

ELT-15

# Research in the Language Classroom

Milestones in ELT

ELT Documents: 133

# Research in the Language Classroom

Editors: Christopher Brumfit  
Rosamond Mitchell

MODERN ENGLISH PUBLICATIONS  
in association with The British Council

## **Milestones in ELT**

The British Council was established in 1934, and one of our main aims has always been to promote the wider knowledge of the English language. Over the last 75 years, we have issued many important publications that have set the agenda for ELT professionals, often in partnership with other organisations and institutions.

As part of its 75th anniversary celebrations, we are re-launching a selection of those publications online. Many of the messages and ideas are just as relevant today as they were when first published. We believe they are also useful historical sources through which colleagues can see how our profession has developed over the years.

## **Research in the Language Classroom**

This book, from 1990, looked at the practical benefits for teachers of classroom research. It consists of papers from practitioners from countries in Europe, North America and Australasia, but with detailed reference to a wider variety of global teaching contexts. Topics range from 'Investigating Learners' Language' to 'Researching Teachers: Behaviour and Belief'.

c/14283 (Rec'd 23.1.91)

£9.95

# Research in the Language Classroom

ELT Documents 133

**Editors: CHRISTOPHER BRUMFIT  
ROSAMOND MITCHELL**

Centre for Language in Education  
University of Southampton

1989

CENTRE FOR LANGUAGE IN EDUCATION  
UNIVERSITY OF SOUTHAMPTON

ELT DOCUMENTS 133



Modern English Publications in association with The British Council

© Modern English Publications and The British Council 1990

*All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without permission in writing from the copyright holders.*

Typesetting by Vine & Gorfin Ltd, Exmouth, Devon  
Printed in Hong Kong

A CIP catalogue record for this book is available from  
the British Library

ISBN 0 333 51216 2

# Contents

## Preface

### I General Issues v

**The Language Classroom as a Focus for Research** 3  
CHRISTOPHER BRUMFIT AND ROSAMOND MITCHELL

**The Teacher as Researcher** 16  
DAVID NUNAN

**Ethnography: Bandaid, Bandwagon, or Contraband?** 33  
LEO VAN LIER

**Researching Classroom Language Learning** 54  
ROD ELLIS

**The I-Language Approach and Classroom  
Observation** 71  
VIVIAN COOK

### II Specific Approaches

**A Look at the Research Process in Classroom  
Observation: A Case Study** 81  
NINA SPADA

**Investigating Learners' Language** 94  
PETER S. GREEN AND KARLHEINZ HECHT

**Researching Teachers: Behaviour and Belief** 117  
TONY LYNCH

<b>Learner Diaries</b>	128
BRIAN PARKINSON AND CHRISTINA HOWELL-RICHARDSON	
<b>Classrooms and Display Questions</b>	141
LYNN BANBROOK AND PETER SKEHAN	
<b>III Special Issues</b>	
<b>Who Should Evaluate L2 Programs?</b>	155
ALAN BERETTA	
<b>Second Language Classroom Research and Teacher Education</b>	161
MICHAEL H. LONG	
<b>Notes on Contributors</b>	171



# Preface

The purpose of this volume is to survey major issues in classroom research, locating them within general linguistic and educational research traditions.

Following a general introductory paper by the editors, a number of contributors address general issues, illustrating them from specific experience of engaging in research. In the second section, a range of particular approaches are explored in more detail, with papers that look at the relationship between teachers and learners framing accounts of investigations of teachers and learners separately. In the final section, issues in programme evaluation and teacher education are addressed specifically from a research perspective.

We have been very lucky to obtain contributions from a wide range of distinguished practitioners in many major centres of research in Europe, North America, the Pacific and Australia. For reasons beyond our control, this volume is appearing rather later than we had intended. We are most grateful to contributors for their patience, and hope they will be pleased with the result.

CJB  
RFM

December 1989



# ***I General Issues***



# The Language Classroom as a Focus for Research

Christopher Brumfit and Rosamond Mitchell

*University of Southampton*

It will be apparent from the varied approaches taken by contributors to this volume that there are many different views on the role of research in language teaching. A comparison between the views of van Lier and Long, or of Cook and Nunan, will at once make clear that there are different views about what constitutes relevant research, what are the most appropriate procedures to use, and who are the best qualified people to carry it out.

Language teaching needs as much serious investigation as possible, of as many different kinds as possible. Whether we call such investigation 'research' is less important than whether the procedures adopted and the people conducting the study are appropriate for the questions being addressed.

There is a sense in which serious practitioners, in any field, will expect to monitor their own performance as objectively as possible. In this sense, language teachers should always be engaging in 'research' if they are to work responsibly and professionally. But beyond this is a whole tradition of educational and linguistic research which could be linked in with the concerns of language teachers. In the rest of this paper, we shall explore some of the major areas of debate which underlie the more specific discussion of later papers.

## **1 Language learning and language teaching**

It is unnecessary to persuade readers of this book that language learning is a crucial phenomenon that deserves study. However, it is worth also emphasising that, for many learners, teaching provides the context for the learning process, and deserves analysis in its own right. Particularly, we have a greater chance of controlling processes of teaching than of learning, so understanding how it works in practice, and how its workings relate to successful language development, is a necessary adjunct to effective education.

At the same time, it must be emphasised that we lack empirical support for much of what is claimed to be 'known' about teaching methodology, as

Long argues in his paper in this collection. We would not wish to claim that this lack invalidates the understandings of teacher trainers or methodologists—there are many valuable practices in language teaching that will never be capable of rigorous falsification in a scientific sense, any more than the success or failure of a particular marriage will be capable of investigation in terms of a single neat and falsifiable explanation. No major area of human relationship can be addressed truthfully by such a mechanistic approach. But that does not mean that we should neglect empirical study as shamefully as we have done in the past. There is a great deal that we do not know, but could in principle know, about practices in classrooms. Until we have built up our knowledge base much more substantially, methodologists and theoreticians will always be open to the attack that they are basing their proposals too much on hunch, the peculiarities of their own personal experience, or a lazy acceptance of tradition. Scholarship and empirical study must combine, but—particularly in the British tradition of language teaching—we have had far more scholarship than empirical investigation. The result of this is that exciting recommendations have abounded in the last twenty years: functional–notional syllabuses, procedural syllabuses, needs analyses, authentic materials, and graded objectives, to name only a few, but there has been remarkably little descriptive work. The exceptions to this generalisation are more concerned with languages other than English, though their work provides models for EFL teachers to adapt: Mitchell, Parkinson and Johnstone (1981), Sanderson (1983), Kasper (1986), Peck (1988) and Mitchell (1988).

Yet the substantial empirical work carried out on second language acquisition and associated areas, predominantly in North America and mainland Europe, does not enable us to answer many of the important questions that language teachers are concerned with. We need many more studies before we can even begin to produce a coherent, empirically-based account of all the important factors in language teaching.

## **2 Educational research, ‘disciplinary’ study and interdisciplinary study**

Language teaching is a form of education, and educational research is inevitably interdisciplinary. Indeed, it is clear that education is a very important field for research, partly because so much money is invested in it, and partly because it raises key questions about the nature of human experience. Human beings are by nature educating creatures, for they are unique in their ability to pass on culturally-acquired characteristics to their descendants, and the medium for doing much of this is explicit education. Understanding educational processes is thus an important part of understanding what makes us distinctive creatures. And central to the educational process is the role of language, and the learning of new

languages, dialects and modes of discourse. Understanding language teaching and learning will contribute to our understanding of language, of education, and of the human condition.

But to say that education is a crucial area of human endeavour is not to say that it is a single formal discipline. Rather, it is a field of study which may be explored from a variety of disciplinary standpoints. Philosophers, psychologists, sociologists, historians, linguists, anthropologists, or economists, indeed any human scientists, may choose to examine educational institutions or educational processes. It will be clear from these papers that psycholinguists, sociolinguists, methodologists and sociologists all feel they have understandings to gain from examining language learning, and useful contributions to make to the theory of language in education.

Some of the traditions represented in these disciplines will be predominantly experimental or observational in character. Indeed, most of the papers in this collection represent such research traditions. But some disciplines (for example, history), have traditionally been concerned more with documentary evidence than with direct observation: others (for example, philosophy) have been concerned with the construction of clear arguments from evidence provided in other disciplines such as psychology or political science, or with the creation and analysis of hypothetical but illuminating scenarios, rather than with empirical study. Within traditions such as these, 'research' means something very different from its meaning for the empirical sciences.

'Education', however, needs all these kinds of research. Serious methodological discussion of language teaching may be regarded as the equivalent of philosophical discussion in other spheres. Just as lawyers gain from the study of jurisprudence, literary critics from critical theory, historians from historiography and scientists from philosophy of science, because these provide them with the tools for avoiding blind following of tradition, so too language teachers gain from methodological discussion by people who understand language teaching from experience, and can interpret and assess the significance of appropriate theory and research.

But we may acknowledge the need for all these types of research, and at the same time feel that the proportion devoted to some types is too great for the good of the profession. Too much interpretation and too little empirical study may result in myths being perpetuated without being tested against recent observation or experiment; too much empirical study and too little interpretation may result in nothing being discussed that cannot be measured by the currently available technology, or only the currently fashionable research areas being considered relevant to the process of language learning. An ideal balance should be sought—but the balance will be dependent on the social conditions in a particular place at a particular time. What is attainable now in the United States, where research students require a PhD degree to qualify for many professions, and consequently are willing to work on projects at this level, will be harder to attain in countries where a PhD is less necessary for career

advancement. What is attainable in one decade, when there is a fashion for funding educationally-related work, may not be attainable in another when external funding agencies divert their attention to work in other areas, or reduce the gross amount of money available for research. What is possible in countries where teachers have a high academic training in research methods may not be possible elsewhere without major, expensive changes in training policy.

Yet, in spite of the different social contexts of language teaching, a research perspective towards our work will always be desirable, in all places and under all conditions. To say this is simply to restate the requirement that language teaching should be carried out responsibly and professionally. How that perspective is realised in practice will depend on funding, the training and experience of teachers and administrators, the relationship between the professional research community (if any) and schools, and other factors that vary from place to place. What is undesirable, though, is the exclusion of some types of enquiry, limiting research to that within one tradition only, and denying recognition, and funding, to others. Variety of approach is essential if the richness of language use and educational opportunity is to be adequately understood.

At the same time, the enormous lack of descriptive work in classrooms makes this the most neglected area of research. It is also, of course, a mode of research to which teachers could contribute most easily. This book concentrates on empirical studies in the classroom context in the hope that much more data will become available for descriptive and comparative purposes.

### **3 Approaches to research**

A 'researching' attitude may be defined as the systematisation of curiosity. Research is essentially a cumulative process, building on what has gone before. Unfortunately, the term 'research' has become debased, so that anyone enquiring into the necessary background facts for a television programme will be listed in the credits as a researcher. However, it is worthwhile to list some of the features that should distinguish serious research from simply the search for information.

Competent research will involve:

1. Careful formulation of the questions to be investigated, to ensure that they are not phrased in such a way as to confuse major issues with minor ones, or to embrace many different questions within one vague, general topic which is incapable of being investigated systematically;
2. Careful exploration of the best means of investigation for the particular question being addressed;
3. Consideration of the major previous attempts to explore the same and closely related questions, in order to borrow and adapt



- appropriate formulations of the questions, and appropriate modes of investigation, and to learn from their mistakes and difficulties;
4. Explicit accounts of the process of question formulation, the criteria for selection of the research techniques, and the reasons for the questions being felt to be important in the first place;
  5. Full documentation of the procedures used, the means by which information has been gathered, and the methods of interpretation and analysis which have been adopted;
  6. Explicit acknowledgement of all previous work which has contributed to the conceptualisation, means of collection, and procedures for analysis of the data collected;
  7. Specific interpretation of the data collected, to assess its usefulness in relation to the initial research questions;
  8. Evaluation of the extent to which the project has achieved its aims, together with an account of the ways in which the process of research has led to changes in the initial formulation of questions—and any other relevant judgements that the researcher may wish to pass on to interested readers;
  9. A willingness to publicise the research, so that it can contribute to further development by others to the exploration of the same, or related, questions.

By the definition we are proposing, research is characterised by being a public, systematic and useful activity. It is public because it needs to be distinguished from simply improving one's own private understanding: it is not another name for personal study. It needs to be public because private work is necessarily inefficient. Public work benefits from having procedures throughout being open to scrutiny by others who will lack the biases of the original researchers (though they will have others of their own), who will bring further understanding to bear on the same problems, and who will be able—above all—to offer public criticism as a result of which methods and formulations can be improved upon in subsequent work. Private research would not benefit from this, and would thus risk being tied to the idiosyncrasies of one set of researchers, being limited to what one person or group of people knew about, and being sympathetic to the interests of those who formulated the proposal, and thus possibly unintentionally biased.

Research must be systematic, because it needs to be explicit about its procedures if it is to be distinguished from mere hunch. What kind of systematicity will be sought depends on the question being investigated, but for every question, the means of exploration will be examined exhaustively to ensure that it is the best that can be devised for the time and resources available, and the formulation and interpretation of the research will be systematically examined as rigorously as possible.

Research must be useful, in a particular sense. It will not necessarily be useful in its immediate results—otherwise we would be committed to producing the results which most fitted with what we most wanted to find

out! But the question to be investigated must have some useful point to it; it must be something that we need to know. A recognition of where particular problems are is a necessary control on irresponsible mind-games. While irresponsibility, divergent thinking and imagination will all contribute to the wide range of potentially fruitful ideas that researchers may exploit, funding, time and expertise are limited. Justifying a particular area for investigation involves necessarily showing how the study will contribute, directly or indirectly, to improving our understanding of language teaching.

But for this understanding to be useful, we do not need to adopt a narrowly utilitarian point of view—all we need to do is to show that we have thought clearly about why we should indulge our curiosity (in a way that is relatively expensive in time and effort) on this topic rather than another. The benefits are the public scrutiny which should improve the effectiveness of what we do by laying us maximally open to advice and criticism. But the right to these benefits needs to be argued by a demonstration that we are not merely being self-indulgent.

Within this framework, there are three main types of research. Often researchers may be dependent on external funding, but individual investigations may fit into these categories also.

*'Pure' research:*

The prime intention of this will be to increase our understanding of the areas of current study that we are least clear about. For language teaching these may be questions in psychology, sociology, linguistics, or any other relevant field—or questions of pedagogy which are motivated by a concern to understand what it is that makes certain procedures appear effective in the classroom. This form of research tends to address questions like 'Why do things happen as they do?' and 'How do we explain particular events and particular relationships?' The cumulative effect of research of this kind is to build up a more or less agreed picture (or 'model') of how things work in a particular field. Such agreed understanding influences the ways in which politicians and decision-makers plan, so research has relevance to any conceivable future activity in the area being investigated.

*Policy-orientated research:*

Such research is concerned with monitoring the effects of policies which have already been decided upon. To know whether a particular programme is realising its stated aims and is therefore effective, whether it is being effective, but in unexpected and unanticipated ways, or whether it is being ineffective, is clearly valuable for sponsors and planners. It enables future planning to be more effective, and sometimes enables adjustments to be made to programmes which are failing to achieve their intended effects. Such research may lead to a questioning of underlying theory, and contribute to formulation of better explanatory models, but that is not its prime intention. Its main purpose is to monitor

the effect of policies. In practice, indeed, many sponsors prefer to discover that their policies are working, especially if a lot of money has been invested in (for example) teaching materials or a new mode of examination. Thus there are potential political risks in this type of research which need conscious guarding against.

*'Action' research:*

It is frequently argued that teachers should devote their time most fruitfully to this type of research. The argument usually made is that this is useful because it is aimed at a particular situation, without the concern for replicability and generalisation which characterises pure research, and without the policy having been previously defined, as in policy-orientated research. It is further argued that action research aims additionally at educating teachers by the process of becoming self-conscious about their work, and that this is as important an aim as evaluating or information-gathering. Certainly, research by teachers, for teachers, on their processes of teaching can only be a good thing. But if obtaining a clearer understanding of teaching processes requires care and rigour in other modes of research, there is no good argument for action research producing less care and rigour unless it is less concerned with clear understanding, which it is not. The real point about action research is that it is closely tied to the particular interests and needs of particular teachers—but these require the same kind of careful support as any other serious investigation.

What distinguishes these three types of research is not so much the procedures used, as the context in which they operate. Typically, pure research will be performed by professional researchers (who may also work simultaneously as higher education teachers), perhaps in collaboration with practitioners. The formulation of the problem will depend on the work of scholars and researchers, and the work may be funded by research councils, private charitable trusts, or government agencies, and will usually carefully describe its relationship with all previous relevant research.

Policy-related research will be performed by professional researchers, sometimes in formally-set-up research institutes attached to official organisations. The work may be collaborative with practitioners. The formulation of the research topic will usually be defined by the policy, and the funding agency will usually be the policy-making body. Relationship with previous research may be explicit, but this is not essential.

Action research will be performed by practitioners, on topics formulated by practitioners. If there is external funding, it may well come from sources whose main concern is with activating the teaching profession rather than with a prime interest in whatever results are obtained.

