年的开学初，让学生在课堂上填写了问卷。问卷的数据输入电脑后，运用社会科学的统计软件(SPSS)对数据进行了分析。分析分四步进行：(1)用描述统计列出各项变量的频率，对学生未作回答的个别题目用该题目的平均数代替；(2)用内部一致性的方法检验各可控因素归类是否能达到统计上的要求（见附录一），其中有12道题因未达到统计要求而被删除；(3)用相关分析的方法，检查变量之间的关系；(4)用$t$-tests检验学生在大学两年中英语学习动机、观念和策略变化的情况。

## 三、研究结果与讨论

### 1. 动机的变化

这一部分将报告学生的英语学习动机（深层动机和表层动机）在三次（刚进大学时，进大学一年后，进大学两年后）调查中，每次的偏爱倾向是什么？这种偏向是否达到统计意义？然后报告纵向比较的结果，即对动机的三次调查结果是否有变化？如有变化，变化的特点和趋势是什么？

### 1） 偏爱倾向

根据表1和表2, 1996年学生对两种学习原因的偏爱程度完全一样，而1997年和1998年，学生的深层动机明显强于表层动机，并且差异具有统计意义：1997年差异程度达到了.02的显著水平, 1998年达到了.000的水平。

| 变量名称 | 深层动机 | 表层动机 |
|---|---|---|
| 1996年(平均数/方差) | 3.34/.68 | 3.34/.70 |
| 1997年(平均数/方差) | 3.63/.63 | 3.41/.62 |
| 1998年(平均数/方差) | 3.83/.68 | 3.44/.73 |

表1：两个动机变量的描述统计结果

| | Mean/SD | | T-value | P |
|---|---|---|---|---|
| 1996年 | 深层动机 | 3.34/.68 | -.04 | .972 |
| | 表层动机 | 3.34/.70 | | |
| 1997年 | 深层动机 | 3.63/.63 | 2.37 | .020 |
| | 表层动机 | 3.41/.62 | | |
| 1998年 | 深层动机 | 3.83/.68 | 3.80 | .000 |
| | 表层动机 | 3.44/.73 | | |

表2：*t*-tests：学生的两种学习动机的变化情况

### 2） 变化的趋势和特点

从纵向比较来看,学生的深层动机在逐年加强, 且每两次之间的变化都具有统计意义(见表3)。深层动机问卷题的平均数1996年是3.34，1997年增加到3.63，1998年增加到 3.84。根据表4, 表层动机也在逐年加强，1996年表层动机问卷题的平均数为3.34, 1997年增加到3.41，1998年增加到 3.44。与深层动机不同的是，表层动机每两次之间的变化都不具有统计意义。

| 深层动机 | Mean/SD | 相关系数 | P | T-value | P |
|---|---|---|---|---|---|
| 第一次与第二次 | 3.34/.68 | .65 | .000 | -4.57 | .000 |
| 比较 (1996年/1997年) | 3.63/.63 | | | | |
| 第二次与第三次 | 3.63/.63 | .56 | .000 | -2.66 | .010 |
| 比较 (1997年/1998年) | 3.83/.68 | | | | |

表3：*t*-tests 的结果：1996-1998年深层动机变化的情况

| 表层动机 | Mean/SD | 相关系数 | P | T-value | P |
|---|---|---|---|---|---|
| 第一次与第二次 | 3.34/.70 | .50 | .000 | -.89 | .376 |
| 比较 (1996年/1997年) | 3.41/.62 | | | | |
| 第二次与第三次 | 3.41/.62 | .41 | .000 | -.27 | .791 |
| 比较 (1997年/1998年) | 3.44/.73 | | | | |

表4：*t*-tests：1996-1998年表层动机变化的情况

## 2．*观念的变化*

### 1）偏爱倾向

从1996年到1998年，在四种不同的观念中，管理观念的平均数最高，依赖母语观念的平均数最低，功能操练观念的平均数位于第二，形式操练观念的平均数位于第三(见表5)。虽然功能观念的平均数高于形式操练的平均数，但只有1998年的差异具有统计意义。

| 变量名称 | 管理观念 | 形式操练观念 | 功能操练观念 | 依赖母语观念 |
|---|---|---|---|---|
| 1996年(平均数/方差) | 4.25/.39 | 3.89/.40 | 3.91/.36 | 2.66/.55 |
| 1997年(平均数/方差) | 4.24/.44 | 3.88/.40 | 3.94/.40 | 2.70/.43 |
| 1998年(平均数/方差) | 4.30/.45 | 3.81/.46 | 3.91/.41 | 2.73/.49 |

表5：不同观念的平均数和标准差

(注：这里每种观念的平均数=每个问卷题的平均数相加后再除以题目数。)

### 2）变化的趋势和特点

从纵向的角度来比较，四种观念的平均数虽有或多或少的变化，但它们都不具有统计意义。这进一步证明了Wenden（1991)提出的观念稳定性，即观念一旦形成，便和其他知识一样储存在长期记忆中，成为知识体系的一部分，不容易改变。或者说，学生在进入大学时，调查中所涉及到的四类观念已经趋向于稳定。

## *3．策略的变化*

### 1）偏爱倾向

横向比较1996年到1998年的各类策略平均数（见表6)。除1998年形式操练策略和功能操练策略之间没有显著性差异以外，其他各类策略之间的差异都具有统计意义。在四项策略中，管理策略的平均数最高，母语策略最低，形式操练策略和功能操练策略的平均数处于中间，这一总体趋势与观念相同。所不同的是，功能操练观念的平均数每次都高于形式操练观念的平均数（1996年，.002；1997年，.006；1998年，.10)；而功能操练策略的平均数在1996年和1997年显著低于形式操练策略（1996年，-.42；1997年，-.24)，只有到了1998年情况才发生了变化，功能操练策略的平均数稍许高出了形式操练策略(1998年，.02)，但差异不具有统计意义。

由此，我们可以得出结论：学生最偏爱管理策略，最不喜欢母语策略；在刚进大学和经过一年学习之后，喜欢形式操练策略的程度明显强于喜欢功能操练策略的程度；但经过第二年的学习以后，学生对形式操练和功能操练两类策略的使用不再有明显的偏爱。许多西方的研究者们把形式操练和功能操练两类策略看成是互为对立的一对矛盾，认为只有使用后一种策略，学习者的外语交际能力才能得到发展。而我校学生在两年内使用这两类策略总趋势的变化似乎表明它们在实际学习的过程中并不互相对立。

| 变量名称 | 管理策略 | 形式操练策略 | 功能操练策略 | 母语策略 |
|---|---|---|---|---|
| 1996年(平均数/方差) | 3.55/.48 | 3.39/.45 | 2.97/.52 | 2.66/.77 |
| 1997年(平均数/方差) | 3.52/.47 | 3.40/.43 | 3.16/.49 | 2.57/.60 |
| 1998年(平均数/方差) | 3.61/.52 | 3.29/.48 | 3.31/.51 | 2.43/.62 |

表6：1996-1998年四种策略的平均数及方差

（注：这里每种策略的平均数=每个问卷题的平均数相加后再除以题目数）

**2）变化的趋势与特点**

如表7所示，1996年到1998年，管理策略的平均数虽有或多或少的变化，但都不具有统计意义。换句话说，学生的管理策略没有受到大学两年生活的影响。或者说，学生的管理策略在高中毕业时已趋向成熟。

| 管理策略 | Mean/SD | 相关系数 | P | T-value | P |
|---|---|---|---|---|---|
| 第一次与第二次 | 3.55/.48 | .57 | .000 | .70 | .486 |
| 比较 (1996年/1997年) | 3.52/.47 | | | | |
| 第二次与第三次 | 3.52/.47 | .56 | .000 | -.1.54 | .128 |
| 比较 (1997年/1998年) | 3.61/.52 | | | | |

表7：*t*-tests：1996-1998 管理策略的变化情况

根据表8，形式操练策略的变化总趋势是先升后降，第一次变化，即上升，没有统计意义，但第二次变化，即平均数降低了.11，达到了.028显著性水平。这似乎表明形式操练策略变化的关键期在二年级，而不在一年级。这一总体趋势与笔者1996年跟踪调查的结果相同，所不同的是，在1996年的跟踪调查中，两次变化都没有统计意义。

| 形式操练策略 | Mean/SD | 相关系数 | P | T-value | P |
|---|---|---|---|---|---|
| 第一次与第二次 | 3.39/.45 | .35 | .003 | -.14 | .889 |
| 比较 (1996年/1997年) | 3.40/.43 | | | | |
| 第二次与第三次 | 3.40/.43 | .63 | .000 | 2.25 | .028 |
| 比较 (1997年/1998年) | 3.29/.48 | | | | |

表8：*t*-tests：1996-1998形式操练策略的变化情况

与形式操练策略不同，功能操练策略的平均数呈直线上升趋势，每次变化都具有统计意义（见表9）。这表明了学生通过大学两年的英语学习，使对功能操练策略的使用有了急剧的增加。这一结果并不出乎意料。英语专业

的教学方法和要求都与中学有明显的不同，学生必须要具备听说读写四方面的交际能力。为达到这一要求，学生一定要增加运用语言进行真实交际的活动。这一结果与笔者1996年的跟踪调查结果基本相同，所不同的是，本项目两次变化都具有统计意义，而1996年的结果只有第二次变化才具有统计意义。

| 功能操练策略 | Mean/SD | 相关系数 | P | T-value | P |
|---|---|---|---|---|---|
| 第一次与第二次 | 2.97/.52 | .40 | .001 | -.2.99 | .004 |
| 比较 (1996年/1997年) | 3.16/.49 | | | | |
| 第二次与第三次 | 3.16/.49 | .65 | .000 | -2.89 | .005 |
| 比较 (1997年/1998年) | 3.31/.51 | | | | |

表9：*t*-tests：1996-1998年功能操练策略的变化情况

与功能操练策略相反，依赖母语策略的平均数呈直线下降的趋势，且每两次的变化都具有统计意义（见表10）。这一结果比较符合逻辑。对母语的依赖程度往往是检验外语水平的标尺。随着外语水平的提高，学生对母语的依赖程度应该逐步降低。以上结果与1996年的跟踪调查结果有相似之处。1996年的结果是：先降后微升，只有第一次变化具有统计意义。

| 依赖母语策略 | Mean/SD | 相关系数 | P | T-value | P |
|---|---|---|---|---|---|
| 第一次与第二次 | 2.66/.77 | .48 | .000 | -.2.99 | .004 |
| 比较 (1996年/1997年) | 2.57/.60 | | | | |
| 第二次与第三次 | 2.57/.60 | .51 | .000 | -2.89 | .005 |
| 比较 (1997年/1998年) | 2.43/.62 | | | | |

表10：*t*-tests：1996-1998年母语策略的变化情况

## *4. 三类可控因素之间的关系及其变化*

### 1）动机与观念和策略

表11显示表层动机与四种观念之间的相关系数均没有统计意义，但与管理观念的正相关有明显增强的趋势。深层动机与管理观念和形式操练观念

分别呈正相关，其中97年深层动机与管理观念的相关系数.30 达到.009的显著水平，96年深层动机与形式操练观念的相关系数.25 达到.032的显著水平；与依赖母语观念呈负相关，且有增强的趋势，其中1997年和1998 年的相关系数(-.24 和 -.27 ) 均有统计意义，显著性水平分别为.05 和.02；与功能操练观念没有联系。

表层动机与管理策略和功能操练策略呈正相关，其中1997年，1998年表层动机与管理策略的相关系数为.29 和.47，分别达到.015 和 .000的显著性水平，1997年表层动机与功能操练策略的相关系数为.26，达到.029的显著性水平；表层动机与其它策略没有统计意义上的联系。深层动机与依赖母语策略的关系最为紧密，1996年，1997年和1998年三次数据均显示它们之间的关系是负相关（-.41， -.31，-.38），且每次相关的程度都达到.001以上的显著性水平。深层动机与管理策略的关系密切程度仅次于依赖母语策略。1997年与1998年它们之间的相关系数为.46 和.34, 显著性水平分别为.000和.004。深层动机与功能操练策略的关系也呈正相关，1997年和1998年两次的相关系数（.26 和.23）分别达到.026 和.049的显著性水平。在四类策略中，唯一与深层动机没有关系的策略就是形式操练策略。

| 表层动机 | 管理观念/策略 | 形式操练观念/策 | 功能操练观念/策略 | 依赖母语观念/策 |
|---|---|---|---|---|
| 1996年相关系数 | .01/.16 | .10/.15 | .17/.09 | .08/.09 |
| 1997年相关系数 | .11/.29* | .19/.17 | .05/-.12 | .01/.08 |
| 1998年相关系数 | .21/.47*** | .18/.08 | .11/.26* | .10/-.03 |
| 深层动机 | 管理观念/策略 | 形式操练观念/策 | 功能操练观念/策略 | 依赖母语观念/策 |
| 1996年相关系数 | .22/.13 | * .25/-.05 | .04/.15 | -.13/-.41*** |
| 1997年相关系数 | ** .30/.46*** | .07/.21 | .01/.26* | *-.24/-.31*** |
| 1998年相关系数 | .21/.34** | .14/.08 | .10/.23* | *-.27/-.38*** |

表11：动机与观念和策略 的关系

(注：*<.05, **<.01, ***<.001)

与动机和观念的关系相比，动机和策略的关系更为密切。前两者之间有统计意义的相关系数只有四个，后两者有统计意义的相关系数有10 个。而在这14个有统计意义的相关系数中，深层动机占有11个，而表层动机只占一个。这意味着深层动机对学习者的观念与策略影响大于表层动机。

在各类观念和策略中，管理观念/策略与动机有统计意义的相关系数占有五个，依赖母语观念/策略与动机的也占有五个，其余三个在功能策略中，一个在形式操练观念中。这意味着动机对管理观念/策略，对依赖母语观念/策略的影响比较大，其次是功能策略，最弱的是形式操练观念；对功能操练观念与形式操练策略几乎没有影响。

在这些相关系数中，特别值得讨论的有三点。第一点是深层动机与依赖母语观念/策略的关系。在它们所有的六对关系中，都呈负相关，其中五对具有统计意义。这说明它们之间的关系方向相反，并且比较稳定。这一结果在以前类似研究中从未发现过，但比较符合逻辑。具有深层动机的学生对学习外语感兴趣，对外国文化感兴趣，对外国的科学技术感兴趣。深层动机愈强，使用目标语言的欲望愈强，因此依赖母语的可能性就越小。

第二个值得讨论的是两类动机与管理策略的关系。在六对关系中（见表11），所有都呈正相关，其中后两年（即1997年和1998年）四对关系具有统计意义。这一结果意味着学生在刚进大学时，动机对管理策略的影响还不明显，但通过大学第一年的学习以后，动机与管理策略之间的关系有了显著的加强。无论是表层动机，还是深层动机，只要想学英语，就都会对管理策略产生积极的影响。

第三个值得讨论的是动机与功能操练策略的关系。六对关系中，五对正相关，其中三对具有统计意义。1996年，当学生刚进入大学时，两种动机与功能操练策略的关系均不具有统计意义。1997年，当学生进入大学一年后，深层动机与功能操练策略的相关系数为.26，达到了.026的显著性水平。到了1998年，即第三次调查时，表层动机和深层动机与功能操练策略的相关系数分别为.26和.23，均达到.05的显著性水平。这似乎意味着通过大学两年的学习，动机与功能操练策略的关系在逐步建立。

从总体上看，相对于其它策略而言，动机与形式操练策略关系最不密切。出现这一结果的原因可能是，与其它几类策略相比，学生对形式操练策略最为熟悉，实践最多，同学之间的差异最不显著。在各类观念中，两类动机与功能操练观念关系最不密切， 其次是形式操练观念。出现这一结果并

不令人奇怪。观念只是学习者头脑中的看法，而不是学习者自身的行为。看法对行为虽有一定的影响，但两者之间一定会有差距，特别是当观念的实现需要学习者克服困难时，这种差距更为明显。在这种情况下，比较普遍的现象是：想得好，做得差。

**2）观念与策略**

| 变量名称 | 管理 | P | 形式操练 | P | 功能操练 | p | 依赖母语 | P |
|---|---|---|---|---|---|---|---|---|
| 1996年相关系数 | .27 | .02 | .12 | .28 | .20 | .09 | .38 | .00 |
| 1997年相关系数 | .36 | .00 | .29 | .01 | .17 | .16 | .40 | .00 |
| 1998年相关系数 | .56 | .00 | .39 | .00 | .43 | .00 | .48 | .00 |

表12：观念与策略之间的关系

从1996年到1998年，管理观念/策略、形式操练观念/策略、功能操练观念/策略与依赖母语观念/策略都呈现正相关（见表12）。在12对相关系数中，除1996年形式操练观念/策略，1996年与1997年功能操练观念/策略三对以外，其它各对都达到了显著性水平。从总体上看，四对观念/策略的关系都有逐年增强的趋势。这意味着学生的某种观念越强，使用相对应的策略可能性越大。出现这一结果的主要原因有两个。第一个原因是心理因素。随着学生的年龄增加，他们逐年成熟起来，观念与策略之间的差距也应该随之缩小。第二个原因与英语水平有关系，即随着学生英语水平的提高，原来比较难以实现的观念现在逐渐变得容易付诸实践。

在这四对观念与策略的关系中，其中依赖母语观念与依赖母语策略的关系最为稳定，1996年，1997年和1998年三次的相关系数分别为.38,.40和.48，差异最小。经过大学第一年的学习后，形式操练观念与形式操练策略之间的关系发生了最为剧烈的变化，1996年与1997年的相关系数分别为.12 和.29，相差.17；其次是管理策略与管理观念之间的关系，1996年与1997年的相关系数分别为.27 和 .36，相差.09；其余两对观念/策略相关系数的差别只有.03和.02。经过大学两年学习后，除形式操练观念/策略以外，三对观念/策略之间的关系发生的变化都远远大于第一年，1996年与1997年的差异分别是：管理观念/策略.09，功能操练观念/策略.03，依赖母语观念/策略.02；而1997年与1998年的差异分别是：.20，.26，.08 。这似乎说明观念/策略之间关系密切程度的变化，关键期在大学学习的第二年。

# 四、结论

## 1. 主要结果

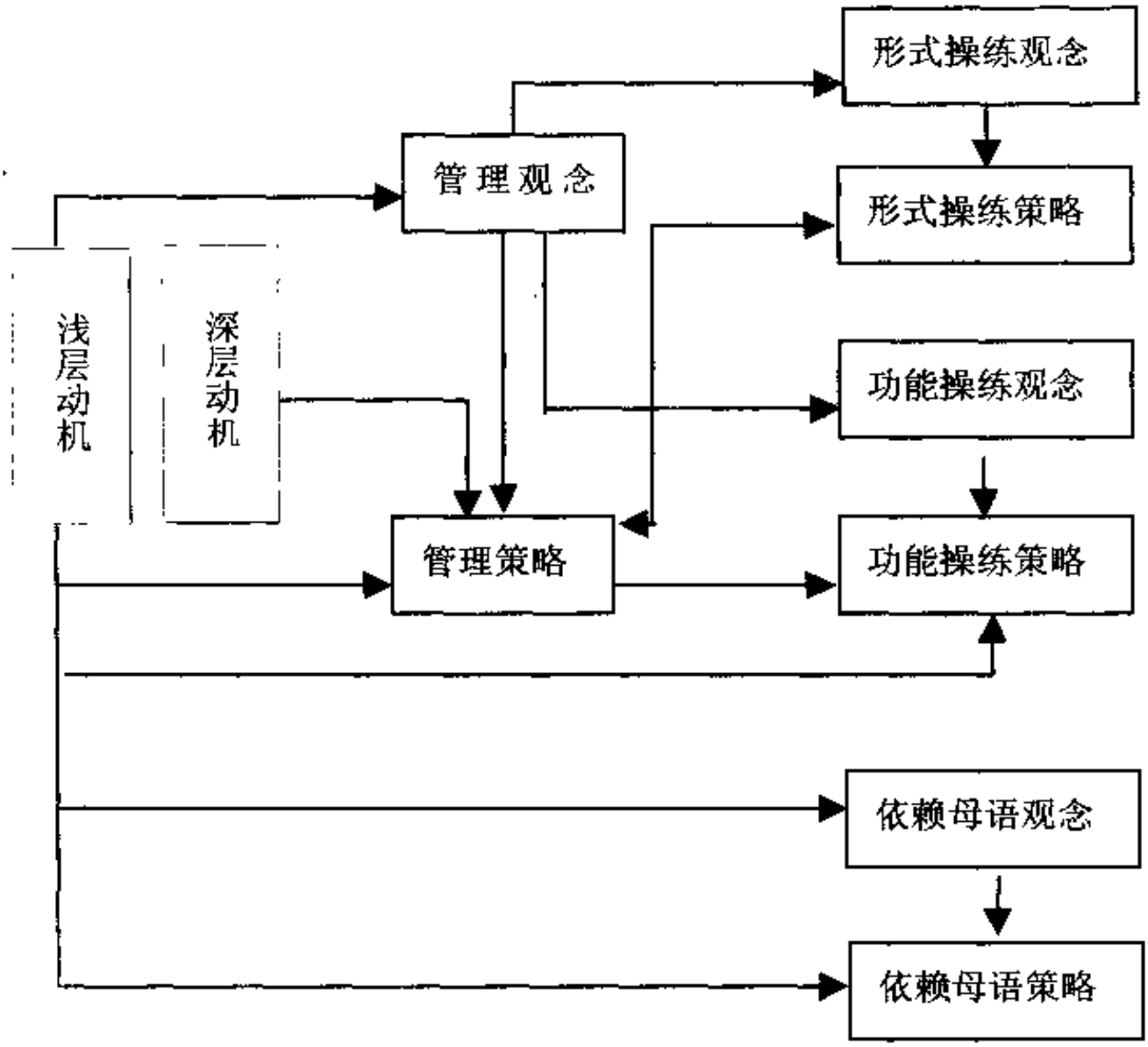
现将该项目的主要研究结果总结如下：

1) 学生在刚进大学时，对表层和深层动机没有明显的偏爱倾向，经过大学第一年学习后，深层动机的强度超过了表层动机；经过大学第二年学习后，深层动机成为明显强势动力源。尽管两种动机的强度都是逐年增强，但深层动机强度的增加幅度远远超过表层动机。两种动机之间的关系由负相关转为了正相关。在四种观念中，学生最赞成管理观念，最不赞成依赖母语观念，对形式操练观念和功能操练观念的赞成程度位于两者之间。这一倾向经过两年的学习没有变化。在四种策略中，学生使用管理策略最多，使用依赖母语策略最少，对形式操练策略和功能操练策略的使用程度位于两者之间。这一总体倾向经过两年的学习没有变化，但就形式操练策略和功能操练策略而言，学生的偏爱倾向有明显变化。1996年刚进大学时，学生使用形式操练策略的程度显著高于使用功能操练策略；1997年时，虽然学生使用功能操练策略的程度明显增加，但仍旧低于形式操练策略的使用程度；1998年时，使用功能操练策略的程度继续增加，而使用形式操练策略的程度反而有所下降，因此使用功能操练策略的程度高于使用形式操练策略的程度。

2) 经过大学两年学习，深层动机和功能操练策略的使用程度呈逐年上升趋势，且具有统计意义；依赖母语策略的使用程度呈逐年下降趋势，且具有统计意义；形式操练策略的使用程度先微升后下降，微升不具有统计意义，但下降具有统计意义；所有四种观念以及管理策略均无显著变化。深层动机、功能操练策略和依赖母语策略的变化属于渐进式，即经过大学第一年和大学第二年的学习，它们都发生了具有统计意义的变化；形式操练策略的变化属于突变式，即第一年没有发生具有统计意义的变化，到了第二年才发生这样的变化。

3) 根据二次跟踪调查的数据，尽管代表各类可控因素的变量可呈上升或下降趋势，各类因素之间关系的密切程度可增强或减弱，但各类因素之间关系的趋向比较稳定（见图1）。深层动机与管理观念/策略、功能操练策略，依赖母语观念/策略有联系；表层动机与管理策略有联系；四个观念与相应的四个策略有联系；管理观念与形式操练观念以及功能操练观念有联系；管理策略与形式操练策略以及功能操练策略有联系。值得指出的是：管理观念/策略、形式操练观念/策略和功能操练观念/策略与依赖母语观念/策略没有联系。唯一与母语观念/策略有联系的是深层动机。

图1：本项目中各变量之间的关系



4) 动机与观念以及策略的关系密切程度的变化没有明显的规律可循，但策略与观念的关系有逐年加强的趋势，其关系密切程度的变化关键期主要在大学学习的第二年。

以上各项结果需要在今后的研究中进一步验证。

## 2. 该项研究结果的意义

本项目的研究结果表明动机、观念、策略之间的关系具有较高的稳定性。动机影响观念和策略；观念影响策略；管理观念/策略影响形式操练策略/功能操练策略，但不影响依赖母语观念/策略。因此在教学中，教师如果要改变学生的语言学习行为，必须依据这些因素之间的关系，从总体上考虑，研究调控措施，避免简单从事。

深层动机、功能操练策略以及依赖母语策略在大学第一年和第二年以后都发生了显著性的变化；形式操练策略的变化主要发生在第二年以后。这一结果说明英语专业的第一年和第二年是学生可控因素发生变化的关键期。如果在这个阶段能够配备得力的教师，对学生因势利导，一定能收到事半功倍的效果。

## 3. 对未来研究的建议

### *1）该项目的不足之处*

该项目的不足之处包括以下几个方面：

1. 被调查的学生人数少，且限于某一个学校，因此代表性不够广泛。

2. 定性调查由于设计不够周密，因此数据不能用于进一步分析。

3. 同一问卷调查了三次，学生可能产生厌烦情绪，影响数据的可靠性。

鉴于以上不足，在应用本项目的研究结果时需要格外谨慎。

### *2）未来研究可控因素变化的几种可能性*

像这样动态跟踪研究可控因素变化的项目，在该领域内尚属首次。我们认为可以选用相关的样本来重复该项研究，检验本次研究的结果。此外，

该项目还需要从以下几个方面继续深入研究：1、运用定性的方法跟踪研究少数学生的可控因素变化情况，当然由于定性研究费时费力，事先一定要精心策划，收集数据时，一定要仔细认真；2、在本次研究的基础上着重深入研究依赖母语观念/策略的变化与深层动机的关系；3、深入研究某些可控因素的变化与教学过程之间的关系。

## 参考文献


Biggs, J. B. (1979). Individual differences in study processes and the quality of learning outcomes. Higher Education, 8, 381-94

Cohen, A. (1998). Strategies in learning and using a second language. London: Longman.