Nonetheless, in spite of these contextual differences, if the research is to be of more than personal value (and hence to justify the term 'research')

at all), they will all be the same in following the general procedures we have outlined above.

#### **4 Being a teacher and being a researcher**

It would be unwise to pretend that the processes of researching and teaching can be combined without effort. There is a particularly strong contradictory pull in that research is a type of contemplation (however systematic) while teaching is a type of action. Pure research will require explicit hypothesis creation, with clear and falsifiable statements that can be rigorously analysed and tested. This process will involve simplifying from the rich context of particular classrooms that teachers are specially trained to respond to. All teachers know that classrooms are confusing places, with different agendas being pursued by different participants, with many different activities and thought processes happening simultaneously, and with different needs being addressed in all directions. Few lessons follow plans exactly, and few plans follow principles as precisely as theorists intend. Teachers work in a world of real people, real motives and conflicting interests, and their prime task is to survive in this world, in order to influence learning and direct it towards the most profitable activities and routines for success. Distancing yourself from this experience enough to be able to see it in relation to the experiences of others in similar situations is part of maturing as a teacher—but it is not easy, and demands sensitivity and commitment.

At the same time, it is vitally important that this distancing is practised by teachers, and that research is not carried out exclusively by outsiders. Understanding what it is to be a teacher, from direct and substantial experience, is different from understanding teaching processes from outside. Knowing how to teach, knowing as an outsider what teachers do, and knowing from experience the practice of teaching, are three different modes of understanding. By combining them, a fuller picture of the teaching process will be built up than by relying on one only. And exactly the same points can be made about knowing language learning processes. The risk for teachers is that research will come too much from outside the profession, so that the contingent and contradictory experiences actually encountered in schools and classrooms are neglected in the efforts of outside researchers to be tidy and concentrated.

For these reasons, the health of research into educational matters is dependent on combinations of research from external perspectives, collaborative research by teachers and outsiders, and research from within teaching itself.

Further, work of this kind is not only justified by the need for research to be sensitive to the social experiences of teachers and learners; it is also justified as a contribution to effective teaching. As with any important

activity, teaching will benefit from being reflected on and examined systematically. We should not forget, however, that many competent teachers reflect very little without failing to be teachers, but we would not define as a teacher someone who reflects or researches all the time. Participant research involves accepting some degree of tension between different modes of thinking and acting. The justification rests in the indirect improvements to practice which will result from increased understanding, and the sense of mastery of our environment which comes from seeing it more clearly.

## **5 Types of research**

The fullest and clearest account of ways of approaching educational research that we know of is Cohen and Manion (1989). Quantitative approaches to applied linguistic research are explained in detail in Hatch and Farhady (1982). Recent books on classroom research include an authoritative survey of recent empirical work on language instruction and its relationship with learning outcomes (Chaudron, 1988), an account of the historical development of classroom research with key texts (Allwright, 1988), and a critique of over-positivist research procedures by van Lier (1988, and see also his contribution to this volume). A shorter summary of major issues in classroom research can be found in Mitchell (1985).

In the final section of this paper we shall briefly describe the general types of research available to teachers. These themes are explored at greater length in Cohen and Manion (1989), and are taken up for language classrooms in a number of papers in this collection.

First, it is useful to distinguish between description, intervention, and experiment.

Descriptive research will aim at providing as accurate an account as possible of what current practice is: how learners do learn, how teachers do teach, what classrooms do look like, at a particular moment in a particular place. In itself such work will only be illuminating if the description is carried out with a particular intention in mind. A description that included everything conceivable that happened in a classroom would be unwieldy and incomprehensible. We have to determine what is and is not relevant information, and to do that we need to know what the purpose of our description is. In practice, then, descriptive studies will usually look at classrooms in relation to particular sets of criteria. These may be, for example, to compare practices of teachers at different levels of schooling, to compare practices with different target languages, to see whether there are differences between native-speaking and non-native-speaking teachers, or to see the relationship between what textbooks recommend and what actually happens in classrooms. The argument for concentration on description is that expectations of teachers, recommendations of teacher educators and

theorists, and the demands of administrators, are often rightly concerned with what ought to be. However, there is little point in constantly pushing for an ideal without any understanding of what in fact happens. We actually know remarkably little about typical practice in language learning, and there is a great need for additional comparative studies.

Interventionist studies are those in which some aspect of teaching or learning is deliberately changed, so that the effects can be monitored. Thus new materials may be introduced, new types of learning activity may be devised or used in an environment where they were not previously used, or teachers may be asked to smile more, use the target language exclusively, or participate in small group discussion. The setting is the normal one for teaching and learning, but the research monitors the effect of changes which have been deliberately introduced.

Experimental studies are similar, but usually involve a much more formal control of variables, thus stopping the classroom from being at all typical. Learners may be put into a language laboratory, or be given highly controlled tasks to ensure that the only factor that they are responding to is the one in which the researcher is interested. Experiments of this kind often operate with a 'control group' of learners who do not receive the special treatment, so that their behaviour can be contrasted with that of the 'experimental group'.

Descriptive studies are a necessary, and rather neglected, base from which discussion of innovation can proceed. However, they do not in themselves result in improvement of teaching or learning, and some researchers consider them an inefficient way of exploring theories of teaching and learning (see Cook's paper in this volume). Others (eg van Lier in this volume) would argue that the apparent rigour of interventionist and experimental studies obscures the close relationship between learning and social context, and that 'rich' descriptions of learning experiences are an essential element in building up a satisfactory theory of language acquisition. Certainly any understanding of teaching must take into account the social context, and we would ourselves accept the view that studies of language acquisition should proceed simultaneously along routes concerned with psychological and internal processing and those examining external and cultural factors.

People who wish to construct idealised models of internal processing will be more attracted to interventionist and experimental procedures. It is certainly the case that such procedures have been developed over a long period, and produced an extensive and sophisticated literature to enable researchers to avoid basic errors in claiming relationships between observations when none actually exist. Courses teaching research methods often include detailed statistical techniques for the analysis of data, and large-scale work usually calls upon the skills of statisticians in interpreting the large bodies of data collected. The risk in such work is that only what can be rigorously measured is examined. Some argue that such procedures restrict us to examining the trivial and simple, and ignoring the complex but much more important real world in

which teachers and learners have to operate. In its strong form this is a foolish debate, because we use simplification all the time in ordinary life in order to be able to cope with the mass of impressions that confront us. We are constantly making judgements about relevance and irrelevance, and interfering with random activity in order to structure a more directed and useful context for change. This, after all, is what teaching consists of, and in this sense formal research is only doing the same thing. But just as teaching can develop its own routines which lose touch with the needs of the outside world that it partly serves, so too can research procedures. A constant debate is necessary about the connection between tidy and structured activity and the more complex and untidy world it is meant to illuminate. Indeed, one of the major contributions that teachers will make to this debate is a close acquaintance with the untidiness of language learning contexts, and an awareness of the successes that can be achieved among (and even possibly because of) the theoretical untidiness.

The debate we have referred to above is often couched in terms of 'quantitative' versus 'qualitative' research. However, a careful consideration of these concepts will make it clear that they cannot really be opposed to each other. If we are examining something that can be objectively described (either numerically, or by explicit and economical records of other kinds), there is no sense in not making use of such data. On the other hand, if the questions we are interested in cannot be quantified simply, we should not avoid them solely on those grounds. We cannot limit observation to what can be measured without ignoring most of the areas that teachers and learners are interested in. It is much more important to break down our questions into those parts for which objective and measurable categories are appropriate, and those for which such categories cannot neatly be devised. As long as the status of our observations is made clear in our reporting, and as long as objectivity is achieved where that is possible, the research will be valuable.

But 'objectivity' is a relative concept. If ten witnesses agree, we will consider something more objective than if only two out of ten do. The view of someone close to the action will be preferred to that of someone a long way away—but someone too close may only see part of what is being observed, while someone a little further away may see the whole picture. Any measurement must clearly be relevant to the question being asked, and subjective impressions of a complete experience may be more valuable than objective measurement of a small and unimportant part. Ultimately, interpreting the value of particular research, and sorting out the best procedures to use, must depend on a determined common sense by practitioners.

The major options in language classroom research include the following:

1. Studies of language learning based on the observed performance of individuals in typical classes. These may be longitudinal

studies of individuals throughout their school careers, or studies of achievement of classes as groups. Much curriculum work makes assumptions about typical progress, with a very small base of knowledge to refer to.

2. Comparisons of learners' language experiences inside and outside the classroom setting.
3. Analyses of the relationship between theoretical ideas, teaching materials and syllabuses, and classroom practices.
4. Analyses of different teaching styles, relating to factors such as materials used, types of learner, size of class, etc.
5. Analyses of the beliefs of teachers about their practices and needs, related to actual classroom practice.
6. Analyses of the beliefs of learners about their practices and needs, related to their degree of success.
7. Case studies of classrooms in particular contexts, attempting to draw upon and synthesise elements of all the above.

For each of these, descriptive studies may be justified on their own, or relationships may be explored through process-product studies relating learners' progress in language learning to different aspects of the classroom learning environment. Such studies may be highly focused, exploring the impact of particular types of interaction on learning, or more comprehensive in coverage (as in Fröhlich, Spada and Allen, 1985).

Techniques used in the study of language classrooms have included:

- (a) Documentary analysis, relating (for example) textbooks or curriculum documents either to analyses of classroom practice, or to the suggestions of methodologists.
- (b) Recordings or actual classroom lessons analysed according to category systems devised for particular purposes (see Mitchell, 1985, for further discussion).
- (c) Maps of classrooms showing the distribution of (for example) teacher questioning, learner response, teacher movement, etc.
- (d) Quantified statements of types of contributions of particular learners as groups, or of individual learners over a long period.
- (e) Diaries kept by researchers over an extended period of observation, consisting of both systematic and impressionistic observation.
- (f) Lessons discussed retrospectively by observers, teachers and learners and the different perspectives analysed.
- (g) Semi-structured interviews in which a core of questions is addressed in a relaxed and improvised way by researchers talking to learners, teachers or other relevant people such as parents or administrators.
- (h) Diary studies (see the paper of Parkinson and Howell-Richardson in this volume) by learners or teachers.
- (i) Tests (of language competence or other relevant knowledge or skills) administered to learners (or teachers).



- (j) Questionnaires exploring attitudes and beliefs about different aspects of classroom experience.

The above list is far from being definitive, however; nor is the validity of the techniques used under many of the above headings as yet clearly established. At present, two of our most striking unresolved methodological problems are (a) how to identify appropriate units of analysis for classifying and categorising behaviours observed in the language classroom (linguistic and otherwise); and (b) how best to access the intentions, plans and strategies of classroom participants which underlie observed behaviour. Classroom research, if it is to remain productive, must retain a questioning attitude not only towards the objects of study, but also towards its own procedures and assumptions.

## References

- Allwright, R. L. (1988) *Observation in the Language Classroom*, Harlow: Longman.
- Chaudron, C. (1988) *Second Language Classrooms*, Cambridge: Cambridge University Press.
- Cohen, L. and Manion, L. (3rd Ed. 1989) *Research Methods in Education* (first published 1980), London: Routledge.
- Fröhlich, M., Spada, N. and Allen, J. P. B. (1985) 'Differences in the communicative orientation of L2 classrooms', *TESOL Quarterly* 19, 1: 27–57.
- Hatch, E. and Farhady, H. (1982) *Research Design and Statistics for Applied Linguistics*, Rowley Mass: Newbury House.
- Kasper, G. (ed) (1986) *Learning, Teaching and Communication in the Foreign Language Classroom*, Aarhus, Denmark: Aarhus University Press.
- Mitchell, R. (1985) 'Process research in second language classrooms', *Language Teaching* 18: 330–352.
- Mitchell, R. (1988) *Communicative Language Teaching in Practice*, London: CILT.
- Mitchell, R., Parkinson, B. and Johnstone, R. (1981) *The Foreign Language Classroom, an observational study*, Stirling: University of Stirling.
- Peck, A. (1988) *Language Teachers at Work*, Hemel Hempstead: Prentice-Hall.
- Sanderson, D. (1983) *Modern Language Teachers in Action*, York: University of York.
- van Lier, L. (1988) *The Classroom and the Language Learner*, Harlow: Longman.

# The Teacher as Researcher

David Nunan

*Macquarie University, Sydney, Australia*

## 1 Introduction

The central theme of this paper is that teachers can and should be involved in researching their own professional practices in their own classrooms, and that this implies extending the concepts of both professional practice and professional development.

In the first part of the paper, I shall present a rationale for the teacher as researcher. I shall then look at some of the issues and questions which are amenable to teacher investigation. This is followed by an examination of a range of research methods and techniques. Finally some of the problems and difficulties of the teacher as researcher will be discussed.

## 2 A rationale for the teacher as researcher

Within the teaching profession there often seems to be an insurmountable gap between theory and practice. Carr and Kemmis cite evidence purporting to show that teachers regard theory and research as esoteric activities 'having little to do with their everyday practical concerns' (Carr and Kemmis 1986: 8). In a similar vein, Beasley and Riordan write that:

... the gulf between research bodies and the teaching profession has ensured that many research programmes are not related to the professional concerns and interests of teachers and students. Priorities for research too often reflect the interests of academic researchers or central office administrators not school people. Teachers and students in the classroom are rarely actively engaged in the research. Within the experimental framework the researcher protects his or her independence for the sake of 'objectivity'. The tacit knowledge of teachers is devalued. Many of the findings are recorded in a form and style which is accessible to the trained researcher but fails to communicate to teachers, school administrators, parents or advisory people. The primary audience for research has been the research community not the practising teacher. Not surprisingly we the practising teachers have come to distrust and reject theoretical research and the researcher who takes but does not give.

(Beasley and Riordan 1981: 60)

Carr and Kemmis draw a distinction between theory, which is aimed at discovering 'truth', and practice, which is aimed at bringing about change. They suggest that the field of educational research is different from many other areas of research because education is essentially a practical rather than a theoretical activity. It is designed to bring about change and get things done. The practical nature of the educational enterprise has a number of important consequences for educational research. In the first place, this research will have different aims from research concerned with the resolution of theoretical problems. Unlike theoretical problems, these practical problems 'cannot be resolved by the discovery of new knowledge, but only by adopting some course of action' (Carr and Kemmis 1986: 108).

One way of bridging the gap between theory and practice is to encourage teachers to adopt a research orientation to their own classroom, and to engage in research projects of one sort or another. In their recent book on action research, Carr and Kemmis point to an emerging trend to broaden the scope of professional development by providing teachers with opportunities for engaging in curriculum theorising and educational research. Evidence for this trend is to be found in school-based curriculum development, research-based in-service education and professional self-evaluation projects.

Similarly, Walker (1985) sees research by teachers as a useful way of ensuring the relevance of teacher education programmes. He suggests that immersion in real research has the advantage of 'providing strong criteria of relevance', and goes on to claim that teacher education programmes with a prominent teacher research component vest much more authority in the trainees than the trainer, as it is the trainees rather than the trainer who set the agenda. They 'will be the people who have first-hand experience of the problem and its context, and the onus will be on them to communicate issues, difficulties and concerns . . .' (Walker 1985: 6).

Beasley and Riordan (1981: 36) list the following advantages of teacher-initiated research:

- It begins with and builds on the knowledge that teachers have already accumulated through research.
- It focuses on the immediate interests and concerns of classroom teachers.
- It matches the subtle organic process of classroom life.
- It builds on the 'natural' processes of evaluation and research which teachers carry out daily.
- It bridges the gap between understanding and action by merging the role of the researcher and practitioner.
- It sharpens teachers' critical awareness through observation, recording and analysis of classroom events and thus acts as a consciousness-raising exercise.

- It provides teachers with better information than they already have about what is actually happening in the classroom and why.
- It helps teachers better articulate teaching and learning processes to their colleagues and interested community members.
- It bridges the gap between theory and practice.

Walker, Beasley and Riordan are writing within a general educational context. Long (1983), however, examines the role of the second language teacher as classroom researcher. He reports on a survey which showed that classroom-centred research was comparatively neglected in MA programmes in Canada and the United States (18% of courses gave it only a minor attention while 82% ignored it altogether). In making a call for a greater role for classroom-centred research in graduate programmes, Long points to three principal roles. In the first place, despite its infancy, classroom-centred research has provided and can provide a great deal of useful information about how classes are taught (as opposed to statements about how they should be taught, or how people imagine they are taught). Secondly, classroom-centred research can promote self-monitoring by classroom practitioners. By training teachers in the use of observation schedules of one sort or another, they will be equipped to monitor their own classes or those of colleagues. The third and final reason advanced by Long for involving teachers in classroom-centred research is that it can help teachers resist the temptation to jump on each language-teaching bandwagon as it rolls past. Straightforward, descriptive studies of what actually does go on in classrooms will help teachers evaluate the competing claims of different materials, syllabuses and methods.

In a slightly different vein, Freeman (1987) outlines four instructional patterns in language teacher education. These are instruction, observation/imitation, guided reflection and self-directed reflection.

In instruction, the trainee is told how to do something. This transmission approach to teacher education characterised an earlier prescriptive approach to teacher education. Observation involves trainees in studying live or recorded lessons and lesson segments. The lessons serve as models which the trainees can attempt to imitate. Guided reflection activities engage trainees in identifying, analysing and hypothesising from their own experience. The final instructional pattern, self-directed reflection, involves trainees in analysing and interpreting experience without external assistance.