Ellis, R. (1994). The study of second language acquisition. Oxford: Oxford University Press.

Gardner, R. & Lambert, W. (1972). Attitudes and motivation in second-language learning. Rowley, Mass: Newbury House.

Huang, X. & M. Van Naerssen. (1985). Learning strategies for oral communication. Applied Linguistics, 8(3), 287-307.

Oxford. R. (1990). Language learning strategies: What every teacher should know. Rowley, Mass.: Newbury House.

Stern, H. (1983). Fundamental concepts of language teaching. Oxford: Oxford University Press.

Wen, Qiufang. (1993). Advanced level English learning in China: The relationship of modifiable learner variables to learning outcomes. Hong Kong University: Unpublished doctoral thesis in

Wen, Qiufang & Johnson, R. K. (1997). L2 learner variables and English achievement: A study of tertiary-level English majors in China. Applied Linguistics, 18 (1), 27-48

Wenden, A. (1991). Learner strategies for learner autonomy. New York: Prentice Hall.

王文宇，1998年，观念、策略与词汇记忆，《外语教学与研究》第一期。
文秋芳，1995年，英语成功者与不成功者在学习方法上的差异，《外语教学与研究》第三期。
文秋芳，1996年，大学生英语学习策略变化的趋势及其特点，《外语与外语教学》第四期。
文秋芳、王海啸，1996a，大学生英语学习观念与策略的分析，《解放军外语学院学报》第四期。
文秋芳、王海啸，1996b，学习者因素与大学英语四级考试成绩的关系，《外语教学与研究》第四期。
文秋芳、乐眉云、丁言仁、王文宇，2000，九五社科项目结项报告：英语学习者动机、观念、策略的变化规律与特点。

# 附录

| 动机名称 | 问卷题内容 | 题目数 | Alpha (1996年) | Alpha (1997年) | Alpha (1998年) |
|---|---|---|---|---|---|
| 深层动机 | 1. 我喜欢学习外国语<br>2. 我对说英语国家的文化感兴趣<br>3. 我要了解外国的科学技术、政治和经济情况 | 3 | .33 | .42 | .62 |
| 表层动机 | 1. 我要找一份好工作<br>2 我要报考研究生<br>3. 我要出国深造 | 3 | .41 | .47 | .54 |

表1：两类动机的问卷题内容、题目数量和 Alpha 值

| 观念名称 | 问卷题内容 | 题目数 | Alpha 1996年 | Alpha 1997年 | Alpha 1998年 |
|---|---|---|---|---|---|
| 管理观念 | 1. 很好地计划自己的学习时间是学好英语的重要保证<br>2. 选择有效的学习方法对学好英语很重要<br>3. 有明确的长期和短期学习目标对学好英语很重要<br>4 不断总结自己的进步并找出存在的问题对学好英语很重要<br>5 学好英语的一个重要方面是了解自己的个性特点，发挥长处，克服短处<br>6. 经常反思自己的学习方法是否有效，对学习外语非常重要 | 6 | .52 | .74 | .76 |
| 形式操练观念 | 1. 记外语句型，对学习外语非常重要<br>2. 记单词对学好外语很重要<br>3. 精读课文对学习外语很重要<br>4. 背诵好的课文对学习外语很重要<br>5. 反复朗读课文对学好外语很重要<br>6 反复模仿好的录音带对练好语音语调很重要 | 6 | .46 | .62 | .70 |
| 功能操练观念 | 1. 进行大量的听力练习对学好外语很重要<br>2. 要学好外语，大量阅读外文报纸、杂志、小说等很重要<br>3. 猜单词和句子的意思是解决阅读困难的最好方法<br>4. 进行大量的口语训练，对学好外语很重要<br>5. 进行大量的写作练习，对学习外语很重要<br>6. 听英语时，碰到没有学过的单词，最好的方法是猜测 | 6 | .42 | .62 | .60 |
| 母语观念 | 1. 理解课文最好的方法是翻译<br>2. 说英语时，最好用中文想好要说的内容<br>3. 要想将所听的英文内容记下来，最好的方法是记中文意思<br>4. 要想写出好的英语作文，最好的方法是先用中文组织好想写的内容 | 4 | .59 | .69 | .79 |

表2：四种观念的问卷题内容、题目数量和 Alpha 值

| 策略名称 | 问卷题内容 | 题目数 | Alpha (1996年) | Alpha (1997年) | Alpha (1998年) |
|---|---|---|---|---|---|
| 管理策略 | 1. 除了老师布置的作业外，我有自己的学习计划。<br>2 为了使自己有足够的时间学习英语，我很好地安排自己的学习日程。<br>3. 我对改进自己的英语学习有明确的要求。<br>4. 我评价自己学习英语进步的情况，从而找出薄弱环节和改进措施。<br>5 我评价自己的学习策略，从而找出存在的问题和解决的方法。<br>6. 我根据学习任务的特点，选择不同的学习策略。<br>7. 我选择适合自己英语水平的材料来学习。<br>8. 我研究自己的个性特点，找出哪些特点有利于自己的英语学习，哪些特点阻碍自己的进步，从而能发挥自己的优势，采取相应的措施，克服弱点。 | 8 | .74 | .78 | .84 |
| 形式操练策略 | 1. 当阅读课文时，我争取弄懂课文里的每一处。<br>2. 课外，我反复朗读课文。<br>3 当我不懂句子的意思时，我分析句子的语法结构。<br>4. 我背诵好的课文。<br>5. 当听英语材料时，我争取听懂每一句话的意思。<br>6. 假如在听英语时碰到生字，我会尽量记住生字的发音，<br>然后根据发音在字典上查找它的拼写和意思。 | 6 | .49 | .61 | .62 |
| 功能操练策略 | 1. 我在课外主动阅读英文报纸、杂志和小说。<br>2. 我课外主动听英语广播。<br>3. 在英语课上，我尽量主动地争取回答问题。<br>4. 我课外主动看英语电视和电影。<br>5. 我课外尽量用英语与同学和老师会话。<br>6. 我课外自己对自己说英语。<br>7. 我主动地用英语记笔记、留言、写信或日记。 | 7 | .60 | .71 | .73 |
| 母语策略 | 1. 为了帮助对课文的理解，我把课文译成中文。<br>2. 我说英语时，首先用中文组织意思，再翻译成英语<br>3. 我用英文写作文时，先用中文组织意思，再译成英语。<br>4. 听英语时，我用中文记住所听的意思。 | 4 | .82 | .72 | .79 |

表3：四种策略的问卷题内容、题目数量和 Alpha 值

# APPENDIX TWO: AN EXPERIMENTAL STUDY

# 输入方式与听力词汇习得——一项听力词汇习得实验的报告

南京农业大学　王艳

**提要**：听力词汇少是造成学生听力理解困难的重要原因。本论文报告了不同输入方式对听力词汇习得影响的实验。结果发现，声音输入组在生词音的辨识和意义的理解上都强于字形输入组，即听入比读入更有利于听力词汇的习得。结果还发现，阅读词汇通过听力练习可以转化为听力词汇。

**关键词**：听力词汇、阅读词汇、词汇输入方式

## 一、引言

笔者在多年听力教学中，通常听到学生反映生词是他们听力理解的主要困难，尤其是大学低年级学生。1999年笔者曾以"英语听力理解困难"为题对大学一年级学生进行了调查。在学生自我报告的困难中，词汇困难占第一位。刘思（1995）在研究中也发现学生的听力词汇量明显低于阅读词汇量，由此可以推断提高听力词汇量是提高听力理解能力的关键之一。

Nation （1990）认为词汇习得的方式是影响词汇学习的重要因素之一，即单词的初始输入是通过听觉渠道还是通过视觉渠道。国外在这方面的实证研究不多。Channell（1988）认为视觉渠道不是学习

生词的最好输入方式。Feitelson, Goldstein, Iraqi, and Share （1993）发现，通过听觉渠道来输入生词不仅易于学会单词，而且能帮助学习语法。国内对词汇输入方式与单词习得关系的研究几乎是空白。本文报告的实验只是对这一问题研究的初步尝试。笔者希望通过实验回答两个问题：（1）听入是否比读入更有利于听力词汇的习得？和（2）阅读词汇可否转化为听力词汇？

## 二、研究设计

### *1. 受试对象*

该研究的受试对象为南京农业大学人文学院英语专业2000年进校的一年级全体学生。他们来自全国15个省份中的40多个不同城市和3个直辖市，共56人， 其中11名男生，45名女生，平均年龄为18.9岁，高考英语平均成绩为112分。

这些学生在进校时随机地分成两个班，每班均为28人。为确认两班学生的英语听力水平相当，实验前进行了一次听力测试，内容为三篇短文，十道选择题，一题一分。测试内容选自外语教学与研究出版社1996年出版的《初级英语听力》。表一为测试结果。

**表1 实验前测试结果**

| | 平均数 | 标准差 |
|---|---|---|
| 1班(*N*=28)=实验组 | 7.4286 | 1.451 |
| 2班(*N*=28)=对比组 | 7.5357 | 1.503 |

经*t*-test 检验，两班成绩没有显著性差异。1班作为实验组，采用听入的方式学习生词（以下称为听入组）；2班作为对比组，采用读入的方式学习生词（以下称为读入组）。

## 2. 实验设计与步骤

实验组与对比组按不同的输入方式学习10个生词。其中，5个为目标生词，即这些词将会在测试的短文中出现；5个为干扰词，从字典上随机选出（见附录1）。10个词混放在一起，受试者需要判断这些词是否要在所听短文中出现。实验分两个阶段进行：（1）词汇学习；（2）测试。

两组受试者在第一阶段采用不同的词汇学习方式：

1. 听入组：通过听的方式学习给定的10个单词，即教师将每个单词念三遍，每次的朗读速度和中间的停顿都保持一致，并用汉语给出中文意思一遍。平均每个词需要9秒钟时间，总时间约为1.5 分钟。10个单词给出的次序事先随机决定。

2. 读入组：采用阅读的方式学习10个生词。教师将单词及其中文意思写在胶片上，用投影仪投映在银幕上。10个单词给出的次序与听入组相同，总时间约为1.5 分钟。

第二阶段为测试，共进行两次。对两组采用的测试方式完全相同。第一次测试时，先播放一篇短文录音（见附录2），然后教师将10个单词依次念出，让受试者判断哪些词在这篇录音材料出现。第一次测试结果交上来之后，进行第二次测试。第二次测试内容分为两个部分。第一部分和第一次测试内容相同，第二部分为新增加的听力理解题，即受试者根据对短文的理解从四个选项中选出一个正确答案，问题和选项均由教师念出。第一部分测试内容的目的是比较两组受试者在掌握目标词的音上是否有差异，第二部分内容旨在通过学生对短文理解的程度来推测两组学生对生词意义的掌握是否有差异。考虑到受试者是一年级学生，其整体听力水平比较低，因此听力理解题安排在第二次测试中。这样，两组受试者在听两遍短文的基础上再做选择题，可降低因随意猜测答案而带来误差的可能性。短文共157个词，掌握5个目标词的意义对短文理解很关键。表2列出了实验步骤和所需时间。

表2 实验步骤、目的和所需时间

| 次序 | 步骤 | | 目的 | 时间（分钟） |
|---|---|---|---|---|
| | 实验组（听入） | 控制组 （读入） | | |
| 1 | 实验前测试 | 实验前测试 | | 20 |
| 2 | 学习生词（听入） | 学习生词（读入） | 显示不同的输入形式 | 1.5 |
| 3 | 实验后第一次测试(听短文,做词汇测试) | 实验后第一次测试(听短文,做词汇测试) | 测试学生通过听短文来辨认已学过单词的能力 | 5.0 |
| 4 | 实验后第二次测试(听相同短文,做词汇与听力理解测试) | 实验后第二次测试(听相同短文,做词汇与听力理解测试) | 1．测试增加听力训练的次数是否提高辨认单词的准确性<br>2．用听力理解题推测学生对生词词义的理解情况。 | 5.0 |

### *3. 数据收集*

两组受试者在听完一遍短文之后，在纸上写下1到10，以代表单词的顺序。教师将10个单词读出来，让受试者做出判断。如果他们认为某个词刚才在短文中出现过，就在相应的题号后打√，如果不是，就打× 。第二次测试中，受试者以同样的方式,做一遍词汇测试，然后做短文的听力理解题。受试者根据自己的理解，选出认为正确的选项。

### *4. 数据分析*

在分析词汇测试的结果时，受试者每正确判断一个词得1分，即对确实从短文中挑选出来的目标词打√的给一分，和对干扰词打×的也给1分。凡是错的均不减分。满分为10分。听力理解选择题选对的得1分，选错的不减分，全对得3分。所有数据输入计算机，用SPSS(社会科学统计软件)进行处理。笔者计算了每个目标单词的平均分和10个单词的总平均分,以及理解题的平均分（见表三），然后将两组的词汇成绩和听力理解成绩进行了独立样本T检验， 最后将听入组和读入组的第一次词汇测试成绩与第二次的分别进行配对样本T检验。

**表3 测试结果**

| | 第一次测试 | | 第二次测试 | |
|---|---|---|---|---|
| 项目 | 平均数 | 标准差 | 平均数 | 标准差 |
| 听入组（*N*=28） | | | | |
| 目标词汇 | | | | |
| *Rhythm* | 0.54 | 0.51 | 0.54 | 0.51 |
| *Crawl* | 0.96 | 0.19 | 0.96 | 0.19 |
| *Spur* | 0.71 | 0.46 | 0.61 | 0.50 |
| *Dismay* | 0.68 | 0.48 | 0.68 | 0.48 |
| *Stroke* | 0.57 | 0.50 | 0.61 | 0.50 |
| 10个目标词总分 | 6.64 | 1.39 | 6.82 | 1.90 |
| 段落理解题得分 | | | 2.60 | 0.57 |
| 读入组（*N*=28） | | | | |
| 目标词汇 | | | | |
| *Rhythm* | 0.50 | 0.51 | 0.39 | 0.50 |
| *Crawl* | 0.68 | 0.48 | 0.82 | 0.39 |
| *Spur* | 0.54 | 0.51 | 0.68 | 0.48 |
| *Dismay* | 0.64 | 0.49 | 0.75 | 0.44 |
| *Stroke* | 0.61 | 0.50 | 0.82 | 0.39 |
| 10个目标词总分 | 5.64 | 1.70 | 6.50 | 1.64 |
| 段落理解题得分 | | | 2.25 | 0.65 |

## 三、结果与讨论

### *1）听入是否比读入更有利于听力词汇的习得？*

词汇的音和义是掌握听力词汇最基本的两个方面。根据表4，在第一次测试中，听入组的词汇平均分（6.64）明显高于读入组(5.64)。经独立样本T检验，两组的得分具有显著性差异 ($t$=2.40，$p$<.05)。

**表4 两组在第一次测试中的总分比较**

| 组别 | 平均数 | 标准差 | p |
|---|---|---|---|
| 听入组(*N*=28) | 6.64 | 1.39 | *.02* |
| 读入组(*N*=28 ) | 5.64 | 1.70 | |

这说明，第一次测试中听入组在单词音的记忆和辨别上强于读入组。可以推断，采用听入的方法学习词汇，能够较好地识别单词的发音。这一结果符合心理学的理论。两种学习方法接受的外界刺激不同，因此在大脑中形成的感觉记忆表象也不同。在外界刺激上，一个是声音，一个是字形。根据记忆的一般原理，人脑对信息的存储是以感觉痕迹形式被记录下来的。通过听觉接受的输入，记录的就是是声音表象；通过视觉接受的输入，记忆的就是图象。心理学家认为声像比图像的延续时间长（朱纯，1994：188）。当原始刺激再度呈现时，受试者需要再认知，即将当前呈现的刺激与头脑中记忆表象相匹配。很显然，当受试者听含有目标词汇的短文录音材料时，听入组的记忆痕迹与听到的目标词更加接近，因此识别任务的完成比读入组要容易得多。由此可以推断，仅仅通过读入方式学习，不能有效地记忆单词的发音，在以后再听到该词时，也不能很好识别。

从表3中可以发现，五个目标单词中，“crawl”这个词的两组得分差距特别大，其差异具有统计上的显著意义（$t$=2.95，$p$<.01）（见表5）。

**表5 第一次测试中两个组对目标词汇“crawl”的得分情况和比较**

| 词汇 | 听入组(*N*=28) | | 读入组(*N*=28) | | p |
|---|---|---|---|---|---|
| | 平均数 | 标准差 | 平均数 | 标准差 | |
| *Crawl* | 0.96 | 0.19 | 0.68 | 0.48 | *.006* |

短文中，“crawl”一词出现的频率为三次，而其他词只出现一次。虽然两个组第一次测试中都只听一遍短文，词汇测试的形式也一样，但是这个词听入组得了0.96分，明显高于其他词；而读入组只得了0.68分，与其他单词的得分相当。这说明目标词出现频率对听入组的影响比对读入组的要大。听入组因依赖其听觉去记忆和辨别单词，听力材料中单词的重复频率对其产生了作用，出现次数越多，越容易识别，因此绝大部分人都判断正确了（全班都答对为1分）。然而，读入组依赖视觉来记忆单词，因此听力材料中单词的重复频率并没有对其产生明显作用。这也证明用读入方法学习词汇不如听入更有利于对词音的辨识。

另外，听入组在对干扰词的判断上，也强于读入组。在五个干扰词中，听入组误选的平均为1.8个，而读入组为平均2.3个。

表6列出了两组在听力理解成绩上的差异。听入组的得分明显高于读入组，且达到统计上的显著意义（$t$=2.20, $p$<.05）。

**表6 两组听力理解测试的得分情况**

| 组别 | 平均数 | 标准差 | p 值 |
|---|---|---|---|
| 听入组（*N*=28） | 2.60 | .567 | |
| 读入组（*N*=28） | 2.25 | .645 | *.032* |

由于5个目标词均为理解短文的关键词，因此我们可以推断，听力理解题得分高的听入组就意味着对目标词词义理解也比读入组好。换句话说，听入组不仅仅是在识别发音上占有优势，在习得词义上也好于读入组。这一结果说明掌握听力词汇不仅可以识别含有生词的语言信息（语音），还有助于对词汇、句子，甚至段落的意义理解。因而，综合前一点的结果，相比较而言，采用听入的方式更有利于习得听力词汇。

### *2）阅读词汇可否转化为听力词汇？*

表7显示两个组第二次测试的结果已趋于接近。这一次，听入组得分的平均值虽然仍高于读入组，但是两组间得分已不再具有显著性差异。

**表7 两组在第二次试后测试的得分比较**

| 组别 | 平均数 | 标准差 | p 值 |
|---|---|---|---|
| 听入组(*N*=28) | 6.82 | 1.90 | |
| 阅读组(*N*=28 ) | 6.50 | 1.64 | *n. s.* |

为了进一步了解第二次听短文对受试者的影响，笔者进行了听配对检测（paired-samples *t*-test）结果显示，第二次听短文对读入组受试者的影响颇为明显，（$t=-2.25$, $p<.05$），而对听入组受试者的影响并不显著（见表8）。

**表8 第二次听短文段落对两个组的影响**

| 组别 | 第一次听短文 | | 第二次听短文 | | p 值 |
|---|---|---|---|---|---|
| | 平均数 | 标准差 | 平均数 | 标准差 | |
| 听入组 (*N*=28 ) | 6.64 | 1.4 | 6.82 | 1.9 | *n. s.* |
| 读入组 (*N*=28 ) | 5.64 | 1.7 | 6.50 | 1.6 | *.033* |

这说明，再接受一次相同输入，对听力组听力词汇习得成绩的提高没有明显作用；但对读入组有显著作用。从这一结果可以推断，只要增加听力训练，阅读词汇可以转化为听力词汇。

## 四、结论

本次实验得到了以下结果：

1．听入比读入更有利于听力词汇的习得。

2．通过训练，阅读词汇可以转化为听力词汇。

事实上，在词汇学习中，言语听觉和视觉并不是孤立存在的，它们和人的其他感觉器官一起发挥作用。问题在于，听觉和视觉分别是听力和阅读的感性基础，在训练不同的英语技能时应有所侧重。长期以来，我们一直重阅读、轻听说，造成听力词汇量远远不足。真正习得一个单词意味着对其书面及口头形式的理解和使用。而观察学生平时的词汇学习不难发现，他们通常的学习方式与读入组非常相似。他们熟谙传统的读背方式，将单词表中的生词朗读出来，记住拼写和中文意思，却常常忽略了发音。下一次看到这个词时，一时难以识别，误以为是生词。如果我们强调用听觉记忆来记住单词，培养学生先记住音和义、再记住拼写的习惯，就会有助于扩大听力词汇量，提高听力理解能力。如果同时配合视觉和动觉等多种输入方式，听力词汇量和阅读词汇量之间的差距就能缩小。

由于该项实验仅是探索性研究，且受试者人数少，测试只涉及10个目标词汇和一篇短文，实验时间也比较短，因此以上结果有待进一步验证。

## 引用文献

Channell, Joanna (1988). Psycholinguistic considerations in the study of l2 vocabulary acquisition. In R. Carter & M. McCarthy (Eds.). Vocabulary and Language Teaching ( 83-96). London: Longman.

Feitelson, D., Z. Goldstein, J.Iraqi, and D. Share (1993). Effects of listening to story reading on aspects of literacy acquisition in a digloss situation. *Reading research Quarterly,* 28, 70-9

Nation, I.S.P. (1990). Teaching and Learning Vocabulary. New York: Newbury House/Harper Row

Pienemann, M., M. Johnston, and G. Brindley. (1988). Constructing an acquisition-based procedure to assessing second language acquisition. Studies in second language acquisition.10, 271-43.
刘思，1995，英语听力词汇量与阅读词汇量——词汇研究调查报告[J]，《外语教学与研究》第1期，61-65页。
朱纯，1994，《外语教育心理学》[M]，上海：上海外语教育出版社。

**附录1：用于实验的10个生词**

*rhythm 节奏，韵律； *crawl 自由式游泳；wrinkle 皱纹； display 陈列，展示； *stride 大步走，跨过； *spur 刺激，鼓舞； yacht 游艇，轻舟；dismay 沮丧，灰心； *stroke …击， 一划； combination 结合，联合
（注：有星号的为目标词）

**附录2：用于词汇和听力理解测试的短文**

Being able to swim has always been a valued skill. But competitive swimming is just over one hundred years old. Racing in the water began in Europe only about fifty years before the first modern Olympics. England was the leader in this new sport. Pools were built and races were held.

Interest was spurred when a group of native Americans went to England in 1844 to swim against an English team. To the surprise and dismay of the English, the Americans beat them easily. At the time the English used the breast stroke, which was thought a good form. The Americans used a kind of free style that was much faster.

A short time later an English family, who had moved to Australia, developed the crawl, an overhand swimming method. One of that same families also took it to American. Charles Daniels, an early user of the crawl, changed the kicking rhythm to develop the American crawl.

Listening comprehension questions:

1. When did racing in water begin in Europe?

   A) About fifty years before the first modern Olympics.

   B) Fifty years ago.

   C) When the crawl was developed.

   D) After the breast stroke was developed.

2. What did a group of Native American swimmers do in 1984?

   A) They learned the English way of swimming.

   B) They beat an English team easily in a competition.

   C) They introduced breast stroke to England.

   D) They failed in the competition.

3. Who developed the crawl?

   A) An Australian family

   B) An American family.

   C) An English family who moved to Australia.

   D) An australian family who moved to England.

# APPENDIX THREE: A CASE STUDY

(Note: The following is a multiple-case study in which six senior middle school students were asked to write a composition based on the pictures given while speaking out what was going on in their minds in order to find out how L1 was involved in the process of L2 writing and what are the specific functions.)

## Thinking in L1 and L2 writing quality:

## A study of the process of English picture composition

Qiufang Wen, Nanjing University
Chunjie Guo, Nanjing University of Aeronautics and Astronautics

The paper reports on a study of the process of English picture composition in an attempt to reveal the relationship between thinking in L1 and L2 writing ability. The subjects involved in the study were 6 senior students from a secondary school in China. They were asked to write an English composition in about 100 words based on the pictures given, while speaking out whatever occurred in their minds hy the inethod of thinking-aloud. The writing task was undertaken individually and the whole process of thinking-aloud was recorded. The data-analysis included two tasks: (1) rating the compositions; (2) transcribing the tapes and analyzing the protocols. Two native speakers were invited to score English compositions independently and the internal reliability between the two raters reached .83. The protocols were analyzed both quantitatively and qualitatively.