These instructional patterns represent a continuum from dependence to independence. Any professional development programme which has as one of its goals the development of professional autonomy will attempt to move the trainee along the continuum from observation, through guided reflection to self-directed reflection. In this paper, I should like to suggest that pre- and in-service programmes aimed at developing the concept of the teacher as researcher will be critically concerned with the notion of professional autonomy, with all that this entails.

### 3 Issues for investigation

In this section we shall look at some of the issues and associated problems which might prove amenable to research. Any element in the curriculum (including teachers, learners, materials/resources, learning tasks, learning styles etc.) can be investigated by the teacher. This is demonstrated in Table 1. The data here were obtained in a workshop on action research in which the teachers were asked to nominate an issue which they intended to follow up as a result of the workshop.

**Table 1 Issues nominated by teachers for investigation (adapted from Richards and Nunan 1989)**

<i>Area</i>	<i>Issue</i>
1 Methodology	Task analysis and the different demands that tasks create. What materials/methods learners do/do not respond to. The learning and teaching of vocabulary.
2 Classroom management and interaction	The occurrence of digressions within a lesson by teachers and students and the extent to which these lead to useful learning outcomes or simply distract, confuse or mislead students. The management of classroom interactions. Effective and ineffective instruction giving. How to increase student talking time. Do students think this is valuable? Does it enhance learning?
3 Professional development and self-evaluation	How do teachers perceive peer analysis? In what ways is it helpful, threatening, inhibiting? How action research can improve cohesion/sense of progression from the students' perspective. Peer teaching/learning for teachers. Promoting personal responsibility for professional development. Using classroom analysis with new teachers to assist them develop their own practices more effectively.
4 Applying skills	Encouraging and monitoring students' use of English outside the classroom. Encouraging the use of English outside the classroom.
5 Affective factors	Student attitudes towards games and drama activities. Student perceptions of language learning.
6 Evaluation and assessment	Evaluating the effectiveness of teaching. Methods of post-learning arrangement assessment.

---

How to develop classroom tests for end of course assessment.

7 Morphosyntax                      Whether plateaux in language learning really exist.

---

I shall confine myself here to issues and problems relating to teachers themselves. I shall suggest that teachers can carry out studies relating to themselves working alone or in collaboration with other teachers. In some circumstances, they can also invite other teachers to collect data on them. The number of questions and issues which might be investigated are almost limitless, and here I shall only be able to present a very few as samples. They can relate to any aspect of teachers' work as they plan, implement and evaluate their programmes.

**Table 2 Issues and sample investigative questions on teachers**

<i>Issue</i>	<i>Investigative questions</i>
Planning	What are the bases on which I select my goals and objectives? What are the major factors I take into consideration when selecting content?
Implementation	What is the relationship between the lesson plans I draw up before class and what actually happens in class? To what extent does my teaching reflect a systematic procedure of specifying objectives, selecting content and learning tasks and evaluating the effectiveness of instruction? What events in the classroom cause me to deviate from my planned lesson?
Classroom management	Some of my learners are disruptive. Is there anything in my behaviour towards them which might account for their behaviour? What effect will modified behaviour on my part have on them? What aspects of learner behaviour do I respond to? How efficient/effective am I at setting up group work?
Talk	How much talking do I do in class? Is this too little or too much? What happens when I alter the amount of talking I do? How clear and/or useful are the explanations I give to my students? What sort of questions do I ask?
Interaction	What typical patterns of interaction are there between myself and my learners? Do I interact with some learners more than others? What happens if I modify my patterns of interaction?

---

The area of teacher talk has been heavily researched in recent years, and in the rest of this section, I shall, for illustrative purposes, review the work which has been done in this particular area, as it is one which is particularly amenable to teacher research.

One aspect of teacher talk which has been intensively studied is the speech modifications made by teachers. A great deal of research has shown that native speakers (teachers and non-teachers alike) simplify their speech to non-natives. It has been suggested that these modifications make the language easier to comprehend, and that this, in turn, should facilitate acquisition. In his review of teacher talk, Chaudron (1988) discusses the amount of teacher talk, the functions of teacher talk and the nature of teachers' explanations.

Pica and Long (1986) investigated the linguistic and conversational performance of experienced and inexperienced teachers. The purpose of the research was to determine whether there was any difference between the speech characteristics of experienced and inexperienced teachers in their classroom interactions. The features they studied included linguistic complexity, the functions of questions, statements and imperatives, question types, and certain conversational adjustments such as comprehension checks and requests for clarification which are hypothesised to facilitate comprehension and therefore acquisition.

Pica and Long found that there were some differences between the experienced and inexperienced teachers (for example, experienced teachers used a wider range of question forms). In general, however, the similarities outweighed the differences.

Not surprisingly, research in language classrooms has shown that teachers tend to do most of the talking. Whether this is a good thing or not will depend on what one believes about the role of input in acquisition. If one believes that learners learn best by doing (ie in this case by practising in the target language), one will probably try to structure classroom activities so that the amount of learner talk is increased at the expense of teacher talk. If, on the other hand, one believes that teacher talk is a valuable source of comprehensible input, one will be much less worried by the quantity of teacher talk. Teachers who record and analyse their classes generally find that they do far more talking than they had imagined. In a recent workshop teachers made the following discoveries about their classroom talk.

- 'I praise but it is rather automatic.'
- 'There was a lot of teacher talk.'
- 'I give too many instructions.'
- 'I need to give clearer instructions.'
- 'There was excessive teacher instruction.'
- 'There was too much teacher talk.'
- 'I need to do more comprehension checking.'
- 'I need to do more eliciting.'

- 'I had tried to avoid the comprehension question list, but ended up asking just as many questions of my own.'
- 'The instructions I gave were unclear.'
- 'Some explanations were nonsensical.'
- 'I gave no praise.'
- 'Instructions and explanations were inadequate.'
- 'Explanations to the class were confusing.'
- 'There were excessive teacher digressions.'

As we can see from these reactions, particular areas of concern are the amount of talking teachers do, and the quality of such talk. In particular, the teachers were dissatisfied with the explanations and instructions they gave in the lessons which they analysed.

One aspect of teacher talk which has received a lot of attention is that of teacher questions. Long and Sato (1983) compared the questions asked by teachers in class with the type of questions used by native speakers when communicating with ESL speakers in non-classroom contexts. They discovered a radical difference in the types of questions asked. In class, questions were overwhelmingly of the display type (that is, questions to which the teacher already knew the answer—'Is this a book?') In non-classroom contexts, virtually all questions were referential (ie the questioner did not know the answer to the question in advance). In a follow-up study, Long and Crookes (1986) trained teachers to ask more referential questions. They found that such training did result in more referential questions being asked in class. However, the use of referential rather than display questions produced rather mixed findings. While display questions actually elicited more student turns, referential questions prompted longer and more extensive responses. There was also a trend suggesting that mastery of lesson content was greater in classes in which teachers made extensive use of referential questions.

Another area of interest is feedback from teachers to learners. Chaudron (1988) points out that feedback in general and error feedback or correction in particular have been widely investigated outside the classroom. In the language class,

. . . the primary role of language teachers is often considered to be the provision of both error correction, a form of negative feedback, and positive sanctions or approval of learners' production. (Chaudron 1988: 132)

From this brief survey, it can be seen that looking at one area only, that of teacher talk, there are many issues and problems which might be investigated. Many of these would be suitable for classroom teachers to investigate in relation to their own classrooms. They can be summarised as follows:



- the amount of teacher talk as a ratio of total talk time;
- the modifications and adjustments to teacher talk eg syntax, vocabulary, rate;
- repetition;
- the range and types of questions;
- the range and types of instructions and explanations;
- the type of error feedback.

#### 4 Research methods

In this section, I shall discuss some of the methods which might be employed in classroom research.

In his book on classroom research, Chaudron outlines four basic research traditions. These are summarised in Table 3.

**Table 3 A comparison between different research traditions, issues and methods in the language classroom (after Chaudron 1988).**

<i>Tradition</i>	<i>Typical issues</i>	<i>Methods</i>
Psychometric	Language gain from different methods, materials, treatments	Experimental method — pre- and post-tests with experimental and control groups
Interaction analysis	Extent to which learner behaviour is a function of teacher determined interaction	Coding classroom interactions in terms of various observation systems and schedules
Discourse analysis	Analysis of classroom discourse in linguistic terms	Study classroom transcripts and assign utterances to pre-determined categories
Ethnographic	Obtain insights into the classroom as a 'cultural' system	Naturalistic 'uncontrolled' observation and description

Each of these different methods will yield a different 'snapshot' of the classroom in action. Consider the following excerpts, each of which provides an analysis of the same interaction.

**Excerpt 1 Interaction Analysis Categories (adapted from Flanders)**

	Tallies	Total
1 Teacher asks a display question (ie a question to which she knows the answer).		
2 Teacher asks a referential question (ie a question to which she does not know the answer).		
3 Teacher explains a grammatical point.		
4 Teacher explains meaning of a vocabulary item.		
5 Teacher explains functional point.		
6 Teacher explains point relating to the content (theme/topic) of the lesson.		
7 Teacher gives instructions/directions		
8 Teacher praises.		
9 Teacher criticises.		
10 Learner asks a question.		
11 Learner answers question.		
12 Learner talks to another learner.		
13 Period of silence or confusion.		

**Excerpt 2 Bowers' Categories of Verbal Behaviour in the Classroom**

*Responding*: any act directly sought by the utterance of another speaker, such as answering a question

*Sociating*: any act not contributing directly to the teaching/learning task, but rather to the establishment or maintenance of interpersonal relationships

*Organising*: any act which serves to structure the learning task or environment without contributing to the teaching/learning task itself

*Directing*: any act encouraging non-verbal activity as an integral part of the teaching/learning process

*Presenting*: any act presenting information of direct relevance to the learning task

*Evaluating*: any act which rates another verbal act positively or negatively

*Eliciting*: any act designed to produce a verbal response from another person

- T: of course i had lunch . . . not enough . . . why? why? (*sociating*)  
 well, like i say, i want to give you something to read (*organising*)  
 – so what you do is, you have to imagine what comes in between, that’s all . . .  
 (*organising*)  
 . . . bring, er, bring your chairs a little closer, you’re too far away, er, ha, not  
 that close (*organising*)
- S: quiss? (*eliciting*)
- T: pardon? (*responding*)
- S: it will be quiss? it will be quiss? quiss? (*eliciting*)
- Ss: quiss . . . quiss (*eliciting*)
- T: ahm, sorry . . . try again (*eliciting*)
- S: i ask you . . . (*eliciting*)
- T: yes?
- S: you give us another quiss? (*eliciting*)
- T: oh, quizz, oh! no, no, not today . . . it’s not going to be a quiss today . . . sorry . . .  
 but, um, what’s today, tuesday is it? (*eliciting*)
- S: yes (*responding*)
- T: i think on thursday, if you like . . . same one as before . . . only i’ll think up some  
 new questions – the other ones were too easy . . . um, ok, er i’ll take some  
 questions from, er, from newspapers over the last few weeks, right? so – means  
 you’ve got to watch the news and read the newspaper and remember what’s  
 going on . . . if you do, you’ll win . . . if not, well, that’s life (*organising*)
- S: will be better from tv (*sociating*)  
 (laughter)
- T: from the tv? . . . what, er, what programmes . . . (*eliciting*)
- Ss: news, news (*responding*)
- T: did you say . . .? oh, ok, we’ll have, er, it’ll be the s . . . , it’ll be the same . . .  
 there’ll be different . . .? er, there’ll be different . . .? different? different? the  
 questions will be on different . . . what? different? (*eliciting*)
- S: talks (*responding*)
- T: tasks? (*evaluating*)  
 what? (*eliciting*)
- S: subject? (*responding*)
- T: different sub. . . (*eliciting*)
- S: subjects (*responding*)
- T: subjects, subjects, thank you . . . right, yes (*evaluating*)

### Excerpt 3 Narrative

The teacher enters the classroom in conversation with one of the students. ‘Of course I had lunch’, he says. ‘Not enough. Why? Why?’

The student gives an inaudible response, and joins the rest of the class who are sitting in a semicircle. There are eighteen students, in all. They are a mixed group in both age and ethnicity.

The teacher deposits three portable cassette players on his table, and slumps in his chair. ‘Well, like I say, I want to give you something to read – so what you do is, you have to imagine what comes in between, that’s all . . .’ He breaks off rather abruptly and beckons with his hand, ‘. . . Bring, er, bring your chairs a little closer, you’re too far away.’ There is some shuffling as most of the students bring their chairs closer. The teacher halts them by putting his hand up, policemanwise. ‘Er, ha, not that close.’ There is some muffled laughter. The

teacher is about to speak again, when a young male student breaks in with a single utterance 'Quiss?' The teacher gives him a quizzical look.

'Pardon?'

The student mutters inaudibly to himself and then says, 'It will be quiss? It will be quiss? Quiss?.' Several other students echo, 'Quiss. Quiss.'

The teacher grins and shakes his head, 'Ahm, sorry. Try again.' The student frowns in concentration and says, 'I ask you . . .' 'Yes?' interjects the teacher. ' . . . You give us another quiss?'

Slowly the light dawns on the teacher's face. 'Oh, quizz, oh! No, no, not today. It's not going to be a quiss today. Sorry . . . But, um, what's today, Tuesday is it?'

'Yes', says the student.

The teacher frowns and flicks through a notebook on his desk. 'I think on Thursday, if you like. Same one as before. Only I'll think up some new questions – the other ones were too easy.' The students laugh, then the teacher, holding up the daily newspaper, continues, 'Um, OK, er I'll take some questions from, er, from newspapers over the last few weeks, right?' So – means you've got to watch the news and read the newspaper and remember what's going on. If you do, you'll win. If not, well, that's life.'

One of the woman students, a Pole in her early thirties says, 'Will be better from TV.'

There is laughter from several of the students.

'From the TV?' echoes the teacher. 'What, er, what programmes . . .'

'News, news,' interject several of the students.

There is an inaudible comment from one of the students. The teacher turns sharply and begins, 'Did you say . . .?' He breaks off abruptly. 'Oh, OK. We'll have, er, it'll be the s . . ., it'll be the same.' He pauses and then adopts an instructional tone, as he attempts to elicit a response from the students.

'There'll be different . . .? Er, there'll be different . . .? Different? Different? The questions will be on different . . . what? Different?'

'Talks,' ventures one of the students near the front.

'Tasks? What?' says the teacher giving a slight frown.

'Subject?' suggests another student rather tentatively. The teacher gives her an encouraging look and says, 'Different sub. . .' He extends his hand and narrows his fingers as if to say 'You've nearly got it.'

'Subjects,' says the student, beaming.

The teacher beams back, 'Subjects, subjects, thank you. Right, yes.'

Each of these excerpts provides us with a different view of the particular lesson under investigation. From the first, we obtain some idea of the amount and type of teacher and learner talk. Without looking at a transcript, we can infer that the excerpt is taken from the beginning of the lesson. Clues here are the number of instructions and directions which are given, the fact that more referential than display questions are asked, and the fact that there are several learner-initiated questions. These are all features which tend to be comparatively rare in the body of an L2 lesson.

The second and third excerpts provide much more detail. Teachers generally report that Excerpt 3 is the most 'user-friendly'. While we get much more information from transcripts such as these, they are

extremely tedious and time-consuming to produce. In addition, while Excerpt 3 might be the most readable, the authorial intrusions may not accurately reflect the tone and mood of the lesson.

Acheson and Gall (1987) describe four techniques for classroom observation. These are selective verbatim, seating charts, wide-lens techniques and the use of checklists and timelines.

Selective verbatim consists of the recording and analysis of selected aspects of verbal interaction. The recording can be done either by a teacher trainer, colleague, or, using audio or video recordings, by the trainees themselves. This method is appropriate for exploring issues such as teacher questions, feedback, directions and structuring statements.

Verbatims can also take the form of diaries and journals. Bailey (1989), who has written extensively on the use of such records in second language teaching and learning, suggests that journals can help us capture some of the complexities of the language classroom, allowing us to look at the cognitive, interactive and affective variables in play. The interaction between such variables might be screened out of more controlled research methods.

One of the problems with free-form diary and journal entries is that it is often difficult to make generalisations from them. Beretta (1989) has tackled the problems associated with the subjectivity of discursive accounts by utilising a levels-of-use method first developed by Loucks and Hall (1977). This allowed him to render his classification of teachers accessible to public scrutiny.

Observational records based on seating charts record physical movements as well as various types of verbal interaction between teachers and learners. They can condense a large amount of information and are easy to construct, use and interpret. Seating charts have been used to document the amount of time on task spent by learners, showing verbal flow diagrammatically (ie indicating who is talking to whom), and documenting movement patterns.

The term 'wide-lens techniques' refers to the free-form notes made by an observer about a class. These form a written 'protocol' or record which can reveal interesting aspects of classroom behaviours which might be screened out by more selective analytical techniques. These direct free-form observations are similar to naturalistic observations made by ethnographers. They can take the form of discursive notes made by an observer, or can take the shape of audio or video recordings.

The fourth category, the use of checklists or timeline coding charts, is probably most familiar to language teacher educators, as a wide variety of schemes have been developed over the last few years. Checklists can focus on almost any aspect of classroom interaction. As we have seen, some coding devices reduce the interaction to a series of rhetorical acts eg 'teacher questions', 'student responds'. A distinction is sometimes drawn between timeline schedules in which observations are made at regular intervals eg every ten seconds, and those which simply record behaviour.

In addition to the usual observation schedules for recording oral

interactions, there are various other instruments teachers can use for recording their instructional behaviour. For example, Koziol and Call (1988) have developed self-report inventories which teachers can use to record a wide range of instructional behaviours. One of the problems with most such inventories, in which teachers record the frequencies of various behaviours, is that they are often couched in vague terms. (How regular is 'regularly'? How frequently is 'frequent'?)

Koziol and Call have attempted to overcome this problem by defining behaviour frequencies in numerical terms. A sample schedule relating to the teaching of grammar is included as an appendix to this paper.

## **5 Some problems and difficulties with the concept of teacher/researcher**

Thus far, I may have given the impression that the notion of teacher as investigator is relatively problem-free. In fact, nothing could be further from the truth, and in this section, I should like to outline some of the problems which are likely to occur.

One fundamental problem is the insular nature of most classrooms, and the tendency of teachers to be rather inward-looking. It is surprising how few teachers actually ever get to watch another teacher teaching, or allow themselves to be observed. This insular climate in which the classroom takes on the nature of a 'black box' (Long 1980) leads to reluctance and often resistance on the part of teachers when it comes to allowing others to sit in on classes, or even to review audio or videotaped excerpts from lessons. While not all of the methods and issues for investigation outlined here require collaboration between a teacher and colleague, trainer or outside researcher, many do. In fact, Kemmis and McTaggart make a strong case for collaboration in classroom research precisely because it does open the investigation to critical scrutiny. I have found that in in-service programmes in which classroom data from participants is used, it is generally necessary before the programme begins to assure participants that confidentiality will be preserved, and that no other participant will have access to their data if they do not wish.

There is also a tendency on the part of teachers to excessive self-criticism when first engaging in the recording and analysis of their own classes. This can be seen in the comments by teachers reported in Section 3. It is therefore necessary to reassure teachers that they are not necessarily as bad as their audio or videotapes would seem to indicate, and that the tendency to be overly critical is a natural one. It is also important, if one uses videotapes of other teachers, that these be genuine classroom extracts rather than idealised models of perfect practice. The artificial, heavily-rehearsed lessons which sometimes form the basis for commercially-made videos can have a demoralising effect on teachers when they compare this teaching to their own.

A problem which can arise at the stage at which teachers are formulating a research proposal is that the proposals themselves often

tend to be rather grand. It is highly unlikely that a classroom teacher will have the time or the research skills to conduct complicated psychometric investigations with pre- and post-treatment tests, randomly assigned experimental and control groups, controlled variables etc. Nor is it likely that they will have the wherewithal for investigating complex issues such as morphosyntactic development. It is therefore important to ensure that the investigation is realistic in terms of the issue or problem selected, the research method and the resources available.

Resources are, in fact, another critical problem. Properly formulated projects are extremely time consuming. The teacher-developer or support person usually needs to work on a one-to-one basis with the classroom teacher — often over the life of the project. This has implications, not only in terms of appropriate personnel being available, but also in terms of time and money.

In one school-based ESL programme in Australia an attempt is being made to develop procedures for providing local support to new teachers. The aim of the project is for experienced teachers with training in curriculum development to be available within teaching centres (rather than making sporadic visits from a centralised curriculum support unit) to observe, consult, and co-teach with new teachers. Such curriculum support personnel need a range of skills. These are summarised below as they are the sorts of skills which research project supervisors or support personnel might need. Skills and knowledge include:

- a sound theoretical background;
- familiarity with a range of materials and their application to a range of learner types;
- specific time allocations for working with new teachers;
- opportunities to talk to other teachers when planning and implementing support strategies;
- the ability to create a non-threatening relationship with the new teacher;
- the ability to assess the extent of the new teacher's ESL awareness;
- the ability to draw conclusions from observation;
- the ability to make decisions about the most immediate areas of need and how to tackle them;
- the ability to perceive opportunities where 'intervention' is appropriate;
- the ability to provide constructive feedback to the newly-appointed teacher;
- the ability to reflect on the overall course with the newly-appointed teacher.

(Adapted from Solomon 1987)

A further problem comes with the completion of the research project. Key questions here are: Is the study to be documented? If so, how is the study to be written up and reported on? Who is the principal audience of the

report? Once again, teacher sensitivities need to be taken into consideration at this final point in the life of a project.

Finally, there are a number of key questions relating to the whole range and scope of teacher research. These have been well summarised by Walker and are included here as they echo and reinforce most of the points which have already been made:

1. Does one begin by immersing teachers in research projects, providing instruction in research methods only when this is requested or seems necessary, or does one begin by providing instruction in research methods followed by gradual practice in actual cases?
2. To what extent should research projects be collaborative or individual exercises?
3. Should projects focus on the peculiarities of specific situations, or be situated in relation to generalised, propositional knowledge (ie should the concern be with cases or samples)?
4. Should one prioritise short-term needs or long-term values?
5. How does one identify the appropriate audiences for the research?

## **6 Looking ahead**

In recent years there has been a change of attitude in both pre- and in-service teacher education. This change in attitude has been marked by a move away from prescription, towards a view of the teacher as an autonomously functioning individual. Programmes concerned to develop autonomy and self-direction will have as their central focus the development of a critical self-awareness on the part of teachers towards their own teaching and the teaching of others. Rather than working exhaustively through a set of prescriptive edicts, or learning teaching routines based on pre-packaged methods such as Suggestopedia, Silent Way, and the Natural Approach, teachers will develop an extensive repertoire of classroom management and teaching techniques. These will be linked in principled ways to beliefs about the nature of language and the nature of learning. Teachers will learn how, when and why to utilise particular approaches, techniques etc, not by following the dictates of a method, but through the close study of the classroom itself.

In this paper, I have tried to show some of the ways in which teachers can develop a critical self-awareness towards themselves, their peers and their classrooms by adopting a research orientation to their work. I have exemplified the process in terms both of issues and research techniques which might be employed when teachers turn a critical eye upon themselves. I have also tried to spell out some of the complexities and difficulties associated with the adoption of such a research orientation.



**References**

Acheson, K. and Gall, M. D. (1987) *Techniques in the Clinical Supervision of Teachers*, New York: Longman.

Bailey, K. (1989) 'Using diaries in teacher development', in J. Richards and D. Nunan (eds) *Second Language Teacher Education*.

Beasley, B. and Riordan, L. (1981) 'The classroom teacher as researcher', *English in Australia*, 55.

Beretta, A. (1989) 'The Bangalore Project: Levels of implementation', in J. Richards and D. Nunan (eds) *Second Language Teacher Education*.

Carr, W. and Kemmis, S. (1985) *Becoming Critical: Knowing Through Action Research*, Victoria: Deakin University Press.

Chaudron, C. (1987b) 'The interaction of quantitative and qualitative approaches to research: a view of the second language classroom', *TESOL Quarterly*, 20, 4, 709-717.

Chaudron, C. (1988) *Second Language Classrooms: Research on Teaching and Learning*, Cambridge: Cambridge University Press.

Freeman, D. (1987) 'Four instructional patterns in language teacher education', paper presented at the Twenty-First Annual TESOL Convention, Miami Beach, March 1987.

Koziol, S. and Call, M. (1988) 'Constructing and using teacher self-report inventories', workshop presented at the Twenty-Second Annual TESOL Convention, Chicago, March 1988.

Long, M. (1980) 'Inside the "black box": Methodological issues in classroom research on language learning', *Language Learning*, 30, 1-42.

Long, M. (1983) 'Training the second language teacher as a classroom researcher', paper presented at the 34th Annual Round Table on Languages and Linguistics, Washington DC: Georgetown University.

Hall, G. E. and Loucks, S. F. (1977) 'A developmental model for determining whether the treatment is actually implemented', *American Educational Research Journal*, 14, 3, 263-276. (Cited in Beretta, 1989.)

Nunan, D. (1987a) *The Teacher as Curriculum Developer*, Adelaide: National Curriculum Resource Centre.

Nunan, D. (1987b) 'Does instruction make a difference? Revisited', *TESOL Quarterly*, 21, 1, 372-377.

Nunan, D. (1988) *The Learner-Centred Curriculum*, Cambridge: Cambridge University Press.

Pica, T. and Long, M. (1986) 'The linguistic and conversational performance of experienced and inexperienced teachers', in Day, R. (ed). *Talking to Learn*, Rowley, Ma. Newbury House.

Richards, J. and Nunan, D. (eds). (forthcoming) *Second Language Teacher Education*.

Solomon, N. (1987) *New teachers: Guidelines for local support*, Sydney: NSW Adult Migrant Education Service.

Stenhouse, L. (1975) *An Introduction to Curriculum Research and Development*, London: Heinemann.

Walker, R. (1985) *Doing Research: A Handbook for Teachers*, London: Methuen.

**Appendix**

Name \_\_\_\_\_ Week of \_\_\_\_\_

Class (language) \_\_\_\_\_ Level \_\_\_\_\_ Period \_\_\_\_\_

This is an inventory that asks you to identify how many times you used a given teaching practice in a particular class in a given week. Please use this key in responding to the following statements relating to different aspects of grammar presentations.

- |                  |  |
|------------------|--|
| 0 = Never        | This is something that I did not do in this particular class this week.  |
| 1 = Infrequently | This is something that I did once this week in this class.               |
| 2 = Sometimes    | This is something that I did two or three times this week in this class. |

3 = Regularly

This is something that I did four or five times this week in this class.

In presenting a grammar teaching point for the first time I:

- \_\_\_\_\_ 1. presented the teaching point both orally and with visual aids.
- \_\_\_\_\_ 2. used pictures and diagrams to convey the meaning of the teaching point.
- \_\_\_\_\_ 3. presented the teaching point indirectly in the context of spoken language, but did not formally teach it.
- \_\_\_\_\_ 4. presented the teaching point indirectly in the context of written language, but did not formally teach it.
- \_\_\_\_\_ 5. presented the teaching point indirectly in the context of spoken language and pointed it out to the students.
- \_\_\_\_\_ 6. presented the teaching point indirectly in the context of written language and pointed it out to the students.
- \_\_\_\_\_ 7. presented the teaching point using only the target language.
- \_\_\_\_\_ 8. reviewed with the students relevant, previously-presented grammatical structures.
- \_\_\_\_\_ 9. gave the students several examples of the teaching point, and guided them in discovering the grammatical rule.
- \_\_\_\_\_ 10. gave the students several examples of the teaching point, before supplying them with the grammatical rule.
- \_\_\_\_\_ 11. translated examples of the teaching point to be certain that the students understood.
- \_\_\_\_\_ 12. assisted the students in participating in a target language conversation, then drew the teaching point from the language that the students themselves had generated.
- \_\_\_\_\_ 13. spoke only in the target language, but modified the structure, vocabulary, and speed so that the students could understand easily.
- \_\_\_\_\_ 14. did not focus on grammar in the teaching of language.
- \_\_\_\_\_ 15. based new teaching points on previously-presented grammatical structure.
- \_\_\_\_\_ 16. gave only one example of the teaching point and did it orally.
- \_\_\_\_\_ 17. embedded the teaching point in a command designed to elicit a non-verbal response from the students.
- \_\_\_\_\_ 18. relied on gestures and mime to convey the meaning of the teaching point.
- \_\_\_\_\_ 19. drew the teaching point from dialogue that the students had memorised.
- \_\_\_\_\_ 20. explained the teaching point in English.
- \_\_\_\_\_ 21. conducted oral drills on the teaching point before presenting it formally.
- \_\_\_\_\_ 22. wrote the grammatical rule on the board/overhead before beginning to explain it.
- \_\_\_\_\_ 23. gave the students the general grammatical rule, then wrote examples of the rule on the board/overhead.
- \_\_\_\_\_ 24. allowed students to look at the explanation in their textbooks while I was presenting the teaching point.
- \_\_\_\_\_ 25. had the students read a grammar explanation in their texts before I presented it in class.

# Ethnography: Bandid, Bandwagon, or Contraband?

Leo van Lier

*Monterey Institute of International Studies*

## Introduction

Much has been written in recent years about ethnography and its actual or potential uses in education. Rather than reviewing the theoretical arguments and describing the methodological options in detail I refer the reader to some of the key surveys (Cazden 1985, Erickson 1985, Hymes 1981, Watson-Gegeo 1988)<sup>1</sup>.

In this paper I first want to place ethnography in the context of scientific enquiry in general and then focus on its current status in SLA research. Finally I will speculate on various ways in which the application of ethnography can be fruitful in teaching praxis, teacher development, and the language learning process. Overall, then, I aim to present a top-down dissection of ethnography, from its theoretical underpinnings to its practical uses for the classroom teacher and learner.

## Ways of doing scientific research

Just the place for a Snark! I have said it twice:  
That alone should encourage the crew.

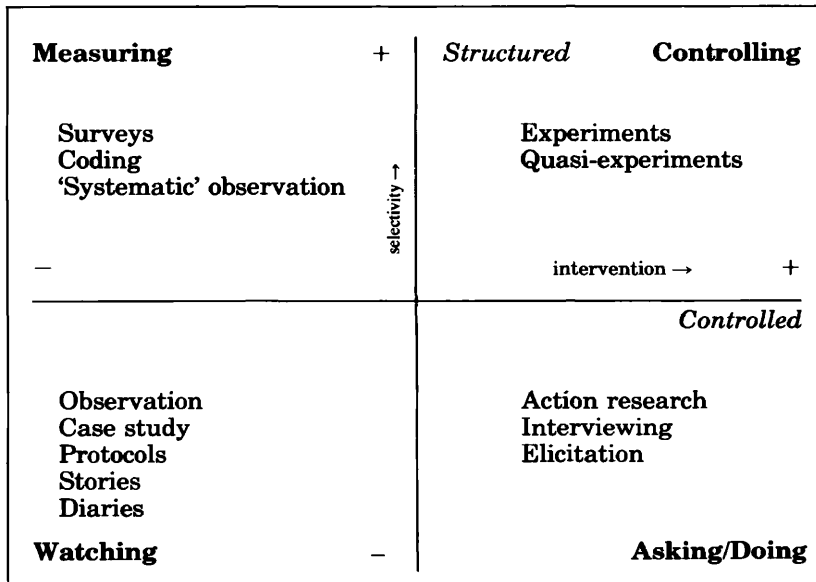
Just the place for a Snark! I have said it thrice:  
What I tell you three times is true.

(Lewis Carroll: *The Hunting of the Snark*)

It is useful to discuss choices in the way empirical research is conducted in terms of two parameters: the degree to which the researcher *intervenes* in the environment, and the degree to which phenomena are *selected* (or isolated) from the context in which they occur. The intersection of these two parameters creates four semantic spaces, as illustrated in Fig. 1, and briefly described on page 34.

---

<sup>1</sup> For two book-length introductions to ethnography, see Hammersley and Atkinson 1983, Saville-Troike 1982; for edited collections of papers illustrating ethnographic work in schools and classrooms, see Adelman 1981, Gilmore and Glatthorn 1982, Trueba, Guthrie and Au 1981, Trueba 1987.



*Fig. 1* Types of research. From van Lier 1988: 57.

### *I Measuring*

One selects certain features, operationally defines them, and quantifies their occurrence, in order to establish a relationship between features, or between features and other things, such as educational outcomes.

### *II Controlling*

One controls significant variables in the environment in order to study selected features in that environment. If features cannot be directly controlled, they can be statistically neutralised by means of randomisation. Once all variables are under control, or are randomised, true experiments can be conducted. A lesser degree of control, partly compensated for by means of statistical tests, leads to various kinds of compromise designs, commonly referred to as quasi-experimental studies.

### *III Watching*

One observes and documents whatever happens in an environment without interfering with that environment. One may progressively focus on selected features in the environment, and thus move from III to I. Quantification may be used, but it is seen as no more than one tool among many, and not inherently superior to any other way of analysing data.

### *IV Asking/doing*

One investigates certain problem areas by probing, trying out minor changes, asking for participants' views and concerns, and so on. After a while it may be possible to pinpoint the problem so precisely that a

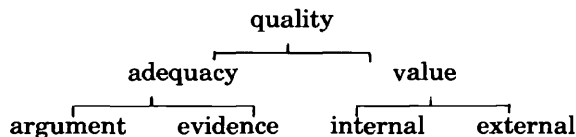
controlled environment can be created in order to conduct an experiment, thus moving from IV to II. On the other hand, increased understanding through interpretation can also make experimentation unnecessary.