The results showed that the students with higher scores on composition used far less L1 than the students with lower scores in terms of the total amount. However, considering the five functions of L1 use

identified from the data (i.e. transformation, confirmation, generating ideas, retrieving L2 forms and controlling the writing procedures), the students with higher scores employed more L1 for transformation and confirmation, but less L1 for the remaining three functions. Furthermore, the mean length of each L1 utterance used by the higher achievers was longer than that used by the lower achievers in the case of transformation and confirmation, but shorter than the lower achievers in the other cases. Due to the small sample size, the results yielded from this study will only serve as hypotheses to be tested in the future.

## 母语思维与外语写作能力的关系：对高中生英语看图作文过程的研究

南京大学　　文秋芳
南京航空航天大学　郭纯洁

本文报告了笔者运用有声思维(think-aloud)的方法纪录了6名高中生英文看图作文的过程，其目的是为了揭示母语思维与外语写作能力之间的关系。结果表明母语思维的参与量与写作成绩之间呈负相关。在英语写作过程中，母语的主要功能为转换中介、内容生成中介、形式检索中介、内容验证中介和程序管理中介。虽然高分组在总体上比低分组少用母语思维，但该组比低分组多使用中文作为转换中介和内容验证中介,少用中文作为内容生成中介、形式检索中介和程序管理中介。就使用中文的平均长度而言，高分组用于转换中介和内容验证中介的中文平均长度比低分组长，但用于内容生成中介和形式检索中介的中文平均长度比低分组短。由于样本人数少，以上结果只是为大样本的研究提供假设。

关键词：有声思维 母语功能 外语写作 二语习得

本文报告了笔者运用有声思维(think-aloud)的方法纪录了高中生英文看图作文的过程。该项研究的目的在于揭示母语思维与外语写作能力之间的关系。具体地说，该项研究首先研究作文得分与母语思维中使用中文词数的关系。第二步，考察母语在外语写作过程中的不同功能以及这些不同功能的动态交替。最后根据所建立的母语功能类别，比较高分组和低分组的差异。由于样本人数少，所得结果只是为今后的大样本研究提供假设。

## 一、研究背景

用外语写作文，特别在外语学习的初级阶段，以母语为中介似乎无法避免，然而大家又都认为，用母语思维会干扰外语学习的进步(Johnson, 1985; Wen & Johnson, 1997; Zamel, 1982)。过去的研究着眼于作文中错误对比分析，以此推断母语的迁移或干扰作用。然而只研究输出结果，不探究输出过程，难以揭示母语参与外语写作的全部作用，更难界定母语参与思维在什么条件下产生促进作用，在什么条件下产生阻碍作用。

从80年代起对写作过程的研究日益增多(Arndt, 1987; Brooks, 1985; Diaz, 1986; Jacobs, 1982; Urzua, 1987; Zamel, 1983)，其中大多数研究都试图检验第二语言的写作过程是否与第一语言的写作过程相同，只有少数研究涉及到母语对外语写作影响这一课题，其研究结果似乎都认为在外语写作过程中借助母语思维对提高外语作文的质量有积极的影响。例如, Lay (1982)拍摄了4个以中文为母语的外语学习者有声思维的写作过程，并就写作背景与态度与被试者进行了面谈。她研究的结果是：在写同一篇作文的过程中，运用母语思维多的学生比运用母语思维少的学生，在作文的内容，结构和细节上都表现出明显的优势。 Zamel(1982)发现在她的8名研究对象中，唯一通过翻译来写作文的，英文写作水平最高，而其余的不仅不用翻译，而且认为翻译对英文写作没有帮助。Cumming (1987)在他的研究报告中指出：6个以法语为母语的加拿大成年人在用英文写作文时，写作水平低的只是通过法语生成英文作文的内容，而写作水平高的不仅用法语生成内容，检查文体，而且更多的是用于选择英文

单词。根据他的看法，英文作文水平高的比水平低的对母语的依赖性更强。Friedlander（1990）研究了母语在外语作文内容的构思中所起的作用。他发现运用母语思维既不影响外语写作的时间，也不影响质量。Kobayashi 和 Rinnert (1992) 比较了日本学生通过翻译写的英文作文和直接用英文写的作文，发现语言水平低的学生与水平高的学生相比，明显得益于翻译。

国内对外语写作过程中母语作用的研究几乎是一片空白，但最近郭纯洁、刘芳（1997）的“外语写作过程中母语影响的动态研究”颇具开拓性。郭和刘通过有声思维的方法研究了12名不同英语水平的学生看图作文的过程。根据对数据的定性分析，他们指出母语在外语输出过程中的主要作用有三个：(1)对作文内容的逻辑推理；(2)对语言形式的分析、判断；(3)对相关外语词汇、短语或句子的检索。他们还指出逻辑推理的过程以母语思维为主。以上结果在过去的相关研究中均未涉及，这无疑对我们全面认识母语在外语写作过程中的作用有着重要的贡献，但该项研究只限于对母语动态影响的一般描述，对母语作用的界定过于简单，不能用于解释所有使用中文的情况。另一方面，该研究还未涉及母语参与量的多少与写作成绩的关系。本文报告的研究就是对郭、刘收集的部分数据作更深入细致的分析，以比较高分组与低分组在外语写作过程中使用母语的差异。

## 二、研究设计

### 1. 研究对象

用于最后数据分析的六名学生来自安徽省临泉一中，其中高二和高三的学生各一半。他们选自20名被邀参加英文看图作文的高中生。挑选标准有两个：(1)学生在面谈中对研究所用的“有声思维”方法态度积极，并认为这种方法基本上反映了他们写作时的思维活动；(2)他们在有声思维的过程中沉默时间不超过总写作时间的20%。

### 2. 测量工具

用于英文看图作文的8幅连环画选自当地出版的教师参考书(见附录)。连环画描写的是一个学生在上学的路上如何帮助盲人过马路的

故事。为了纪录学生有声思维的过程和所用的总时间以及沉默的时间，我们准备了一台高质量的录音机和计时器。

3. 数据收集

数据收集分二个阶段进行：(1)“有声思维”方法训练；(2)正式测试。

为了使被试者明确有声思维的要求和获得有声思维的亲身体验，他们接受了必要的训练。首先由研究者进行有声思维示范，他一边读古文，一边说出自己理解古文的思维过程，并在示范基础上进行了必要的解释。接着，对学生试测。试测时，每个学生需用有声思维的方法完成两项任务：一是读一段古文，二是辨认一个不完整的图画或汉字。这两项试测任务经过精心挑选，均不涉及两种语言的交替使用，目的是希望被试者不要误认为在英文看图作文时必须要中英文交替使用。

正式测试的要求包括：(1)根据所给的图画，用英文写出一篇100字左右的文章。(2)必须自始自终说出写作过程中头脑中出现的任何想法。想法是以汉语出现，就说汉语；想法以英文出现，就说英文。说的过程中，不允许有较长的停顿和沉默。(3) 要边说边写。研究者对每个学生看图作文时的有声思维情况不仅认真地做了观察纪录，而且进行了录音。

4. 数据分析

数据分析分涉及两项任务：(1)对作文进行评分；(2)逐字纪录有声思维的情况，并对其进行分析。我们聘请了在南京大学任教的两位外籍教师，对被试者的六篇作文分别进行了等级评定。他们根据对文章的内容、结构和语言三方面的总体印象，将六篇作文的质量由高到低派出名次等级，6级为质量最高, 1级为质量最低。两位教师所给成绩相关系数为 .83 (p =.04)。每个学生作文的最后等级根据二位教师所给等级平均数的排序。表一列出了每个学生的作文得分情况：

表一：6名学生英文作文得分情况

| | 教师A | 教师B | 平均等级 | 最后等级 |
|---|---|---|---|---|
| 学生1 | 2 | 1 | 1.5 | 1 |
| 学生2 | 5 | 5 | 5.0 | 5 |
| 学生3 | 3 | 2 | 2.5 | 3 |
| 学生4 | 6 | 6 | 6.0 | 6 |
| 学生5 | 1 | 3 | 2.0 | 2 |
| 学生6 | 4 | 4 | 4.0 | 4 |

对有声思维录音数据的分析分三步进行。首先，我们统计每个学生在看图作文的思维过程中使用中文词数占总词数的比例，然后计算该比例与写作成绩的关系。中文词的确定根据商务印书馆1996年出版的《现代汉语词典》修订本。统计中英文总词数时，没有排除重复出现的词。

第二步，我们从每个学生有声思维的纪录中分离出一系列使用中文的事件，每一个事件必须与某个功能相联系。每个中文事件功能的界定以郭和刘（1997）提出的三大功能为基本依据，但在此基础上作了较为详细的补充和归类，目的是为了能描述有声思维中出现的所有使用中文的事件。建立类别的过程比较费时耗力，原因是类别的建立需要反复修订和验证。第三步，根据建立的母语功能类别，比较高分组和低分组的差异，以此推断在外语写作过程中用母语所产生积极和消极作用的条件。

三、结果与讨论

该项研究的结果主要用于回答下列三个问题。第一个问题是母语参与思维量的多少与外语写作能力之间的关系。根据以前的研究结果，我们的假设是：作文得分高的学生比得分低的学生多依赖于母语思维；或者说多用母语思维的学生比少用的学生作文得高分的可能性更大。第二个问题是母语在外语作文的过程中究竟有哪些功能？这些不同的功能是如何在英语句子的写作过程中交替发挥作用的？第三个问题是作文得高分者和得低分者在使用母语上的差异如何体现在所建立的功能类别上。

1．中文词的数量与作文成绩的关系

根据表二，高分组学生使用中文词数的平均比例(37%)大大低于低分组的学生(51%)。统计结果表明，中文词数所占总词数的比例与作文成绩之间的相关系数为 -.77（p=.07）。虽然相关系数比较高，但由于样本太小，不具有统计意义。这里负相关的含义是：写作思维过程中使用中文词数比例越高，学生作文得低分的可能性越大。由此我们可以提出假设：写作过程中多依赖母语为中介影响作文的质量；或者说英文作文水平低的学生不得不多用中文思维。究竟使用中文思维是导致作文水平高低的原因还是结果，相关分析不能回答这个问题。这里我们只能推断：写作思维过程中汉语使用比例的高低可以在很大程度上预测作文的得分。

上述结果刚好与Lay（1982），Zamel(1982)，Cummings(1987) 和 Friedlander（1990）的相反。造成差异的原因可能是多方面的。其中重要的原因是，他们的结果都不是来自精确的量化数据。仅凭被试者自己的报告和研究者的印象，所得结论很可能不准确。当然这里不能排斥另一种可能性，即上述提到的研究，包括本文报告的研究在内，样本人数都很少，所得结果缺少可靠性。

表二: 高分组和低分组使用中文词数比例的比较

| | | 中文总词数 | 英文总词数 | 中文词数占中英文总词数的比例 | 作文等级排序 |
|---|---|---|---|---|---|
| 高分组 | 学生4 | 448 | 842 | 35% | 6 |
| | 学生2 | 596 | 769 | 43% | 5 |
| | 学生6 | 247 | 580 | 30% | 4 |
| | *平均数* | | | *37%* | |
| 低分组 | 学生3 | 567 | 936 | 38% | 3 |
| | 学生5 | 587 | 524 | 53% | 2 |
| | 学生1 | 958 | 542 | 64% | 1 |
| | *平均数* | | | *51%* | |

2．母语在外语看图作文过程中的功能

根据对有声思维数据的分析，母语在某一语言单位形成过程中，主要有5个功能：(1)转换中介；(2)内容生成中介；(3)形式检索中介；(4)内容验证中介和(5)程序管理中介（见表三）。

表三：母语功能的分类及其定义

| 母语功能的名称 | 定义 |
|---|---|
| 1. 转换中介 | 与后面紧跟着的英文有直接联系的中文，是作者转换英文的依据。 |
| 2. 内容生成中介 | 在形成内容过程中所使用的母语，用于对内容的推理、判断、选择、监控、联想和提问。 |
| 3. 形式检索中介 | 在形式检索过程中所使用的母语，用于对形式的判断、推理、监控、评价、联想和提问。 |
| 4. 内容验证中介 | 把英文倒译成的母语，用于检查所用的英文形式是否表达了自己想要表达的内容。 |
| 5. 程序管理中介 | 在决定写作程序时所用的母语。 |

下面我们首先从静态的角度，分别举例说明母语5种不同的功能；然后以动态的方式说明在外语句子/从句形成的过程中，这5种不同功能如何交替发挥作用的。

1）对母语功能的静态描述

(1) 转换中介

转换中介指与英文作文内容有直接联系的中文内容，是作者转换成外语语言形式的依据。判断转换中介的标准只有一个，即中文和与之相对应的英文内容紧紧连在一起，没有被其它与之没有联系的中英文所阻隔，例(a)和例(b)中划线的部分就是转换中介：

a) 吃过早饭以后，after having breakfast, after having his, after having his, affter having his breakfast. (学生4)

b) 上气不接下气 quite out of breath, 可能上课过了吧 the class had begun. (学生5)