Clearly, most research does not fall neatly into one of the four boxes created here. Rather, the researcher combines different features according to a particular *research design* or just in response to problems and possibilities, constraints and resources, since research is in practice 'messy' rather than 'neat'. However, researchers tend to have certain preferences which are in part based on assumptions of value, quality and truth. For example, one of the most prominent assumptions among researchers is that the ultimate aim of research is to find laws of cause and effect (so-called 'natural laws'). In this view, truth (that which has been proven without doubt to accord with facts) is causal truth, and the value of a particular research effort increases as it comes closer to establishing causal relationships. The quality of research is measured in terms of adherence to methods which promote the search for causal truth. Thus, a range of assumptions which are tacitly accepted by a vast majority of producers and consumers of research, is based on a particular definition of truth as *that which expresses a causal law*.

We must not forget another assumption inherent in the above classification, namely that research must essentially be empirical. This view does not take into account the potential fruitfulness of more speculative forms of enquiry, such as the hypothetico-deductive method of research. In particular, scientists are increasingly aware that in scientific discovery interpretation and experimentation go hand in hand. Painstaking empirical studies are most fruitful and cumulative when they are steered by theoretical positions which suggest that a certain direction is a useful one. In that sense, therefore, theory, which is always in part speculative and intuitive, guides both interpretive and normative enquiry.

In order to understand the discussions between ethnographers and cause/effect researchers, it is necessary to explore the notions of truth and understanding, cause and effect, however philosophical and abstract these notions may seem. Failing to do so places all arguments on a superficial and trivial plane. In the next section, therefore, I will look at the notion of *research quality*<sup>2</sup> in terms of the basic ideas of truth and

2 The concept of *quality* as a superordinate concept is crucial. It encompasses both reliability and validity. However, since these latter terms are associated with experimental and statistical norms, I prefer to use the terms *adequacy* (of argumentation and evidence) and *value* (within a theory, ie internal, and to human affairs in general, ie external), as shown in the diagram below (see Ravetz 1971).



understanding. I do not presume to have any superior knowledge of these matters, but I do stress that it is essential that we think about them and discuss them intensely.

## Truth and causality

We feel that even when all *possible* scientific questions have been answered, the problems of life remain completely untouched. (Wittgenstein 1961: 73)

Much of classroom research is based on a belief that everything that happens has a cause. One of the things that happens in the world is that people learn second languages. The belief in causes asserts that this event-series, whenever it occurs, is caused by a set of conditions. If we can find out what these conditions are, then we can place people who want to or need to learn second languages in these conditions, and the desired result will occur.

The belief that everything that happens has a cause is known as the Causal Principle. To paraphrase Hospers (1967: 308), the Causal Principle states that for every event in the universe there is a set of conditions such that if the conditions, C, are all fulfilled, the event, E, invariably occurs (or, in a weaker version, E 'will probably occur').

But what if the Causal Principle is false? Or if L2 development is only partly caused by antecedent conditions? In these cases, very clearly, research based on causal models of science is certainly incomplete and possibly misguided. We will need to look for other ways of studying L2 learning, ways which are not based on a piecemeal sifting through all the conditions which may be relevant, but which simply try to understand the circumstances in which complex processes take place.

Supporters of the Causal Principle will object that there is no reason to believe that L2 learning does not fall under the domain of causal events, and that in fact we cannot know this until we have proved it. This is true, but of course the same reasoning would require that the causality of events can also be *disproved*, and this can never be done. If we find causes, then indeed the Causal Principle is proved. However, if we don't find causes, the Principle is *not disproved*. We may say, for example, that there are causes but that we haven't been able to find them yet. Or we may say that the causal conditions are so complex that it will take a very long time to find them. So, although the principle of causality must be confirmed by empirical evidence, there is no empirical evidence which requires us to abandon it (thus, causal studies fail the 'falsifiability requirement' of Popper, see eg Phillips 1987).

If it is assumed *a priori* that L2 learning is caused by certain *sufficient* conditions, the researcher's job is to circumscribe those conditions so that, whenever they obtain, the occurrence of L2 learning can be accurately *predicted*. This is an extremely complex task at best, but it may also be an imperfect and insufficient route to knowledge and understanding. This

can be illustrated by taking a very straightforward and simple class of events: trees blowing down. It is clear that not every time the wind blows against a tree, that tree will fall down. When we study the phenomenon, we must add qualifications and amendments which are endless: the wind must blow hard enough, the tree must be fragile enough, the roots must grip the soil insufficiently, the soil must be loose enough, etc. In addition, we must take into account the position of the tree among buildings, other trees, and so on. It would probably be impossible to lay down all the conditions which would ensure a guaranteed tree-falling-down event. So, even if we are able to say: 'The wind caused that tree to fall down', we are still not able to specify exactly what it will take for another tree, say the orange tree in the back yard, to fall down.

It is obvious that L2 learning is an event which is vastly more complex than a tree blowing down. It is therefore hard to see how a listing of conditions and a demonstration of their causal relevance to L2 learning can ever add up to an accumulation of understanding about L2 learning which is sufficiently predictive. And if L2 learning is not straightforwardly causal (in a linear sense), the search for causal truth will be even less profitable.

But there are other possibilities for researching an understanding of L2 learning, which do not depend exclusively on establishing cause-effect relationships.

For example, Benn and Peters (1965: 29) point out that we need to know the rules of chess in order to understand the *point* of a chessplayer's move. They go on to argue that human actions are not due to causes but to reasons. Causes are antecedent conditions, but reasons require rules, norms, and standards. The attempt to find causes for human actions will necessarily leave us forever unsatisfied because 'we should be looking not in the realm of causes but in the realm of reasons' (Hospers 1967: 342). Human knowledge (and why not include second language development?) is, in the words of Kant 'a compound of that which we receive through impressions, and that which the faculty of cognition supplies from itself' (1934: 25).

Further, Bateson (1979) suggests that learning is a stochastic process. He describes a stochastic process as one in which a random series of events interacts with a selective process. The notion of randomness in biological processes is also described in convincing terms in Lorenz 1987.<sup>3</sup>

Lastly, the notion of linear (unidirectional) causality (ie, the causation of an event by an antecedent event), has been challenged repeatedly, eg by Bertrand Russell in 1921 and, more recently, by Bandura in his

---

3 Whether or not randomness actually exists in the universe is an issue of debate among theoretical physicists. However, whether or not randomness exists among particles or in space, may not matter when discussing life forms. Indeed, it is quite possible that one of the defining differences between life and non-life is that, in the former, randomness occurs. To assume that the same laws that apply to inorganic matter must also apply to organisms is undoubtedly reductionist.

proposal of *reciprocal determinism* (1978) and Lincoln and Guba's (1985) notion of *mutual shaping*.<sup>4</sup>

If it is true that significant doubt can be cast upon the notion that L2 learning is caused directly by certain conditions, then the axiomatic primacy of the causal paradigm of scientific explanation is brought into question. Its explanatory power will be further diminished if the contention is true that causes, if shown to exist, add nothing but relatively trivial information to our quest for an understanding of language learning. Most of our efforts at doing experiments or quasi-experiments, with all the attempted controls of variables and randomisations of treatment, may be doomed to failure (especially given the complexity of language learning processes).<sup>5</sup>

As a result, it is arguable that statistical measurement, resting as it does on laborious efforts of inching towards a notion of causality or, as a way stage, a demonstrated or *probable* association between two variables, may be no more than an elaborate hoax played on us by our own deterministic minds.

There is no time here to discuss the issue of *cause* versus *reason*, or the parallel debate of *determinism* versus *free will* (Lehrer 1965; Minsky 1985; Russell 1921). I will just simply argue that a simple causal view is inappropriate in classroom research for one very uncontroversial reason, namely, that *teaching does not cause learning*. Many times learning takes place without teaching, and, perhaps equally often, the teaching event is not followed by a learning event. Many years ago, von Humboldt stated that teaching language was not possible, one could only create the conditions for learning to be possible. Lest the attentive reader reply that 'teaching is no more than creating the conditions that cause learning', von Humboldt's conditions are clearly *enabling*, not *determining* conditions. In the final analysis, I suppose, even the most ideal external conditions might fail if the learner-internal conditions (including the will to learn) were not propitious.

## Current practices in L2 classroom research

We turn to using quantities when we can't compare the qualities of things.  
(Minsky 1985: 284)

The predominant view of scientific progress in L2 classroom research, as in other domains, is that understanding is only achieved when the causes

4 In 1963, Konrad Lorenz wrote the following:

'In an organism . . . there are, so to speak, no single and straight causal chains, but an all but inextricable network of causal relationships in which, to make matters worse, the effect usually exerts an influence backwards on its cause.' (p. 11)

5 Of course, experimentation may be of immense practical utility when a straightforward link between two phenomena can be established, such as eg between smoking and cancer. It is very doubtful that such simple 'necessary connections' can be conceived to exist in learning, unless we revert to an extreme form of behaviourism.



of effects are known. Thus, in an important paper Long (1984) states matter-of-factly that 'a causal relationship must be established between program X and ESL development' (p. 410). On one particular aspect of the classroom, Chaudron (1986: 713) says that 'questioning would not be of great interest unless we believed that it contributed in a causal, quantitatively verifiable way to L2 production and development.'

Within the parameters of a causal view of teaching and learning, researchers have examined a number of conditions in order to gauge their effect on L2 development. Long (1985) studied input modifications by comparing comprehension on two lecturettes, one of which was especially adapted for ESL learners. Pica, Young and Doughty (1987) compared the amount of comprehensible input available in tasks with or without the possibility of interacting. Brock (1986) looked at the effect of display and referential questions on learner responses. The particular findings are irrelevant here, but what all these studies have in common is that one or a small set of classroom actions is selected in accordance with some theoretical or methodological assumption of relevance, and an attempt is made, largely through quantification, to measure the effect of this condition on L2 learning.

Features selected for attention include referential and display questions, wait time, various kinds of negotiation moves (requests for clarification, repetitions, confirmation checks, etc), self-repair and other-repair, features of simplified discourse, etc. The reasons for the selection of these features vary enormously. Some are hunches or ideas, while others are derived from certain theoretical or methodological positions (eg, the assumption that negotiation of information promotes comprehensible input — and hence second language acquisition — and that certain behaviours, such as requesting clarification, are instances of negotiation). Once these features are selected, they are either manipulated (ie, controlled, eg, by training some teachers to produce many of them) and then counted, or they are just counted. The quantified features are then compared to some product measure, eg language output, comprehension on a task, etc. If a statistical relationship can be established between a feature and a measure, this indicates that differences in the product are (partly) 'accounted for' by differences in the process.

Authors of causally oriented studies frequently conclude their discussions using strong terms from normative science ('show', 'demonstrate', 'find', 'establish', — but always just short of 'prove', of course) mixed with extremely hedged statements ('appears to contribute', 'some aspects of performance', 'may be due to') and this results in a perplexing blend of boldness and modesty. By contrast, in a very thought-provoking paper, Long (1985: 388) gives us perhaps the boldest statement in the entire literature, when he claims 'the existence of an indirect causal relationship between linguistic and conversational adjustments and SLA'. The relationship is 'indirect' because, if it is shown (as Long did), that the adjustments promote comprehension, and if it is shown

(actually, it is *assumed*) that comprehensible input promotes acquisition, then it follows that the adjustments promote acquisition. Several objections can be raised against such argumentation, one of them being that 'promoting' something is not equivalent to 'causing' something.

The process-product orientation exemplified in the above-mentioned studies appears to be dominant (especially in the USA), but it is of course not the only way to do research, nor is it the only way to do process-product research.<sup>6</sup>

Case study (Ellis 1984, Felix 1981), descriptive research (van Lier 1988), process research (Mitchell 1985), action research (Nixon 1981), ethnography (Watson-Gegeo 1988) and classroom discourse analysis (Sinclair and Coulthard 1975) are all different ways of studying classroom realities, and need not be concerned with the establishment of correlations or statistical significance. Rather, these varied approaches have focused on finding out what it is that teachers and learners actually do, and on creating plausible bases for well-founded judgments regarding the quality of what is done. As a next step, several researchers have begun to look at ways in which improvements in the things that are done can be achieved. Note that, in order to do these things, it is not necessary to take a causal view of actions in a classroom. It is sufficient to say that the people involved can make an effort to create optimum conditions so that learners can get on with the business of learning in the best way that they see fit, and can help each other in the process. This approach to research can be called *interpretive*, in opposition to causally oriented research (with its preferences for controlling and selecting variables) which is often referred to as *normative*. The remainder of this paper will be devoted to a description of ethnography as a principled way of doing interpretive research.

## The principles of ethnography

They sought it with thimbles, they sought it with care;  
They pursued it with forks and hope;  
They threatened its life with a railway-share;  
They charmed it with smiles and soap.

(Lewis Carroll: *The Hunting of the Snark*)

Much more has been written about what ethnographers *do* than about what ethnography *is*. As regards the latter question, a dictionary definition will suffice: 'ethnography is the description of the races of

---

6 In Long's recommendations for process-product research (1984), the process-end of the research is conducted through ethnographic study (eg, participant observation, or at least the gathering of contextualised data). In practice, at least thus far, the process-end cuts significant corners by simply coding and counting the features that are *considered to be relevant*. This etic approach makes process-product research incompatible with ethnography and actually trivialises Long's very useful recommendations.

mankind' (Longman Modern English Dictionary). As regards the former question, the emphasis on the activities of ethnographers (in this sense, anthropologists doing descriptive field work) should be no surprise, given that they traditionally engage in a wide variety of rather unpredictable work. The classic anecdote is the one attributed to Kroeber who, upon being asked what to do by a prospective field worker, took the fattest tome off his shelf and said: 'Go forth and do likewise' (see Hammersley and Atkinson 1983: 27). Traditionally, ethnography has been a craft that was learned on the job. Malinowski (1967) and Bateson (1958), among others, provide insightful descriptions of the insecurities and frustrations of the job and emphasise that, during their first attempts at field work, their failures were perhaps as significant as their successes. Serendipity and hindsight tend to play a large part initially, but the craft is perfected by doing, and a period of apprenticed experience is one of the most important assets an ethnographer can have.

This heuristic quality of ethnography makes it an inherently insecure pursuit, since there are no firm external rules and guidelines for proper scientific conduct. The worker in the field is essentially alone, and inevitably learns as much from opportunities missed, false leads too strenuously pursued, and insights by-passed in inexplicable ways, as from routine description and categorisation.

Gradually, ethnography has expanded its sphere of application from field work among unknown ethnic groups to the investigation of groups of people (however identified) in industrialised countries and urban settings, and from there has moved beyond urban anthropology into the social sciences, and finally into education, where at times the classroom is treated as an identifiable group with its own cultural characteristics. This expansion of the scope of ethnography has resulted in its adoption as a method of doing research by educational and linguistic researchers with no formal anthropological training or extensive experience of field work in exotic settings. This development gives rise to the two main criticisms that can justifiably be levelled against ethnography in education:

- (i) ethnography, not being an established scientific discipline with clearly defined parameters of scientific conduct, lacks the rigour required in terms of reliability and validity to be a valid alternative to standard research designs which emphasises statistical tests of significance and strict requirements of control of variables;
- (ii) ethnography is being conducted in an irresponsible fashion by people who are not suitably trained in it, and thus becomes a euphemism for anecdotal reports of subjective observational activities (Rist 1980). The only qualification required to be an ethnographer appears to be a dislike for statistical methods.

These criticisms need to be addressed in a fundamental way if we seek to define the potential role and value of ethnography in second language

education. If we fail to address them we may contribute to an unnecessary polarisation of research interests and a consequent trivialisation of the problems that face the profession of language education, and education in general.

The charge of lack of scientific rigour in ethnography can be refuted if we consider that one research method cannot discredit another research method simply because the latter does not adhere to the rules of conduct of the former. The blacksmith cannot criticise the carpenter for not heating the piece of wood over a fire. However, the carpenter must demonstrate a principled control over the materials used. In the case of ethnography, we need to establish the principles of conduct an ethnographer must adhere to in an uncompromising fashion. These principles must then be related to the goals of the task at hand.

It is clear that ethnography entered educational research because of widespread dissatisfaction with predominant measurement-oriented research modelled on the exact sciences. What, then, does ethnography have to offer that exact science cannot deliver? To answer this question we will discuss the two basic principles that are generally agreed to underly ethnographic enquiry:

- (i) an *emic* viewpoint
- (ii) a *holistic* treatment of cultural facts or, in other words, a concern with context.

Both these principles can be seen to be a result of the ethnographer's refusal to control the environment or to isolate features for specific scrutiny and quantification on the basis of pre-established criteria or vested interests. However, as an observer the ethnographer is of necessity selective in his/her observations, intrusive by virtue of his/her presence, and predisposed to interpret happenings in a certain way by virtue of training, beliefs, assumptions and theoretical persuasions. These are irksome hindrances in the ethnographer's desire to do interpretive science, and much of the methodological literature (see eg Garfinkel's arguments for ethnomethodology, 1967) addresses proposals to overcome them. Let me briefly discuss the two basic principles of ethnography and show how an adherence to them will influence research on second language classrooms.

### *The emic principle*

The term *emic* stands in contrast to the term *etic*. Both terms are from Pike (1964), and are clippings from the words *phonemic* and *phonetic*. *Phonemic* is an adjective pertaining to the sound system, the rules of sounds and their combinations, of one particular language. For instance, l and r are different phonemes in English (compare the words *lot* and *rot*) but not in Japanese, where they are *allophones* of one another and do not change meaning. These are phonemic observations. Phonetically, however, we can describe l and r in terms of articulation, voicing, retroflexity, or laterality, regardless of how these sounds happen to be

used in any particular language. The notions retroflex, lateral, voiced, and so on, are describable without focusing on any particular language.