(2) 内容生成中介

所谓内容生成中介，就是指作者在表述作文内容生成过程中的各种思维活动时所用的母语，例如用母语进行推理、判断、监控、联想、选择、提问等。它与转换中介的不同之处有两点。第一点是用于内容推理、监控、联想、选择和提问的母语，都与要写的外语作文没有直接联系，最终不进入转换阶段，但它有助于作者对图画的理解，最终形成要表达的内容；第二点是，用于判断内容的母语中虽有一部分与英文作文内容有直接联系，但与之相关的英文内容没有紧跟在相应的中文后面，而被其它中英文的思维活动所阻隔。下面用一些例子来说明这个问题。

在例(a)中划线部分是对内容的推理，它与英文作文内容没有直接的联系。斜体字部分“起来晚了”是对内容的判断，与之相联系的英文在后面出现了，但在“起来晚了”和“he woke up late”之间，有一系列与“起来晚了”没有直接联系的中英文交替活动。设想he woke up late 紧接在“起来晚了”后面，“起来晚了”就是转换中介。例(b)和例(c)中的划线部分“还没有把他的意思表达出来”“这太罗嗦了”是作者对自己所写内容的评价或者监控；例(d)中的划线部分是作者对图画内容的联想，或者说是对别人行为的评价；例(a)中“这一幅画到底咋说呢？”和例(e)中“汤姆怎么跟他说话呢？”及“汤姆跟他说的是什么”称为“内容提问”，这一类问题表明作者在内容的形成上碰到了暂时的困难；例(f)中画线的部分是对内容的选择。

a) *“肯定是来晚了，肯定是起来晚了*(内容判断中介)，要不然他咋能会，咋能会救了人，救了一个盲人就迟到呢(内容推理中介)。*起来晚了，起来晚了*(内容判断中介)。穿衣服，穿衣服Tom put on clothes, clothes quickly...从clothes quickly, quickly 这再加上，加上从句吧for he 用哪个系动词呢？ for he, for he, wa wake up late, 算了， woke up, woke up late, woke up late.（学生1）

b) 还没有把他的意思表达出来(内容监控中介)，这一幅画到底咋说呢？(内容提问中介)（学生1）

c) 到了学校之后，他发现已经晚了，and he found when go to his school, he run, 这太罗嗦了(内容监控中介)。（学生2）

d) ...车快来了，这开车的也是，你看不见前面有人吗？哪为啥非要往前开呢？道德品质极其恶劣，这个道德品质好(内容联想中介)，小明过去了，冲上去rushed up and pulled him back (学生5)

e) 汤姆怎么跟他说话呢？（内容提问中介）汤姆要走了必定要跟他说句话（内容推理中介），汤姆跟他说的是什么？（内容提问）（学生3）

f) when he realized what had happened, he was, 当他意识到了发生了什么，可能很高兴（内容判断中介），还是又激动又高兴呢？（内容选择中介）（学生5）

(3) 形式检索中介

形式检索中介用于表述在外语形式检索过程中所产生的各种思维活动，其中包括对语言形式的判断、选择、监控、评价、提问和联想，下列例子中画线的部分说明了母语在此过程中的不同功能：

a) ....a good deed是复数还是单数？（形式选择中介）did a good deed, 作了一件，一件好事（形式推理中介），对，did a good deed（学生2）

b) This time, as he was going to cross the street, He saw, He 应该小写（形式监控中介），这是一个主句，主句（形式推理中介）。（学生4）

c) 再加个，加个 with，再加个with（形式判断中介），no thought for his safety, with no thought for his own safety怪好哩，正好管用，老师讲的这词组正好管用（形式评价中介）。（学生1）

d) 穿过人行道咋说呢？（形式提问中介）（学生5）

e) 用啥词儿呢？他不顾自己的安危，对，上一次老师才讲的完形填空。（形式联想中介）（学生1）

(4) 内容验证中介

有声思维纪录揭示，作者有时在一个外语语言单位形成后，又把它倒译成中文，有时是全译，有时只是译其中的一部分。笔者设想，这里倒译的原因是为了要检验所形成的外语语言单位是否符合

自己要表达的内容，因此这种倒译时使用的母语，我们就把它称之为内容验证中介，例如下列例子中画线的部分：

a) almost strike him 差一点撞着他了（内容验证中介），就在这一霎那间，冲上去，rushed up and pull him back 拉了回来（内容验证中介）。(学生5)

b) when he explained why he 他为什么来晚了why he was late to his teacher 当他给老师解释他为什么来晚了（内容验证中介），when he explained,....（学生3）

c) When Tom passed through the street, he suddenly saw 突然看见（内容验证中介），suddenly saw a blind man 一个盲人（内容验证中介）（学生2）

当然这里不能排斥另一种可能性，即作者倒译是为了照顾前后内容的联系。由于当时没有在有声思维结束后立即和被试者面谈这个问题，因此这里的分析只能是一种假设，有待今后进一步研究。

(5) 程序管理中介

程序管理中介用于决定写作的程序，例如作者用母语思考究竟先写，还是先看下一幅图画；究竟先修改，还是先往下写等活动，此时使用的母语称之为程序管理中介，它与英文作文内容没有联系。由于被试者要求完成的是看图作文，就其程序上的思考在整个写作过程中并不多，但笔者考虑到对母语功能描述的完整性，还是将其单独列为一大类。另外出于将来研究的需要，学生用外语进行命题作文时，用于程序管理的母语肯定是不可缺少的一部分。请看下面例子中画线的部分就是我们所说的程序管理中介。

a) 先别忙，先写上一点(程序管理中介)。想起来，想起来好像，电视里边，跟电视里边放的普通生活中，肯定是...。（学生1）

b) 马上再修改，先写完(程序管理中介) a car went up with high speed（学生5）

2）母语功能的动态描述

根据学生有声思维的数据，我们概括出母语在外语句子写作过程中的动态功能（见图一）。如图一所示，学生接受了图画刺激后，可从阿拉伯数字标出了三种选择中挑选任何一种：(1)图画⟶外语；(2)图画⟶转换中介；(3)图画⟶内容生成中介，开始英语句子的写作。图中所有箭头都用虚线连接，以表示箭头所标出的路径只是可选择的路径之一。凡是箭头标向带虚点的方框，就表示写作过程中有母语参与，其具体功能如框中所注。箭头标向带实线的方框，表示语言形式是外语。双向箭头表示所标出的路径可以形成环路。由于篇幅的限制，我们不可能用例证说明图中标出的所有可能组合的路径。下面我们以例子说明三种途径产生英语句子/从句的过程中母语功能交替使用的情况。

在例（a）中，作者选择了路径1，即在三次接受了图画刺激后，头脑中首先出现的都是英语：(1) After he arrive, got to his school, when; (2)he found that; (3) class has, class, classes, class has, class had, class begin。有趣的是这三个英语语言单位的后面都出现了中文。“当他”和“他发现”是对相应英文的部分或全部倒译，目的在于验证所要表达的内容。思维中出现的“过去完成时”是对形式的监控。

a) After he arrive, got to his school, when 当他（内容验证中介），when he got to school, he found that 他发现（内容验证中介），he found that, class has, class, classes, class has, class had, class begin, he found that 过去完成时（形式监控中介），he found that class had, had begun. (学生2)

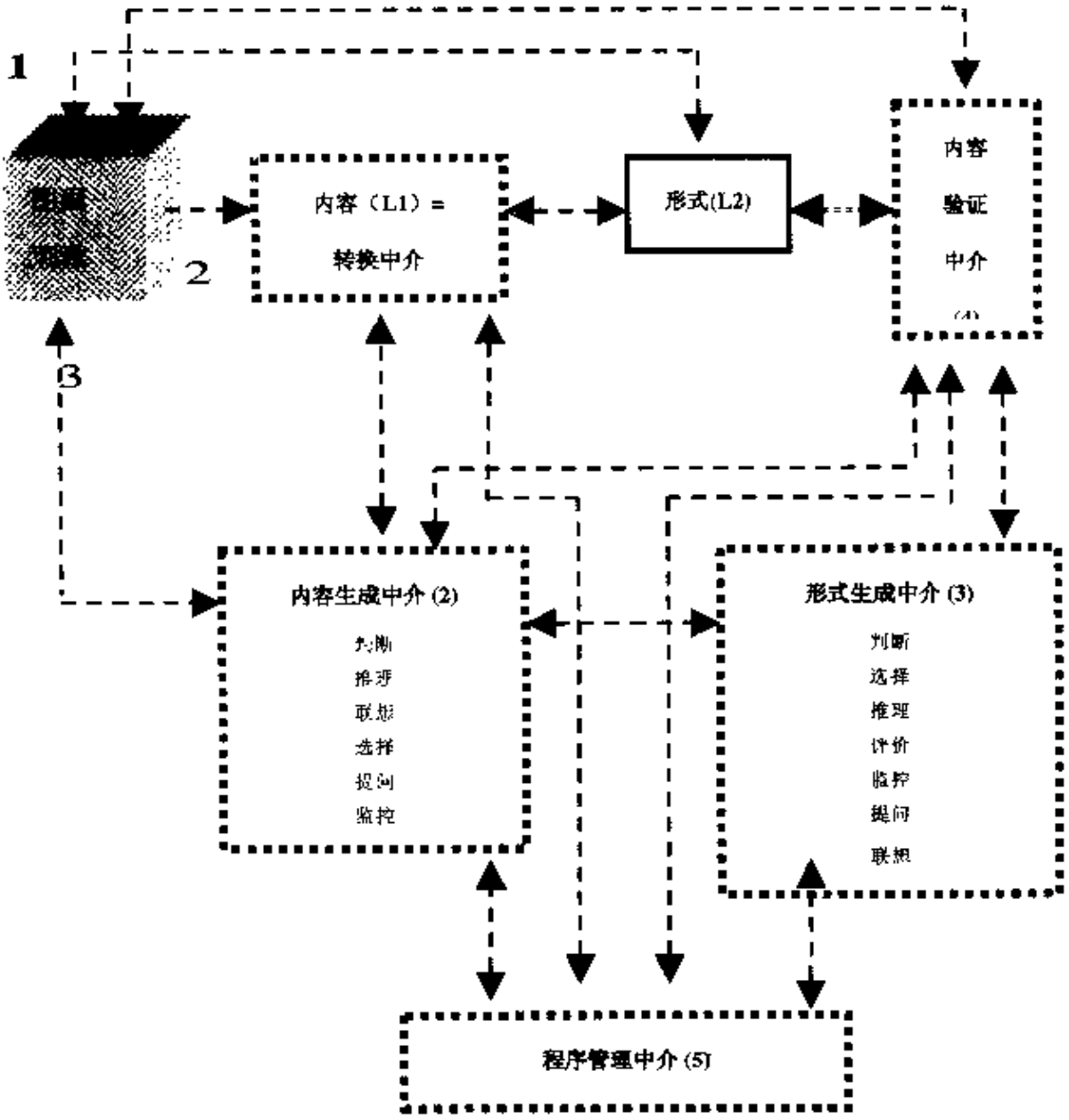
例(b)中作者选择的是路径2，即看了图画后，将要表达的内容先用中文表述出来，然后再以中文为中介分两次转换成英文。当第二次转换结束以后，作者又把英文倒译成中文。

b) 他又穿过马路，就在他上学的路上，他要穿过一条马路 （转换中介） on her, on his way to school, way, to school, s-c-h-o-o-l, on his way to school, he 他要穿过一条马路 （转换中介） he had to, he had to cross, c-r-o-s-s, he had to cross a street 他要穿过一条马路 （内容验证中介） （学生5）。

例(c)的作者选择了路径3：看到图画以后，没有用中文直接形成要表达的内容，而是用中文对图画的内容进行分析，推理以及对写作程序作出决定（见划线部分），其中除了“他往前走”以外，其它部分都与英文作文的内容没有直接关系。但他将“他往前走”转换到一半时，对“when”的用法产生了疑问，于是用中文作为形式判断和推理的中介，解决了“when”的问题，此后作者又对内容“toward school”产生了问题，他先认为可以用toward school, 但立刻又否定了这一想法，提出的理由是“离学校还远着呢”，这时作者不能马上想出解决的方法，只好决定了“先这样写”。

c) 他到底是在等汽车还是在穿马路，这有个路标，路标？哦，不对，这他的腿在往前走着哩。可能是往前走，可能是往前走，正当，可能是正当他，正当他，正当他，正当，他慢慢走，前面有quickly, 他现在慢慢走？（内容推理中介）一会儿再改，先写上，先写出来再讲（程序管理中介）。他往前走（转换中介），when he, he, 用when好些 （形式判断中介）， when表示突然（形式推理中介）， he was walking, he was, he was walking, he was walking toward school, toward school, 用school现在就说（内容判断中介），不对（内容监控中介），离school还远着呢（内容推理中介），先这样写（程序管理中介），toward school （学生1）