The etic is therefore that which describes or generalises without regard to any particular context. Etic descriptors are tools for investigators. Often they are classificatory grids, technical descriptions, etc. Examples include, in addition to the International Phonetic Alphabet, ways of charting kinship terminology, basic parts-of-speech distinctions, universals of grammar and pragmatics, and so on.

Etic characteristics, on the other hand, refer to the rules, concepts, beliefs and meanings of the people themselves, functioning within their own group.

It would appear, superficially speaking, that the emic and the etic are conflicting tendencies in ethnographic work, and to some extent this is true. For instance, a classroom researcher who employs an etic checklist such as Flanders's FIAC schedule, will find it hard to reconcile this with the participants' own views of what happens in the lesson (see Delamont 1976).

On the other hand, etic and emic considerations can also be complementary, especially when etic stands for *tools* and *skills*, and for whatever it is that is hypothesised as being *universal* (the phonetic specification of speech sounds, certain syntactic and morphological processes, features of politeness (Brown and Levinson 1978), and so on). Working with both emic and etic categories, the ethnographer continually walks a fine line between naive observation and externally imposed interpretation. However, this is perfectly acceptable, so long as the researcher remains aware of and committed to the requirement to analyse all observations and scrutinise all interpretations and inferences rigorously. In order to resolve emic–etic conflicts, it is necessary to adhere to the second basic principle of ethnography: the *holistic* principle.

### *The holistic principle*

Very often when we look at something that is happening (anything, anywhere), this event only makes sense when we see it in connection to its surroundings. Let us say that we see someone standing on a street corner, screaming and shouting and waving his fists in the air. At first it might appear to us that another poor deranged person is venting his frustration at all and everything. However, if we see a car all smashed up by the side of the road, and another damaged car just speeding off round the corner, then we suspect that there may have been a hit-and-run accident. Acts of screaming and shouting can therefore be understood only if we know the context of their occurrence. We have to bear this in mind if we want to compare big cities, say, London and New York, in terms of the screaming and shouting that goes on in them. Counting all the screams and shouts on an average day, and finding that there are half a million of them in New York, and only a quarter of a million in London, does not get us very far in *understanding* the two cities, and *judging* them according to frustration level, number of deranged pedestrians, and

so on. Indeed, basing ourselves merely on quantity of screams and shouts, we may well draw very erroneous conclusions. It may for example turn out that fifteen per cent of the screams and shouts in New York are aimed at attracting the attention of taxi drivers, and another fifteen per cent just hailing acquaintances across the street. At which point do vocal emissions become screams and shouts? When do they signal frustration? Does acoustic amplitude co-vary with strength of emotion? And so on, and so forth.

It is unfortunately the case that most of the so-called process-product research in second-language classrooms referred to above is rather similar to the scream-and-shout counting just mentioned. The focus in such studies is more on correctly identifying a certain class of behaviours (say, repetitions), and obtaining a high inter-coder reliability, than on examining the behaviours in their context in order to find out what they mean, why they occur when they occur, and the range of functions they may fulfil. This leads to much counting for the sake of counting, and the researcher comes to be in danger of resembling that famous character from *Sesame Street*, Count Dracula, who spends his entire life compulsively counting everything in sight.

### *A unitary view of ethnographic research*

Scientists in all walks of life need to conform to certain standards by which the peer group evaluates them. This is no different in ethnographic research. I have earlier pointed out that one of the problems of an ethnographic approach to classroom research is that this research often has to be done by people with no formal training in ethnography. People inclined to do ethnography may therefore have no clear idea as to what would count as 'good' or 'bad' ethnography. Although this problem also exists in normative types of research, workers in the latter tradition have the advantage that most graduate degree programmes have substantial components of quantitative research design and statistics training, whereas training in ethnography is rarely available.<sup>7</sup>

Moreover, it appears that ethnography is a craft learned by doing and by example, so that it is difficult to become a skilled ethnographer just by taking courses in it.

There is thus a danger that we will see a significant amount of bad ethnography before classroom researchers reach a consensus about how to judge the quality of an ethnographic study. This would not matter a great deal (if we assume that even bad ethnography helps to establish guidelines for proper scientific conduct) were it not that funding agencies, journals, employers, and so on, tend to emphasise cost-effectiveness rather than exploration for the sake of scientific advancement, and it

---

7 There are exceptions to this, of course. There are several universities which offer excellent cross-disciplinary courses preparing students for doing educational ethnography (see, eg, Levine et al. 1980). Eventually, such expertise will become more available to L2 classroom research than it is at present.

would thus be very easy for ethnography to become discredited and to fall into disrepute.<sup>8</sup>

It is therefore important for the language teaching profession, not just for proponents of interpretive research, but equally for normative researchers (who need a counterbalance to their proposed findings), to make an effort to put ethnography on the map. To do this, several different things need to be borne in mind.

1. First, every study needs to be scrutinised for its adherence to the *emic and holistic principles* outlined above.

2. Secondly, the notion of *context* needs to be examined in great detail, and the role of context in interpretation must be made explicit. It must be remembered that the context is potentially as wide as the whole world: a line must be drawn somewhere. Whether context is seen as primarily the context of interaction (micro-context) or the context of the classroom in society (macro-context), clear indications must be given as to the *relevance* of using contextual features. There can of course be too much, as well as too little, examination of context. Occam's razor, or the law of parsimony, applies here as it does elsewhere.<sup>9</sup>

3. Thirdly, ethnographic research must be *open*, that is, it must examine and report its own processes of inferencing and reasoning, so that its procedures can be profitably discussed. The opposite of this requirement of openness or transparency would be the opacity which characterises some normative research, where it may for example be unclear why certain — but not other — statistical tests were applied, variables controlled, hypotheses formulated, and so on.

4. Fourthly, analysis must be either *broad* (longitudinal) or *deep* (micro-ethnographic). Critics of unsatisfactory ethnography (eg Rist 1980) often complain of a tendency to do 'Blitzkrieg ethnography', quick in-and-out raids on classrooms or schools, in order to get some data, and write a paper or thesis. This criticism is no doubt very valid. However, we must not make the mistake of assuming that all ethnographic work must be longitudinal. Much painstaking analysis of minute pieces of data is also required (see van Lier 1988, Erickson 1981, Sacks 1972, for examples of micro-ethnography or conversation analysis), so that length of time spent in a research site alone does not determine quality. Ethnography requires intensive immersion in the data, whether this is the daily language use of an entire culture, or one small story told by a child.

One thing is relatively clear. It is very difficult to conduct a responsible ethnographic study in the limited timespan usually allowed in the cycle of

---

8 Lewin's action research, first proposed in the 40s, is a case of a very useful research tool which never became popular at the time due to the enormous prestige of psychometric methods; now it is enjoying a very belated comeback (see Sanford 1981, van Lier and Bailey 1989).

9 For a principled and well-argued approach to examining data in context, see Mehan 1979. Erickson and Shultz 1981 is also very insightful. The first study seriously tackling the concept of context in SLA is Selinker and Douglas 1985.

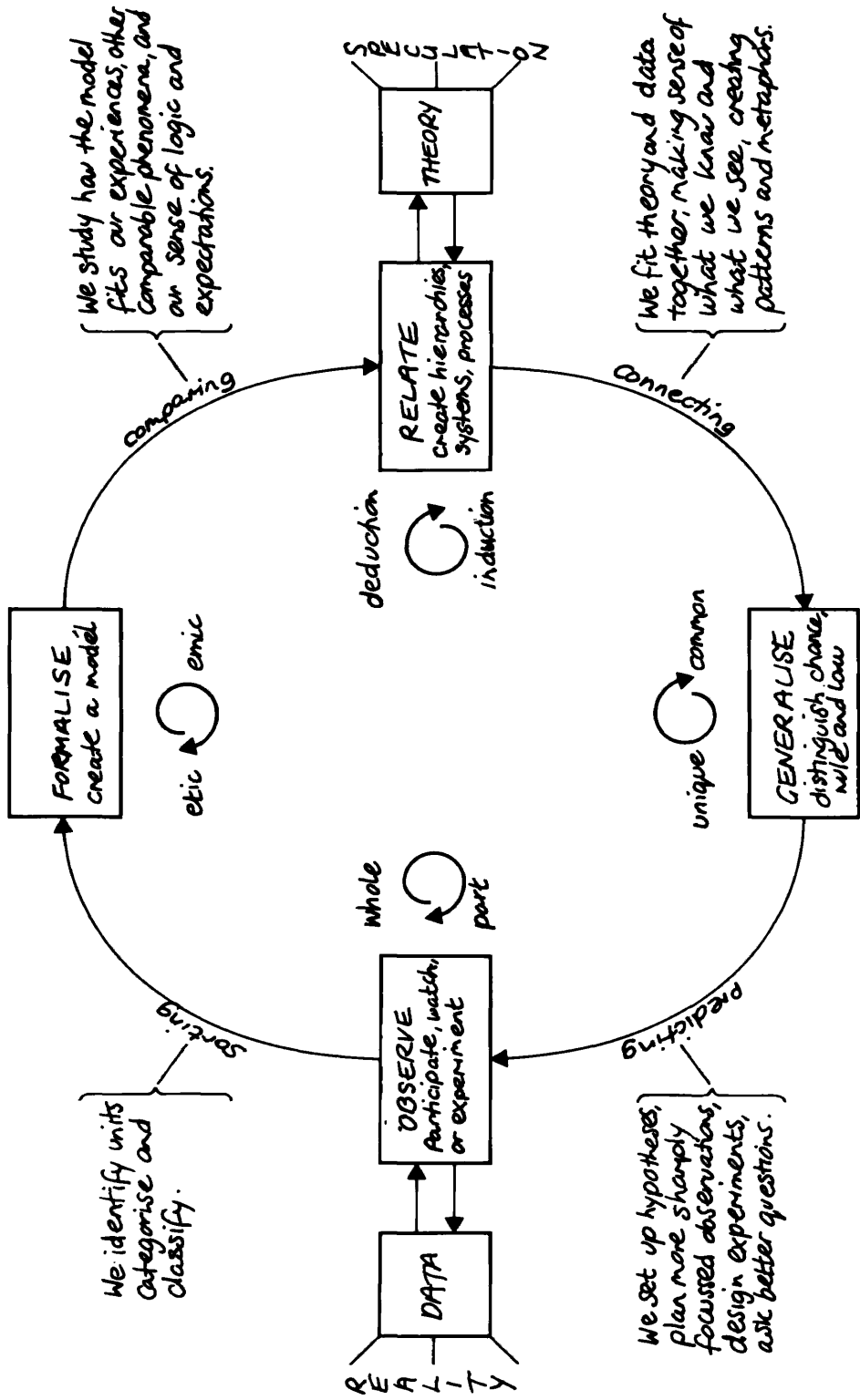


Fig.2 Cycles of research



conference presentations: a brief treatment period, a testing session, a twenty-page write-up, all probably wrapped up in about three to six months from start to finish. Ethnography is not conducted that quickly. If it needs to be fitted into the conference cycle, we will get a great many 'preliminary ethnographies' which lack depth and probably do not deserve to be called ethnographic research. A good ethnographer will never claim to have found sufficient or even adequate clues to knowledge about the people studied. Every insight generates further questions, and every question suggests further avenues for exploration.

Ethnography is thus a cyclical enterprise. While it is clearly data-based and data-oriented (ie, empirical), it also draws on and fuels theories and speculative assumptions. The diagram above shows how the cycle of activities in ethnography can be conceptualised. It is important to realise that we do not start in any particular place (eg, at the data-end); rather, we are already *in* the cycle, as members of society, we merely sharpen our vision as we continue to progress.

## **Applications of ethnography in second language classrooms**

I said it in Hebrew – I said it in Dutch –  
I said it in German and Greek:  
But I wholly forgot (and it vexes me much)  
That English is what you speak!

Lewis Carroll: *The Hunting of the Snark*

Given the current popularity of educational ethnography, it is surprising how little ethnographic work has been conducted in L2 classrooms. We know very little of what we really mean by 'instructed L2 situations', or 'classroom interaction'. Only very recently the L2 profession has begun to examine the general educational tradition in ethnography, and to take tentative steps towards applying some of the insights and experience gained there to the language classroom. This in spite of a veritable chorus of calls for more information about the context of language learning, and the nature of interlingual interaction, from almost all researchers. Quite obviously, the profession has a number of things to learn. We do not seem to be comfortable sailing between the Scylla and Charybdis of the following extreme positions regarding ethnography:

- (a) anyone can do it, all you have to do is watch and see what you can see;
- (b) you need to be at least as well trained in ethnography as normative researchers are in statistics and research design, otherwise they will just ridicule your efforts.

Basically, this is a credentialling problem: who decides that a researcher can call him/herself an ethnographer, or that a piece of work can be called ethnographic? I believe that the profession as a whole has not decided what the appropriate criteria might be, in the way that at

least a reasonable consensus exists for the judging of normative work. It is my hope that the four requirements described above will help in building such a consensus.

We can also learn from the pioneering work of some second language researchers to search for fruitful directions in second language classroom research (for more detailed discussions, see Mitchell 1985, van Lier 1988). One of the earliest exploratory studies in second-language classroom interaction was conducted by Dick Allwright (1980), who used a recorded and transcribed second-language lesson to describe patterns of participation, in the process developing useful insights into the turn-taking system. Particularly innovative, and characteristic of an ethnographic (emic) approach was his decision not to distinguish *a priori* between teacher and student roles, but rather to allow patterns of control and initiative to emerge from the data. This stands in contrast to the approach of eg Sinclair and Brazil (1982), and McHoul (1978), who assume characteristic teacher and pupil roles as givens, and thus illustrate an etic-before-emic approach. Allwright also demonstrates the fruitfulness of justifying inferences in terms of their 'plausibility', an approach recommended by Erickson (1985), and also visible in Kumaravadivelu's (1988) study of 'learning opportunities' based on a micro-ethnographic analysis of video-taped classroom data.

In addition to micro-ethnographic analysis of transcribed data, the use of diaries as data can be very fruitful. In second-language settings diary studies have been conducted by eg Schumann and Schumann (1977), Bailey (1983; see also Bailey and Ochsner 1983) and Schmidt and Frota (1986). Such studies are particularly useful for the description of individual cognitive and affective factors accompanying the learning process. Expanding somewhat, the use of techniques to elicit introspective and retrospective data is rapidly becoming a field of substantial interest, after early pioneering work of Hosenfeld (1979), Cavalcanti (1982) and Mann (1982; see also a recent collection edited by Faerch and Kasper 1987).

Another way to approach ethnographic study is through an examination of the basic underlying concepts. Thus, Selinker and Douglas (1985) attempted to find out what we mean by the notion of 'context', by examining in detail tutoring sessions which differed in topic and participants' background knowledge.

These are useful and promising beginnings. However, work approaching the sophistication of some of the best work in general education, such as Mehan (1979), Hymes et al. (1981), Heath (1983), and studies reported in Trueba et al. (1981) and Trueba (1987) still needs to be done in SLA (see van Lier (1988) for an exploration of the parameters of ethnographic work in L2 classrooms).

## **A look into the future**

Is ethnography only exploratory? Or is it also theory building and theory

testing? In the former case, ethnography serves as preparatory ground-clearing for causal research, in the latter case its methods carry ethnography all the way to a full understanding of social issues. There are those who take one or the other position (see Hammersley and Atkinson 1982 for a detailed discussion). It is in the interest of our profession to pursue both options, and at all times to maintain a conversation between normative and interpretive kinds of research, without assuming that differences are trivial. They are not, but in studying the differences a greater depth of understanding may be reached.

There are a number of practical areas of concern in which the use of ethnographic methods in research can be of assistance. For example, the role of evaluation in a learner/learning-centred curriculum (as advocated eg in Candlin 1987) can be fruitfully explored through a longitudinal programme of ethnographic monitoring (see Hymes 1981).

A related area is programme evaluation. Beretta, studying the effect of the Bangalore Project (see Prabhu 1987), illustrated an unusual aspect of ethnographic procedures: the use of historical data or documents, complemented with interviews. From this analysis, the notion of 'ownership' has emerged as a powerful determining factor in the teachers' application of Dr Prabhu's principles.

The curriculum itself can also profit from ethnographic research. Watson-Gegeo, by comparing learning in school to learning in the community, proposed innovative changes in the curriculum in order to make school culturally compatible with society (see the Kamehameha project, and the use of similar procedures in the Solomon Islands; Watson-Gegeo, 1988). In a similar way, Heath (1983) suggests ways of curriculum renewal through a detailed comparison of questioning at home and in school.