图一：在外语句子写作过程中，母语的动态功能




从上述3个例子可以看出，无论学生选择哪一种途径，几乎都逃脱不了母语的参与。

3．高分组和低分组在使用母语功能类别上的差异

在问题一中，我们已经得到结论：在英文看图作文的过程中，高分组学生使用母语参与思维的比例比低分组的学生低14%，这一结果揭示了高分组和低分组使用母语思维的总体趋势，但我们并不能就此推断出这两组学生在使用每一类母语功能上的具体差异。因此，下面我们根据列出的母语功能类别，逐一比较这两组学生使用中文的情况，比较的内容包括三个方面：（1）母语功能类别数量上的差异，即哪一个组使用的功能类别多；（2）每个母语功能类别所用中文词数占中文总词数比例上的差异；（3）母语功能出现的频率和使用中文的平均长度。

1）使用母语功能类别的数量差异

从母语功能的5个大类来看，高分组只有前4类，没有第5类（见表四）。这就是说在看图作文的过程中，高分组中没有人用母语来描述写作的程序，而低分组用了10次，所用的中文词46个，占总中文词数的2%。事实上，低分组需要对程序作出抉择，一般是因为对前面所写的内容发生了疑问，想修改但又不能立刻拿出很好的解决方法。

从内容生成中介和形式检索中介所包括的小类来看，高分组没有“内容选择”，“形式评价”和“形式联想”这三小类。问题是低分组在这几个小类上使用的中文词数占总中文词数的比例也很低。笔者猜想这很可能不是高分组和低分组之间的真正差异。如果样本扩大了，这种情况仍旧存在，我们才能比较有把握地作出结论：高分组使用母语的功能类别比低分组少。

2）每个母语功能类别所用中文词数占中文总词数比例上的差异

表四中列出了某母语功能使用词数占中文总词数的百分比，具体的计算方法如下：

高分组/低分组使用某个母语功能时

$$\frac{\text{所用的中文词数}}{\text{该组使用中文的总词数}} \times 100 = \text{某个母语功能类别所占的百分比}$$

高分组在有的功能类别上所占比例比低分组高，在有的功能类别上，比低分组低。这一结果再不像问题一中所得到的结论那么简单了。高分组使用母语作为转换中介的占中文总词数的37%，而低分组只占15%；高分组在内容验证上使用的中文词数占总中文总词数的16%，是低分组5%的3倍多；高分组用母语作为内容生成中介的比例只有28%，而低分组高达57%；用于形式检索中介和程序管理中介的比例，高分组略低于低分组。现在我们可以对问题一中作出的结论作进一步的修订，即在英文看图作文的思维过程中，虽然高分组总体上使用母语比低分组少，但在作为转换中介和内容验证中介上，比低分组使用的母语却多得多，少只是少用于内容生成中介、形式检索中介和程序管理中介上；虽然低分组在总体上使用母语比高分组多，但用于转换中介和内容验证中介上的母语比例却比高分组低得多。

在5个母语功能类别中，只有用于转换中介的母语跟英文作文的内容有直接联系，因此高分组多用母语作为转换中介，对英文作文质量的提高可能有积极的帮助。内容验证中介是用于检验所形成的外语形式是否与作者要表达的内容相一致，因此它可能对英文作文质量的提高也有促进作用。低分组在内容生成过程中使用了大量的母语，很显然这对写好英文作文帮助不大，因为所用母语中的绝大部分都与作文内容没有任何关系。

那为什么低分组要在这一类活动上花这么大气力和时间呢？是低分组不知道其中的奥秘吗？不能排斥有一部分学生确实缺少这方面的知识和技能，因此对他们需要进行必要的训练。笔者认为问题可能不会如此简单。低分组在内容生成过程中多用母语只是现象，其真正的原因可能是，低分组对图画的理解能力以及用母语写作的能力不如高分组，因此在内容的组织上需要用母语进行反复推敲。当然联想活动的产生，不能说明内容生成上有困难，但过多的联想至少说明作者不知道如何将有限的时间用在写作上，很可能他们在母语写作时也有类似的情况。上述原因仅是笔者的推测，需要进一步的研究。

3)使用中文的平均长度

表四分别列出了高分组和低分组使用每个母语功能的频率以及使用中文的平均长度。所谓使用中文的平均长度就是指每次使用某个母语功能时平均需要的中文词数，具体计算方法为：

$$\text{使用中文平均长度} = \frac{\text{使用某个功能使用的总中文词数}}{\text{某个功能出现的频率。}}$$

以高分组的转换中介为例。根据表四，高分组使用转换中介的频率为87次，共用中文词数463，那么就转换中介而然，使用中文的平均长度为5.3个词，即463除以87。就转换中介和内容验证中介来说，低分组的中文平均长度分别为3.9和3.3，低于高分组的5.3 和3.7。在使用内容生成中介和形式检索中介时，低分组的中文平均长度分别为9.4和3.6，高于高分组的6.2和3.0。由于两组在内容验证中介和形式检索中介上使用的中文平均长度差异不大，因此不作进一步讨论。从逻辑上说，转换中介使用的中文平均长度越长，得到与之内容相近的英文形式也可能越长。如果上述推理能够成立，那么高分组转换的效率就比低分组高。为什么这两组的学生在这方面表现不一样？原因值得进一步研究。 内容生成中介就其功能而言，是为了解决内容形成过程中碰到的问题，作者用中文进行推理、判断、选择、监控、联想和提问。所用中文的平均长度越长，说明问题解决得越不顺利，花的时间越多。从这个意义上说，低分组解决问题的能力比高分组低，其原因已在前面分析过，这里不再重复。

表四：高分组与低分组在使用母语功能类别上的差异

| 功能 / 母语 类别 | | 低分组 | | | | 高分组 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 使用中文的总词数 | 某母语功能使用词数占中文总词数的百分比 | 某母语功能出现的频率 | 平均长度 | 使用中文的总词数 | 某母语功能使用词数占中文总词数的百分比 | 某母语功能出现的频率 | 平均长度 |
| 1.转类 中介 | | 307 | 15% | 79 | 3.9 | 463 | 37% | 87 | 6.3 |
| 2.内养 检证 中介 | | 115 | 5% | 36 | 3.3 | 208 | 16% | 56 | 3.7 |
| 3.内容生成中介 | 1) 判断 | 433 | 20% | 43 | 10.1 | 184 | 14% | 27 | 6.8 |
| | 2) 推理 | 323 | 15% | 20 | 16.2 | 79 | 6% | 8 | 9.8 |
| | 3) 选择 | 26 | 1% | 4 | 6.5 | 0 | 0 | 0 | 0 |
| | 4) 联想 | 147 | 7% | 11 | 13.4 | 11 | 1% | 2 | 5.5 |
| | 5) 监控 | 182 | 9% | 32 | 5.7 | 85 | 7% | 18 | 4.7 |
| | 6) 提问 | 88 | 4% | 18 | 4.9 | 32 | 2% | 6 | 5.3 |
| | | 1199 | 56% | 128 | 9.4 | 371 | 28% | 59 | 6.2 |
| 4.形式检索中介 | 1) 判断 | 182 | 9% | 66 | 4.3 | 139 | 11% | 50 | 2.8 |
| | 2) 选择 | 48 | 2% | 13 | 3.7 | 17 | 1% | 6 | 2.8 |
| | 3) 推理 | 40 | 2% | 6 | 6.7 | 35 | 3% | 8 | 4.4 |
| | 4) 评价 | 45 | 2% | 11 | 4.1 | 0 | 0 | 0 | 0 |
| | 5) 联想 | 14 | 1% | 2 | 7.0 | 0 | 0 | 0 | 0 |
| | 6) 监控 | 54 | 3% | 17 | 3.2 | 30 | 2% | 12 | 2.5 |
| | 7) 提问 | 62 | 3% | 18 | 3.4 | 28 | 2% | 7 | 4 |
| | | 445 | 22% | 123 | 3.6 | 249 | 19% | 83 | 3.0 |
| 5.程序 管理 中介 | | 46 | 2% | 10 | 4.6 | 0 | 0% | 0 | 0 |
| 合计 | | 2112 | | 375 | 6.2 | 1291 | | 285 | 4.8 |

四、结论

该项研究的结果可以归纳如下：

1. 从总体上看，外语写作能力高的学生比能力低的学生对母语依赖性小，但高能力组比低能力组多使用中文作为转换中介和内容验证中介；而低能力组比高能力组多用中文作为内容生成中介、形式检索中介和程序管理中介。就使用中文的平均长度而言，高能力组用于转换中介和内容验证中介的中文平均长度比低能力组长，但低能力组用于内容生成中介和形式检验中介的中文平均长度比高能力组长。由此可见，写作能力高的学生对母语的使用是有选择的。
2. 母语在外语看图作文过程中一般作为转换中介、内容生成中介、形式检索中介、内容验证中介和程序管理中介，它们在外语句子/从句形成的过程中交替出现。
3. 学生接受了图画刺激后，可以有三种选择开始英语句子的写作：(1)图画 ⇒ 外语；(2)图画 ⇒ 转换中介；(3)图画 ⇒ 内容生成中介。学生无论选择哪一种途径，几乎都避免不了使用母语。

以上结果，由于样本数量小，只能作为今后研究的假设。此外该课题还可以从以下几方面作进一步深入研究：（1）将有声思维和面谈结合，让学生自己陈述使用母语的原因；（2）研究学生进行英语命题作文时母语参与思维的情况；（3）将大学生英语看图作文的情况和高中生进行比较；（4）对学生写作过程进行必要的指导是否能控制母语的消极作用，从而发挥其积极作用。

参考文献：

Arndt, V. (1987). Six writers in search of texts: A protocol-based study of L1 and L2 writing. ELT Journal Volume, 41(4), 257-67.

Brooks, E. (1985). Case studies of the composing processes of five "unskilled" English-as-a-second-language writers. Dissertation Abstracts International, 47(1),164A.

Cumming, A. (1987). Decision making and text representation in ESL writing performance. Paper presented at 21st Annual TESOL Convention, Miami, FL.

Diaz, D. (1986). The adult ESL writer: The process and the context. Paper presented at the 76th Annual NCYE Convention, San Antonio, TX.

Friedlander, A. (1990). Composing in English: effects of a first language on writing in English as a second language. In B. Kroll (Ed.). Second language Writing. Cambridge: Cambridge Applied Linguistics.

Jacobs, S. (1982). Composing and coherence: The writing of eleven pre-medical students. Linguistics and Literacy, 3. Washington, DC: Center for Applied Linguistics.

Johnson, C. (1985). The composing process of six ESL students. Dissertation Abstracts International, 46(5), 1216A.

Kobayashi, H. & C. Rinnert. (1992). Effects of first language on second language writing: Translation vs. direct composition. Language Learning, 42(2), 183-251

Lay, N. (1982). Composing processes of adult of ESL learners. TESOL Quarterly, 16(2), 406.

Urzua, C. 1987. "You stopped too soon": Second language children composing and revising. TESOL Quarterly, 21(2), 279-304.

Wen, Qiufang & R. K. Johnson. (1997). L2 learner variables and English achievement: A study of tertiary-level English majors in China. Applied Linguistics, 18(1), 27-48. Oxford University Press.

Zamel, V. (1982). Writing: The process of discovering meaning. TESOL Quarterly, 16(1), 195-209.

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. TESOL Quarterly, 17(1), 165-187.

郭纯洁 刘芳 (1997)，外语写作过程中母语影响的动态研究，《现代外语》，第4期, 30-38页。

# APPENDIX FOUR: QUESTIONNAIRE DESCRIPTION

[Note: the following part is taken from my own thesis (Wen, 1993: 102-107) to illustrate how a questionnaire is described in the methodology chapter]

5.4.1 Language Learner Factors Questionnaire

In the following section, the structure of the questionnaire will be described first. It is followed by brief descriptions of the content of the questionnaire items. Finally, sources of the questionnaire items will be given.

1) Descriptions of the structure of the questionnaire

The Language Learner Factors Questionnaire consists of three parts. Table 5.3 presents the structure of the questionnaire along with the number of items for each part. Part A can be further divided into two subparts: one is about personal details (i.e., name, sex, date of birth, the secondary school where they studied, the scores of English and Chinese matriculation tests and an estimate of time spent in studying English outside of class). The other section of Part A is to establish students' reasons for learning English. Students rated a range of reasons given on a five-point scale: from not important (1) to extremely important (5). Part B consists of statements of beliefs about language learning. The students indicated their opinions in terms of a five-point scale: from strongly disagree (1) to strongly agree (5). Part C contains statements concerning learning strategies. Some are management strategies and the rest, language learning strategies. Students again responded on a five-point scale: from 'This statement is never or almost never true of me (1)' to 'This statement is completely or almost completely true of me (5)' (See Appendix 1).

| | Content | No. of items |
|---|---|---|
| Part A | (1) Personal details: name, sex, date of birth<br>(2) Score on the Matriculation English Test<br>(3) Score on the Matriculation Chinese Test<br>(4) Number of hours spent outside class<br>(5) Reasons for learning English | 7 (A01 to A07) |
| Part B | Beliefs | 31 (B01 to B31) |
| Part C | (1) Management strategies<br>(2) Language learning strategies | 22 (C01 to C22)<br>56 (C23 to C80) |

Table 5.3: Structure of the Language Learner Factors Questionnaire

2) Descriptions of the content of the questionnaire

The following section will briefly describe the content of the questionnaire items. Descriptions will be first about the items concerning students' perception of learning purposes, then about the items representing beliefs, and finally about the items used as indicators of management strategies and language learning strategies.

a) Learning purposes

Seven items addressing learning purposes fall into two conceptual categories: extrinsic learning purpose and intrinsic learning purpose. Extrinsic learning purposes are represented by items such as "I need Englisb for my future career.", "I need English for my studying abroad", "I need English for my further study in China" and intrinsic learning purposes by items such as "I like learning a foreign language", "I am interested in the English culture".

b) Beliefs

31 belief items were constructed centring around four conceptual areas: (1) beliefs about the importance of effort (i.e. attribution beliefs); (2) beliefs about the importance of management (i.e. managment beliefs); (3) beliefs about the importance of traditional strategies (i.e. traditional beliefs) and (4) beliefs about the importance of non-traditional strategies

(i.e. non-traditional beliefs). Indicators of attribution beliefs are items such as "Learning a foreign language requires painstaking effort", "Effort is more important than inborn ability in learning a foreign language". Items like "Planning your study time is important for success" and "Setting your learning goals and evaluating your progress is important for success" represent management beliefs. Traditional beliefs are learners' opinions about importance of repetition, memorising vocabulary, sentence patterns and texts, intensive study of texts and using the mother tongue as a mediator. Questionnaire items describing traditional beliefs include "It is important to repeat a lot", "Memorising sentence patterns is important for learning a foreign language", "Intensive study of texts is important for learning a foreign language", "To understand a text well, the best way is to translate it". Non-traditional beliefs are learners' views about importance of communicative activities and thinking in English, which are expressed in questionnaire items such as "Reading foreign newspapers, magazines, novels, etc. is important for learning a foreign language", "Listening to a wide range of materials is important for learning a foreign language", "When using English, it is best to keep Chinese out of your mind".

As was described in Chapter Two, the basic assumption underlying the questionnaire items concerning traditional and non-traditional beliefs and strategies is Stern's three continua (i.e. L1-L2, explicit-implicit and code-communication). Specifically speaking, these two groups of questionnaire items can be placed at the two extremes of the three continua (See Section 2.3.4, Chapter Two).

c) Management strategies

Management strategies were assumed to include planning, goal-setting, self-evaluation, searching for learning resources and options, study habits and affective-control. Questionnaire items as indicators of these strategies were constructed accordingly, for example, "I plan my schedule so I will have enough time to study English", "I have goals for improving my English skills", "I evaluate my learning strategies to find

out the ineffective ones and improve them", "I ask other people to tell me good strategies for learning English", "When I come across a problem in learning English, I try to solve it myself before I ask the other people for help" and "When I do less well than my fellow students, I double my effort".

d) Language learning strategies

Questionnaire items concerning language learning strategies are divided into two general categories: some items representing traditional strategies and the others non-traditional strategies. The former includes "When reading texts, I try to understand everything in it", "I make grammatical analyses of sentences when I cannot understand the meaning", "After class I read texts alound until I can read them fluently", "When reading a text, I translate it to help me understand its meaning", etc. The latter contains "I read (e.g. English newspapers, magazines or novels) on my own initiative", "I listen to English broadcasts on my own initiative", "I try to avoid thinking in Chinese when communicating in English".

3) Sources of the questionnaire items

The specific items concerning the two types of learning purposes were constructed partly based on the purposes for foreign language education prescribed by the State Education Commission and partly from Gardner's motivation Battery (1985). The questionnaire items related to the management component are derived from the instrument designed by Oxford (1990) and Biggs' Study Process Questionnaire for use with tertiary students (1986).

The questionnaire items concerning the two approaches were derived from various sources. The items regarding traditional approach were primarily derived from three sources. The first was from interviews in which four successful Chinese language learners who are now university English teachers were asked to describe their beliefs about how English was best learned and their learning strategies from their own experience. The second was from papers concerning English language teaching and

learning in China in which claims are made about what constitutes a traditional Chinese approach to language learning (Cheng, 1982; Hao, 1982; Harvey, 1985; Huang, 1984; Maley, 1984; Penner, 1991; Sampson, 1985; Scovel, 1982; Shen, 1984; Tai, 1989; Yuan, 1982). The third was from instruments designed by western researchers, in particular, Horwitz's Beliefs about Language Learning Inventory (1987), Politzer's Questionnaire of Learning Behaviours (1983) and Oxford's Strategies Inventory for Language Learning (1990). The items related to the Non-traditional approach were primarily selected from the instruments developed by the Western researchers just mentioned.

The items in the questionnaire were presented in English. Upon completion of the initial questionnaire, 18 English teacher-trainers from China who were visiting scholars in the Faculty of Education, Hong Kong University were asked to read through the questionnaire items and find out possible difficult words and phrases for the second-year English majors. Upon their suggestions, eight words and phrases were given Chinese equivalents in brackets. The questionnaire items were then tested in a pilot study undertaken in 1990 and revisions were tuade. The summary of the pilot study is presented in Appendix 2.

# APPENDIX FIVE: RANDOM NUMBERS

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 12651 | 61646 | 11769 | 75109 | 86996 | 97669 | 25757 | 32535 | 07122 | 76763 |
| 81769 | 74436 | 02630 | 72310 | 45049 | 18029 | 07469 | 42341 | 98173 | 79260 |
| 36737 | 98863 | 77240 | 76251 | 00654 | 64688 | 09343 | 70278 | 67331 | 98729 |
| 82861 | 54371 | 76610 | 94934 | 72748 | 44124 | 05610 | 53750 | 95938 | 01485 |
| 21325 | 15732 | 24127 | 37431 | 09723 | 63529 | 73977 | 95218 | 96074 | 42138 |
| 74146 | 47887 | 62463 | 23045 | 41490 | 07954 | 22597 | 60012 | 98866 | 90959 |
| 90759 | 64410 | 54179 | 66075 | 61051 | 75385 | 51378 | 08360 | 95946 | 95547 |
| 55683 | 98078 | 02238 | 91540 | 21219 | 17720 | 87817 | 41705 | 95985 | 12563 |
| 79686 | 17969 | 76061 | 83748 | 55920 | 83612 | 41540 | 86492 | 06447 | 60568 |
| 70333 | 00201 | 86201 | 69716 | 78185 | 62154 | 77930 | 67663 | 29529 | 75116 |
| 14402 | 53536 | 07779 | 04157 | 41172 | 36473 | 42123 | 43929 | 50533 | 33437 |
| 59911 | 08256 | 06596 | 48416 | 69770 | 68797 | 56080 | 14223 | 59199 | 30162 |
| 62368 | 62623 | 62742 | 14891 | 39247 | 52242 | 98832 | 69533 | 91174 | 57979 |
| 57529 | 97751 | 54976 | 48957 | 74599 | 08759 | 78494 | 52785 | 68526 | 64618 |
| 15469 | 90574 | 78033 | 66885 | 13936 | 42117 | 71831 | 22961 | 94225 | 31816 |
| 18625 | 23674 | 53850 | 32827 | 81647 | 80820 | 00420 | 63555 | 74489 | 80141 |
| 74626 | 68394 | 88562 | 70745 | 23701 | 45630 | 65891 | 58220 | 35442 | 60414 |
| 11119 | 16519 | 27384 | 90199 | 79210 | 76965 | 99546 | 30323 | 31664 | 22845 |
| 41101 | 17336 | 48951 | 53674 | 17880 | 45260 | 08575 | 49321 | 36191 | 17095 |
| 32123 | 91576 | 84221 | 78902 | 82010 | 30847 | 62329 | 63898 | 23268 | 74283 |
| 26091 | 68409 | 69704 | 82267 | 14751 | 13151 | 93115 | 01437 | 56945 | 89661 |
| 67680 | 79790 | 48462 | 59278 | 44185 | 29616 | 76531 | 19589 | 83139 | 28454 |
| 15184 | 19260 | 14073 | 07026 | 25264 | 08388 | 27182 | 22557 | 51501 | 67481 |
| 58010 | 45039 | 57181 | 10238 | 36874 | 28546 | 37444 | 80824 | 63981 | 39942 |
| 56425 | 50996 | 86245 | 32623 | 78858 | 08143 | 60377 | 42925 | 42815 | 11159 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 82630 | 80466 | 13592 | 60642 | 17904 | 99718 | 63432 | 88642 | 37858 | 25431 |
| 14927 | 40909 | 23900 | 48761 | 44860 | 92467 | 31742 | 87142 | 03607 | 32059 |
| 23740 | 22505 | 07489 | 85986 | 74420 | 21744 | 97711 | 36648 | 35620 | 97949 |
| 32990 | 97446 | 03711 | 63824 | 07953 | 85965 | 87089 | 11687 | 92414 | 67257 |
| 05310 | 24058 | 91946 | 78437 | 34365 | 82469 | 12430 | 84754 | 19354 | 72745 |
| 21839 | 39937 | 27534 | 88913 | 49055 | 19218 | 47712 | 67677 | 51889 | 70926 |
| 08833 | 42549 | 93981 | 94051 | 28382 | 83725 | 72643 | 64233 | 97252 | 17133 |
| 58336 | 11139 | 47479 | 00931 | 91560 | 95372 | 97642 | 33856 | 54825 | 55680 |
| 62032 | 91144 | 75478 | 47431 | 52726 | 30289 | 42411 | 91886 | 51818 | 78292 |
| 45171 | 30557 | 53116 | 04118 | 58301 | 24375 | 65609 | 85810 | 18620 | 49198 |
| 91611 | 62656 | 60128 | 35609 | 63698 | 78356 | 50582 | 22505 | 01692 | 36291 |
| 55472 | 63819 | 86314 | 49174 | 93582 | 73604 | 78614 | 78849 | 23096 | 72825 |
| 18573 | 09729 | 74091 | 53994 | 10970 | 86557 | 65661 | 41854 | 26037 | 53296 |
| 60866 | 02955 | 90288 | 82136 | 83644 | 94455 | 06560 | 78029 | 98768 | 71296 |
| 45043 | 55608 | 82767 | 60890 | 74646 | 79485 | 13619 | 98868 | 40857 | 19415 |
| 17831 | 09737 | 79473 | 75945 | 28394 | 79334 | 70577 | 38048 | 03607 | 06932 |
| 40137 | 03981 | 07585 | 18128 | 11178 | 32601 | 27994 | 05641 | 22600 | 86064 |
| 77776 | 31343 | 14576 | 97706 | 16039 | 47517 | 43300 | 59080 | 80392 | 63189 |
| 69605 | 44104 | 40103 | 95635 | 05635 | 81673 | 68657 | 09559 | 23510 | 95875 |
| 19916 | 52934 | 26499 | 09821 | 97331 | 80993 | 61299 | 36979 | 73599 | 35055 |
| 02606 | 58552 | 07678 | 56619 | 65325 | 30705 | 99582 | 53390 | 46357 | 13244 |
| 65183 | 73160 | 87131 | 35530 | 47946 | 09854 | 18080 | 02321 | 05809 | 04893 |
| 10740 | 98914 | 44916 | 11322 | 89717 | 88189 | 30143 | 52687 | 19420 | 60061 |
| 98462 | 89822 | 71691 | 51573 | 83666 | 61642 | 46683 | 33761 | 47542 | 23551 |
| 60139 | 25601 | 93663 | 25547 | 02654 | 94829 | 48672 | 28736 | 84994 | 13071 |

(Black, K. 1992: A2-A3)

# Subject Index

# Author Index

Nunan 6, 199, 203, 224
Oxford 121
Oxford and Crookall 18
Patton 95, 188
Punch 58, 89,.97, 106, 140, 150, 159, 292
Qin 295
Raimes 191
Rivers 200
Rosenberg 161
Rudestam & Newton 8, 42, 55, 305, 306, 327
Schmidt and Frota 202
Skehan 46
SPSS Inc. 229, 235, 249, 255
Stake 180
Stern 120
Strauss and Corbin 274, 282
Strauss 281,282
Su 295
Tashakkori and Teddlie 90, 93, 97, 102, 106
Theodorson & Theodorson 179
Wang 26, 27, 70, 179, 295
Wen 16, 117, 121, 179, 200, 295, 298, 336, 337, 346
Wen & Wang 101
Wen & Wu 15
Wen and Gao 24, 266
Wen and Guo 329
Wu 14, 59, 126, 295
Xu 60, 252
Ye 14
Zamel 191
Zhu 61, 295