A crucial way in which ethnography differs from normative research is that, if it is done right, it actively encourages the participation of teachers and learners. Currently there is much discussion about action research, the empowerment of teachers, teachers 'helping themselves' to research, and related sentiments. At the same time many researchers emphasise the need for case studies (Yin 1984), which involve longer-term collection of data and much detail of description (see eg Schmidt and Frota 1986), and which allow for a finer-grained examination of context than is possible in the process-product research. In a practical sense, ethnographic research can help solve immediate problems, such as those inherent in the employment of foreign teaching assistants (FTAs) in tertiary education. For example, Bailey (1985) uses extensive data collection and analysis to provide a useful categorisation of types of teaching which can be helpful in in-service training. Also in the context of FTAs' classroom work, Rounds (1987) examines the functions and uses of silence in mathematics lectures.

Finally, several curriculum theorists advocate using ethnography as part of a task-based curriculum, that is, designing tasks which encourage

learners to use the methods of ethnography to discover significant aspects of target language use (Heath 1986).

It is important to pursue controlled studies of the process-product type. However, it is also essential that, if only in order to safeguard against the potential triviality and misinterpretation that accompanies all piecemeal empiricism, ethnographic studies are conducted which examine all actions and interactions in classrooms within their context, both wide and narrow. Only in this way can research on classroom language learning move forward: each research type, normative and interpretive, serves as a safeguard against misinterpretations and false directions in the other. Rather than saying that the two types of research are combinable (and perhaps implying that the differences are minor), therefore, we regard them as alternative ways of knowing, both of which are necessary to arrive at a better understanding of the reality of the language classroom. This is the sense in which Smith and Heshusis (1986) speak of a 'conversation' between the two approaches to science, a conversation which we must make the effort to preserve, but which is 'closed down' if it is assumed that the differences between them are minimal.

It is possible that interpretive and normative research programmes will sometimes provide the same sorts of findings. It is more likely, however, that they will yield different kinds of information, information that may either be compatible, or contradictory. Whichever way things turn out, a diversity of research programmes is essential to promote an enrichment of theoretical and professional knowledge.

### *Acknowledgement*

I am grateful to Kathi Bailey and Lyn Goldstein for commenting in detail on a previous draft of this paper.

### **References**

- Adelman, C. (ed.) (1981) *Uttering Muttering*, London: Grant McIntyre.
- Allwright, R. L. (1980) 'Turns, topics and tasks: patterns of participation in language learning and teaching' in Larsen-Freeman, D. (ed.) *Discourse Analysis in Second Language Research*: 165-87, Rowley, Massachusetts: Newbury House.
- Bailey, K. M. (1983) 'Competitiveness and anxiety in adult second language learning: looking at and through the diary studies' in Seliger, H. W. and Long, M. H. (eds) *Classroom Oriented Research in Second Language Acquisition*: 67-103, Rowley, Massachusetts: Newbury House.
- Bailey, K. (1985) 'If I had known then what I know now: performance testing of foreign teaching assistants' in Hauptmann, P. C. LeBlanc, R. and Wesche, M. B. (eds.) *Second Language Performance Testing*: 153-180, Ottawa: University of Ottawa Press.
- Bailey, K. M. and Ochsner, R. (1983) 'A methodological review of the diary studies: windmill tilting or social science?' in Bailey, K. M. Long, M. H. and Peck, S. (eds.) *Second Language Acquisition Studies*: 188-98, Rowley, MA: Newbury House.
- Bandura, A. (1978) 'The self system in reciprocal determinism', *American Psychologist* 33.
- Bateson, G. (1958) *Naven*, Stanford, CA: Stanford University Press.
- Bateson, G. (1979) *Mind and Nature: A Necessary Unity*, London: Fontana.

- Benn, S. I. and Peters, R. (1965) *Principles of Political Thought*, New York: Free Press.
- Beretta (1988) 'Implementation of the Bangalore project', presentation at 22nd Annual TESOL Convention, Chicago.
- Brock, C. A. (1986) 'The effects of referential questions on ESL classroom discourse', *TESOL Quarterly* 20: 47-59.
- Brown, P. and Levinson, S. (1978) 'Universals in language usage: politeness phenomena' in Goody E. (ed.) *Questions and Politeness: Strategies in Social Interaction*: 56-289, Cambridge: Cambridge University Press.
- Candlin, C. N. (1987) 'Towards task-based language learning' in Candlin, C. N. and Murphy, D. F. (eds.) *Language Learning Tasks*, Englewood Cliffs, NJ: Prentice-Hall: 5-22.
- Cavalcanti (1982) 'Using the unorthodox, unmeasurable verbal protocol technique: qualitative data in foreign language reading research' in Dingwall S. and Mann, S. (eds.) *Methods and Problems in Doing Applied Linguistic Research*: 72-85, University of Lancaster.
- Cazden, C. B. (1985) 'Classroom discourse' in Wittrock, M. C. (ed.) *Handbook of Research on Teaching*: 432-63, New York: Macmillan.
- Chaudron, C. (1986) 'The interaction of quantitative and qualitative approaches to research: a view of the second language classroom', *TESOL Quarterly*, 20: 709-17.
- Delamont, S. (1976) 'Beyond Flanders' fields' in Stubbs, M. and Delamont, S. (eds.) *Explorations in Classroom Observation*, Chichester: John Wiley.
- Ellis, R. (1984) *Classroom Second Language Development*, Oxford: Pergamon Press.
- Erickson, F. (1981) 'Some approaches to inquiry in school-community ethnography' in Trueba, H. T. Guthrie, G. P. and Au, H. P. (eds.) *Culture and the Bilingual Classroom: Studies in Classroom Ethnography*: 17-35, Rowley, MA: Newbury House.
- Erickson, F. (1985) 'Qualitative methods in research on teaching' in Wittrock, M. C. (ed.) *Handbook of Research on Teaching*: 119-61. New York: Macmillan.
- Erickson, F. and Shultz, J. (1981) 'When is a context? Some issues and methods in the analysis of social competence' in Green, J. and Wallat, C. (eds.) *Ethnography and Language in Educational Settings*: 147-60, Norwood, NJ: Ablex.
- Faerch, C. and Kasper, G. (eds.) (1987) *Introspection in Second Language Research*, Clevedon: Multilingual Matters.
- Felix, S. (1981) 'The effect of formal instruction on second language acquisition', *Language Learning* 31: 87-112.
- Garfinkel, H. (1967) *Studies in Ethnomethodology*, Englewood Cliffs, NJ: Prentice-Hall.
- Gilmore, P. and Glatthorn, A. A. (eds.) (1982) *Children In and Out of School*, Washington, DC: Center for Applied Linguistics.
- Hammersley, M. and Atkinson, P. (1983) *Ethnography, Principles in Practice*, London: Tavistock Publications.
- Heath, S. B. (1983) *Ways with Words: Language, Life, and Work in Communities and Classrooms*, Cambridge: Cambridge University Press.
- Heath, S. B. (1986) *Children at Risk? Building Investment in Diversity*, Plenary Address, 20th TESOL Convention, Anaheim, California.
- Hosenfeld, C. (1979) 'Cindy: a learner in today's foreign language classroom' in Born, W. (ed.) *The Foreign Language Learner in Today's Classroom Environment*, Northeast Conference Reports, Northeast Conference on the Teaching of Foreign Languages, Middlebury, Vermont.
- Hospers, J. (1967) *An Introduction to Philosophical Analysis*, Englewood Cliffs, NJ: Prentice-Hall.
- Hymes, D. (1981) 'Ethnographic monitoring' in Trueba, H. T. Guthrie, G. P. and Au, H. P. (eds.) *Culture and the Bilingual Classroom: Studies in Classroom Ethnography*: 56-58, Rowley, MA: Newbury House.
- Hymes, D. et al. (1981) *Ethnographic Monitoring of Children's Acquisition of Reading/Language Arts Skills In and Out of the Classroom*, Vols I, II, III, Final Report to NIE, Washington, DC, (ED 208 096).
- Kant, I. (1934) *A Critique of Pure Reason*, London: Dent.

- Kumaravadivelu, B. (1988) 'Creation and utilization of learning opportunities', 22nd Annual TESOL Convention, Chicago, 8–13 March, 1988.
- Lehrer, K. (ed.) (1965) *Freedom and Determinism*, New York: Random House.
- Levine, H. G., Gallimore, R., Weisner, T. S., and Turner, J. L. (1980) 'Teaching participant-observation research methods: a skills-building approach', *Anthropology and Education Quarterly* XI: 38–54.
- Long, M. H. (1984) 'Process and product in ESL program evaluation', *TESOL Quarterly* 18: 409–22.
- Long, M. H. (1985) 'Input and second language acquisition theory' in Gass, S. and Madden, C. (eds.) *Input and Second Language Acquisition*, Rowley, MA: Newbury House.
- Lorenz, K. (1963) 'A scientist's credo' in Lorenz, K. *Studies in Animal and Human Behaviour*, Vol. II: 1–13, London: Methuen.
- Lorenz, K. (1987) *The Waning of Humaneness*, New York: Little Brown.
- McHoul, A. W. (1978) 'The organization of turns at formal talk in the classroom', *Language in Society* 7: 183–213.
- Malinowski, B. (1967) *A Diary in the Strict Sense of the Term*, London: Routledge and Kegan Paul.
- Mann, S. (1982) 'Verbal reports as data: a focus on retrospection' in Dingwall, S. and Mann, S. (eds.) *Methods and Problems in Doing Applied Linguistic Research*: 72–85, University of Lancaster.
- Mehan, H. (1979) *Learning Lessons: Social Organization in the Classroom*, Cambridge, MA: Harvard University Press.
- Minsky, M. (1985) *The Society of Mind*, New York: Simon and Schuster.
- Mitchell, R. (1985) 'Process research in second-language classrooms', *Language Teaching* 18: 330–52.
- Nixon, J. (ed.) (1981) *A teacher's guide to action research*, London: Croom Helm.
- Phillips, D. C. (1987) *Philosophy, Science, and Social Inquiry*, Oxford: Pergamon Press.
- Pica, T., Young, R. and Doughty, C. (1987) 'The impact of interaction on comprehension', *TESOL Quarterly* 21, 4: 737–58.
- Pike, K. L. (1964) *Language in Relation to a Unified Theory of Structures of Human Behaviour*, The Hague, Mouton.
- Prabhu, N. S. (1987) *Second Language Pedagogy*, Oxford: Oxford University Press.
- Ravetz, J. R. (1971) *Scientific Knowledge and its Social Problems*, Oxford: Clarendon Press.
- Rist, R. C. (1980) 'Blitzkrieg ethnography: on the transformation of a method into a movement', *Educational Researcher* 9: 8–10.
- Rounds, P. (1987) 'Characterizing successful classroom discourse for NNS Teaching Assistant Training', *TESOL Quarterly* 21: 643–71.
- Russell, B. (1921) 'A note on cause' in *Mysticism and Logic*, London: Longmans, Green and Co.
- Sacks, H. (1972) 'On the analyzability of stories by children' in Gumperz, J. J. and Hymes, D. (eds.) *Directions in Sociolinguistics*: 329–45, New York: Holt, Rinehart and Winston.
- Sanford, N. (1981) 'A model for action research' in Reason, P. and Rowan, J. (eds.) *Human Inquiry: A Sourcebook of New Paradigm Research*: 173–81, Chichester: John Wiley.
- Saville-Troike, M. (1982) *The Ethnography of Communication: An Introduction*, Baltimore: University Park.
- Schmidt, R. W. and Frota, S. N. (1986) 'Developing basic conversational ability in a second language: a case study of an adult learner of Portuguese' in Day, R. R. (ed.) *Talking to Learn: Conversation in Second Language Acquisition*, Rowley, MA: Newbury House.
- Schumann, F. E. and Schumann, J. H. (1977) 'Diary of a language learner: an introspective study of second language learning' in Brown, H. D. Crymes, R. H. and Yorio, C. A. (eds.) *Teaching and Learning: Trends in Research and Practice* 241–9, Washington, DC: TESOL.
- Selinker, L. and Douglas, D. (1985) 'Wrestling with "context" in interlanguage studies', *Applied Linguistics* 6: 190–204.

- Sinclair, J. McH. and Brazil, D. (1982) *Teacher Talk*, Oxford: Oxford University Press.
- Sinclair, J. McH. and Coulthard, M. (1975) *Towards an Analysis of Discourse*, Oxford: Oxford University Press.
- Smith, J. K. and Heshusius, L. (1986) 'Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers', *Educational Researcher* 15, 1: 4-12.
- Trueba, H. T., Guthrie, G. P. and Au, H. P. (1981) *Culture and the Bilingual Classroom: Studies in Classroom Ethnography*, Rowley, Massachusetts: Newbury House.
- Trueba, H. T. (ed.) (1987) *Success or Failure: Learning and the Language Minority Student*, Rowley, Massachusetts: Newbury House.
- van Lier, L. (1988) *The Classroom and the Language Learner: Ethnography and Second-Language Classroom Research*, London: Longman.
- van Lier, L. and Bailey, K. 'Innovation and action research in pre-service teacher education', 23rd Annual TESOL Convention, San Antonio, 7-11 March, 1989.
- Watson-Gegeo, K. (1988) 'Ethnography in ESL: defining the essentials', *TESOL Quarterly* 22, pp. 575-92.
- Wittgenstein, L. (1961) *Tractatus Logico-Philosophicus*, London: Routledge and Kegan Paul.
- Yin, R. K. (1984) *Case Study Research: Design and Methods*, Beverly Hills: Sage Publications.

# Researching Classroom Language Learning

Rod Ellis

*Temple University, Japan*

## Introduction

In this article I shall describe and evaluate two approaches for investigating the relationship between instruction and second language (L2) learning. One approach — the *linguistic* or *psycholinguistic* — takes as its starting point either a theory of language or a theory of L2 acquisition and seeks to test hypotheses based on the theory using data obtained from L2 learners. An example of a linguistic theory which is currently attracting the attention of researchers who follow this approach is Chomsky's Government and Binding Theory (Chomsky, 1981). An example of a psycholinguistic theory which has attracted considerable attention in the 1980s is Krashen's Monitor Model (Krashen, 1981). Researchers who follow this approach may or may not be concerned with improving pedagogic practice. The main aim of some researchers is theory construction. The classroom merely serves as a convenient setting in which to carry out empirical work. Other researchers in this tradition, however, are motivated by a desire to increase instructional efficiency, believing that progress is best assured if it is research-led and if the research is based on a strong theory (Long, 1985).

The second approach is *educational* or *pedagogic* (Ellis, 1988a). In this case the research question derives from some educational principle (for example, 'Does discovery learning lead to deeper and better-remembered understanding?') or from some pedagogical technique (for example, 'Does controlled practice promote acquisition?'). The aim of this research is to subject the principle or technique to critical scrutiny in order to (1) identify the assumptions that underlie it and (2) evaluate these assumptions in the light of what is currently known about language learning. Educational/pedagogical research of this kind can be both rationalist in nature (as in the work of Widdowson, 1978; 1984) or it can be empirical (as in Seliger, 1977, or Day, 1984). Often enough it is a mixture of the two, as in much of my own work (Ellis, 1984) or Allwright's (1980; 1984).

In this paper I shall limit myself to a discussion of the empirical research based on these two approaches. My aim is to sketch out the kinds of issues investigated in both 'schools', to examine the procedures



which have been used and to discuss some of the problems that have arisen. It is *not* my intention, however, to argue that one approach is better than the other. Both provide the teaching profession with useful information which is of value when considering the content and methodology of language pedagogy. However, I shall argue that, irrespective of the research approach, the results obtained do not constitute 'facts' to be incorporated into pedagogy but rather provide 'insights' which can help to illuminate pedagogy. In other words, the research constitutes only one source of information which can guide pedagogy. It needs to be considered in conjunction with teachers' ongoing experience of what works and what does not work. It should contribute to, but never direct, decision-making.

### Problems of definition

The aim of both approaches is to investigate in what ways and to what extent *instruction* contributes to *L2 learning*. There are, however, problems with defining both the independent and dependent variables for research purposes. I shall begin, therefore, by examining some of these problems.

Instruction can be defined in two ways, reflecting the two approaches. (1) Long (1988) has proposed that instruction be defined in terms of 'psycholinguistically relevant design features'. (2) Ellis (1988a) has suggested that an educational approach views instruction in terms of methods or specific instructional procedures.

By 'psycholinguistically relevant design features', Long means those characteristics of classroom environments that can be distinguished in accordance with a strong theory of L2 acquisition. One important distinction, which Long identifies, is that between a *focus on form* and a *focus on meaning*. The central research question derived from this distinction concerns which mode of instruction works most efficiently, there being three possibilities:

1. An exclusive focus on form;
2. An exclusive focus on meaning;
3. A combination of a focus on form and meaning.

There are a number of problems with such a distinction.

First, it is not clear to what unit of instruction the form *vs* meaning distinction is to be applied. It is tempting to distinguish the two types of instruction at the level of 'lesson', ie by labelling some lessons 'form-focused' and others 'meaning-focused'. This may be possible in the case of lesson-planning, but when it comes to the interactions through which any single lesson is enacted, it is almost certain that the lesson will be comprised of both form and meaning-focused classroom behaviours. At the level of implementation, then, lessons will differ not absolutely but only in the degree to which they lean towards one or the other design feature. Lesson transcripts show that a teacher typically moves to and

from a focus on form or meaning, irrespective of what the overall aim of the lesson is. Ideally, then, the distinction should be applied to some smaller unit than that of 'lesson' — perhaps 'exchange' or 'sequence' — and instructional treatments be differentiated in terms of the *frequency* with which exchanges directed at form or meaning occur or, as Chaudron (1988) suggests, in terms of the amount of time devoted to each.

Second, it is not clear whose perspective the researcher is supposed to take — the teacher's or the learners'. McTear (1975) observed that there is often a conflict in perspective. For example, a question which a teacher asks in order to elicit a specific linguistic feature (= a focus on form) may be treated by a learner as a request for real information (= a focus on meaning). It is the learner's perspective that is crucial if the intention is to examine the effect of the design feature on acquisition, yet it is rare that classroom researchers ascertain what is the actual focus of the learners' attention.

Third, related to the point above, viewing instruction as a focus on form as opposed to focus on meaning ignores the fact that the same instruction can often mean different things to individual learners in the classroom (Cohen and Apeh, 1981). Thus, it is possible that at any one moment in a lesson one learner will be focused on form (in accordance with or despite the intention of the teacher) while another will be focused on meaning.

Problems such as these testify to the difficulty of operationalising the construct of 'instruction' in terms of 'design features' such as those proposed by Long. However, such problems are not insuperable, although they do call for a degree of flexibility in research design and methodology that is sometimes lacking (see discussion below).

Instruction viewed as 'method' is even more problematic. Early comparative method studies (eg Scherer and Wertheimer, 1964; Smith, 1970) failed to show that different methods resulted in differing levels of achievement. More recent studies (eg Harley, Allen, Cummins and Swain, 1987) have also been unable to demonstrate that 'method' is a significant factor in determining learning outcomes. There is a general feeling among L2 acquisition researchers that 'method' cannot provide an adequate definition of the independent variable. Allwright (1983; 1988) talks of the 'failure' of the method construct. Larsen-Freeman (1988) argues that 'future research should not attempt to compare methodologies at the global level, but rather should focus on local practices'.

This raises the question as to what is meant by 'local practices' and how these can be operationalised for the purposes of research. One aspect of language teaching that has received attention is that of language practice (cf. Ellis, 1988b). But what serves as a perfectly adequate pedagogic construct is problematic when it comes to conducting research. 'Language practice' proves to be a very loose notion, covering a variety of activities which result in diverging patterns of classroom behaviour. In order to research 'language practice' it is necessary to arrive at a tight, narrow definition, which may then turn out to be of limited interest to

language teachers. What is true for 'language practice' is true for other pedagogical procedures. It is often very difficult to identify specific classroom behaviours that relate to the instructional techniques mentioned in methodology handbooks.

The problems of *a priori* definitions of instruction have led some applied linguists to reject them as a basis for classroom research. Allwright (1988), for instance, claims that:

... any *a priori* characterization of learning and teaching environments is bound to be suspect for fundamental research purposes.

He goes on to argue:

... the characterization of teaching and learning environments is something that must *emerge from research* rather than something that can be *imposed* on research.

For this reason, Allwright favours research that is data rather than theory-led. His position, therefore, appears to be directly opposed to that of Long's.

I have sympathy with Allwright's position. Exploratory research based on detailed classroom observation is obviously needed to illuminate what takes place during teaching-learning. But I am reluctant to abandon *a priori definitions*, as to do so is to cut the ties between what language teaching methodologists believe to be important and what researchers do. Allwright's arguments do not warrant a dismissal of hypothesis-testing research — whether the hypotheses derive from a linguistic-psycholinguistic theory or from educational/pedagogic constructs. However, it is obviously necessary to ensure that (1) 'instruction' is operationalised explicitly and rigorously and (2) process-information is obtained through classroom observation and learner introspection in order to determine whether external definitions of instruction are realised in identifiable classroom behaviours. As we shall see, however, there are very few studies that satisfy both of these criteria.

The dependent variable — learning — is even more problematic. There is no direct window through which the researcher can peer to discover what the learner knows. There is not even a consensus regarding what the term 'knowledge' of an L2 means. When researchers seek to relate instructional treatments to learning outcomes, they need instruments with which to measure what learning has taken place. The problem facing the L2 acquisition researcher is really the same as that facing the language tester — how to provide valid and reliable measurements of what the learner knows. The solutions to this problem are various — which is another way of saying again that there is no consensus among researchers. On the one hand there are researchers who operate within a competence/performance framework and who employ instruments designed to capture the learner's intuitions about well-formedness in the L2. The principal instrument they use is some kind of grammaticality judgement task (ie a task that requires learners to judge whether

sentences are grammatically correct or not). On the other hand, there are researchers who base their study of L2 acquisition on cognitive theories of language processing and who employ instruments designed to capture samples of 'naturally occurring' speech (eg informal interviews or information-gap activities). Clearly, the concept of 'learning' that underlies these two approaches to measurement is very different. There have been classroom studies based on both approaches; not surprisingly, they have produced diverging results, which are hard to reconcile.

I see no way out of this problem. As long as there are competing theories of language and language learning there will be differences in opinion regarding how learning should be measured. It might be argued, however, that there are some ways of measuring learning that have greater validity for the L2 classroom researcher who is concerned with investigating how language pedagogy can be made more efficient (as opposed to the L2 researcher who is more concerned with theory-building). It is surely more important to pay attention to research that addresses what learners can *do* in real-life situations rather than what they *know* but might not be able to perform. In other words, learning should be measured by means of tasks that elicit natural language behaviour rather than by tasks that elicit grammaticality judgments. The latter may result in arcane behaviours which do not correspond to what most teachers recognise as the goal of language instruction.

It is important to recognise these problems. They constrain the kinds of research that have taken place. All classroom L2 researchers must take decisions regarding the operationalisation of 'instruction' and 'learning'. Good research is research that tries to address the problems discussed above in an explicit manner. However, in reaching solutions, researchers run the risk of oversimplifying and thereby trivialising what is, in actuality, an immensely complex phenomenon — the relationship between instruction and learning.

### **The linguistic/psycholinguistic approach**

Research based on the linguistic/psycholinguistic approach has sought answers to three key questions in recent years:

1. To what extent is it possible to teach learners grammatical structures? This question relates to the role of instruction when the focus is on form.
2. What kinds of communication promote L2 acquisition in the classroom? This question relates, in particular, to the role of instruction when the focus is on meaning.
3. Is L2 grammar learnt more efficiently through a focus on form or a focus on meaning?

I shall consider a number of studies illustrative of the research which has tackled these questions. My aim is to examine the methodological

procedures which these studies have followed rather than to report results in detail.

The first question has been researched in two ways. Data collected from naturalistic learners (ie learners with no exposure to formal instruction) is compared with data collected from classroom learners with a view to discovering whether grammatical properties of the L2 are acquired in the same or in a different sequence. The underlying rationale of this approach is that if instruction has a direct effect on acquisition this ought to become manifest in a different sequence. Conversely, if no difference in developmental sequences is apparent, then this indicates that language teaching is unable to influence the natural process by which a learner acquires grammatical knowledge of the L2. A good example of this research is Pica's (1983) study of morpheme sequences in groups of naturalistic, classroom and mixed learners. What is missing from this line of research is a process element, ie studies such as Pica's provide no information about how the learners were exposed to the L2 or how different learners react to this exposure. We are told nothing about what actually happens in the natural and classroom environments. Thus, the research rests entirely on an analysis of products. The conclusions based on such research, therefore, are necessarily high-inference and must be treated with circumspection.

The second kind of research investigating the role of grammar teaching makes use of an experimental (or rather pseudo-experimental) design. Experimental studies enable researchers to test for the effects of instruction *directly*. It is possible to investigate whether instruction in feature *x* leads to acquisition of *x*. One of the most frequently cited studies of this kind is Pienemann's (1984). Pienemann based his research on the finding from naturalistic studies of the acquisition of L2 German that learners learn word order rules in a fixed sequence. His study was designed to investigate what happened when a group of Italian children aged 7–9 years was taught a difficult German word order rule (ie one that was typically acquired late by naturalistic learners). The study followed a standard pre-test, treatment, post-test design. First, data relating to the learners' existing stages of acquisition were collected by means of 'linguistic interviews' and 'hidden recordings' made in the children's playing environment. Next, the subjects received several days instruction directed at the target feature (INVERSION). Finally, further samples of the learners' speech were collected using the same means as before. The data were then analysed for evidence of acquisition of word order rules. Pienemann presents results for only two children, one of whom acquired INVERSION as a result of instruction and one of whom did not. The crucial factor, according to Pienemann, was whether the learners had reached the stage of word order development immediately preceding that of INVERSION. If this was the case, the instruction worked; if it was not the case, the instruction failed.

Pienemann's study is an interesting example of classroom experimental research based on a fairly well-articulated psycholinguistic

theory. However, it raises a number of questions. First, as reported by Pienemann, the study fails to provide explicit information about the nature of instruction to which the learners were exposed. Pienemann merely hints that the instruction was of the audiolingual kind, involving drills and dialogues. Also, we are told nothing of the individual learners' participation in the instruction. It is possible, therefore, that the differences in the results obtained for the two learners reflected not so much their current levels of development as the extent to which they individually responded to the classroom events. As Allwright (1988) observes, 'each lesson is a different lesson for each learner'. In short, what is missing from Pienemann's study is a process-element.

Research directed at investigating the effects of grammar teaching on L2 learning is promising. This is an area in which experimental studies of the kind carried out by Pienemann are particularly welcome. Through such studies we are likely to refine our understanding of the constraints that govern formal language instruction — developmental and otherwise. However, this research needs to become more sensitive to a number of factors before it is likely to be taken seriously by teachers. First, there is a need to replicate studies, varying the form of the 'treatment' in order to discover whether the way in which the instruction is conducted affects learning outcomes. Second, as Long (1980) proposed, it is essential to build a process-element into the design in order to discover how learners react to the instruction. Third, it is necessary to recognise that the effects of instruction may not be immediate; instruction provided at one time may have a delayed effect, bearing fruit weeks or even months later (Ellis, forthcoming). It is perfectly feasible to design a research project that takes these variables into account and it is surprising, perhaps, that no such study has taken place.

The empirical study of form-focused instruction has proved much easier to design than the study of meaning-focused instruction. The second question that has motivated psycholinguistic studies of classroom language learning remains largely unanswered. There is a strongly articulated theory, but almost no research that addresses the central premise of the theory.

The theory states that learners construct their interlanguage with the help of input obtained from meaning-focused interaction (Krashen, 1985; Long 1983a; Allwright, 1984). It is also hypothesised that certain kinds of interaction facilitate acquisition to a greater extent than others. Long, for instance, claims that communication containing high frequencies of interactional adjustments (eg requests for clarification and comprehension checks) brought about by the need to negotiate meaning in conversational exchanges where a problem of understanding arises is especially likely to promote acquisition. Hatch (1978), in a review of case studies of naturalistic L2 acquisition, claims that grammatical structures are internalised as a result of learning how to communicate. She suggests that restrictions regarding what is conversationally possible at any stage of development regulate the frequency with which grammatical forms appear in the input and so influence what is learnt.

The evidence which Hatch provides to support her hypothesis is of a qualitative, illustrative nature. She analyses protocols of native-speaker/non-native speaker conversation to show how specific linguistic forms are modelled for the learner and how conversations enable learners to construct grammatical structures, first vertically and then horizontally. Ellis (1984a; 1985) adopts a similar procedure when using data collected in an ESL classroom in a London Language Unit. This approach, which is 'exploratory-interpretative' in nature, sheds light on how interaction contributes to learning but does not demonstrate a necessary connection between the two. It fails to show that interaction *causes* learning to take place.

One way of investigating the link between interaction and learning is to be found in the L1 acquisition research of Wells (1985). Wells examined the interactions of a number of parent-child dyads in order to discover what relationship, if any, existed between specific interactional features of the parents' speech and the children's rates of acquisition. He found a strong and consistent correlation between a number of measures (eg directives) and acquisition. Intriguingly, Wells also found that parents increased the frequency of specific linguistic forms (eg auxiliaries) in their input just before these forms began to appear in their children's speech — almost as if they anticipated their acquisition! It is worthwhile asking whether the paired-subjects design Wells used for studying L1 acquisition can be adapted for classroom L2 acquisition. Clearly a classroom with its one-to-many participant setting is very different from the one-to-one setting of the home. An equivalent design would have to compare the differing interactional features of *classrooms*. But there are obvious drawbacks to such a proposal. First, the method studies suggest that global comparisons of the kind that would be needed do not work. Second, even if it can be demonstrated that two (or more) classrooms differ interactionally to a significant level there remains the problem already mentioned — the interactional environment found in any single classroom may not be the same for all the learners within it. Whether for these or other reasons, there has been no L2 classroom study equivalent to Wells' research.

If it is not possible to investigate the connection between interaction and acquisition by direct means, the alternative is to try to do so indirectly. Long (1985) outlines how this can be done in three stages:

- Step 1:* Show that (a) linguistic/conversational adjustments promote (b) comprehensible input.
- Step 2:* Show that (b) comprehensible input promotes (c) acquisition.
- Step 3:* Deduce that (a) linguistic/conversational adjustments promote (c) acquisition.

There have been a number of studies which have investigated whether linguistic and/or conversational adjustments facilitate comprehensible input, but none, to the best of my knowledge, that have examined whether comprehensible input promotes acquisition. This is not surprising, as it is not clear how such studies can be designed. Krashen's

(1981) contention that comprehensible input containing  $i + 1$  enables the learner to progress to the next stage in the acquisitional sequence is probably not testable, as a number of commentators have argued (Gregg, 1983; White, 1986). Therefore, Long's proposal for an indirect approach to studying the relationship between meaning-focused interaction and acquisition is of doubtful value because of the difficulty of accomplishing Step 2.

There are other ways of proceeding, however. Slimani (1987) undertook a detailed analysis of a small number of lessons involving a group of Algerian students of L2 English. She asked the students to complete 'uptake charts' immediately after the end of each lesson and again a few hours later. These charts required the students to enter anything 'new' that they thought they had learnt from the lesson. Slimani then looked through transcripts of the lessons, trying to discover what features of the interaction appeared to prompt learners to record 'uptake'. She considered a number of features — interactional adjustments involved in the negotiation of meaning, amount of learner participation and topicalisation. Only the latter proved to be significant. Learners were likely to record the uptake of items that occurred in interactional sequences whose topic had been nominated by other students (rather than by the teacher or themselves). This study provides an ingenious solution to the problem of how to investigate the relationship between interaction and learning. It raises a number of questions, however — not least to do with the measure of the dependent variable. We cannot be certain how valid and reliable a measure of acquisition 'uptake' is. Slimani does not cross-validate it by obtaining more conventional measures of acquisition. An analysis of the kinds of items typically reported by the learners in her study shows that they are mostly lexical rather than grammatical. The use of such an approach may be limited to the study of vocabulary learning, therefore. However, this is a worthy attempt at solving an intractable research problem and merits replication. Its great strength is that it treats classroom learners as individuals by recognising that the same interactional event can have different outcomes for different learners.

Not all research based on the focus on form *vs* focus on meaning distinction has investigated the acquisition of new linguistic knowledge. There is a considerable body of research that has looked more generally at whether instruction of one kind or the other results in more rapid acquisition or higher levels of achievement. Long (1983b) surveyed eleven studies that had sought an answer to this question by comparing the levels of achievement of classroom learners with that of naturalistic learners. Long argued that if it could be demonstrated that classroom learners learnt more rapidly and progressed further than naturalistic learners this would indicate that focusing on form was beneficial for acquisition. He assumed that classroom learners would be required to focus on form while naturalistic learners would experience meaning-focused communication. The results of his survey lent support to the



claim that form-focused instruction aids acquisition. This approach, however, cannot provide conclusive evidence, as it fails to control for potentially confounding variables such as motivation. Learners who enrol for language classes may simply be more motivated than those who are content to do their learning in the street.

Hulstijn (forthcoming) tries to evaluate the differential effect of instruction directed at form and meaning experimentally. The pre-test, which Hulstijn used, consisted of a sentence-copying task, ie subjects were shown single sentences, all of which reflected a single underlying grammatical structure. They were asked to write the sentences down. The subjects were then assigned to one of three treatments involving exposure to twelve sentences containing the target structure. One group (the Form-focused Group) had to perform an anagram task that directed their attention to word-order without any need to consider meaning. The second group (the Meaning-focused Group) were shown the same sentences one by one on a screen and asked to respond meaningfully to them by saying 'Yes', 'Perhaps' or 'I don't know'. The third group (the Form and Meaning-focused Group) were simply told to pay attention to both form and meaning but were given no actual task to perform. The post-test consisted of two tasks — a sentence-copying task similar to the pre-test and a task requiring cued recall of the same sentences used in the learning task. Hulstijn calculated the gain scores for the three groups. The results gave support to the hypothesis that 'for implicit and incidental learning of structural language elements to take place, attention to form at input is a sufficient condition'. However, although the Form-focused Group outperformed both the Meaning-focused Group and the Form and Meaning-focused Group, the other two groups also showed improvement. In other words some learning of formal elements took place even when the focus was entirely on meaning — contrary to Hulstijn's expectations.

This study has been described at some length because it is a good example of how experimental studies can be designed to test hypotheses based on a psycholinguistic theory. The study is of interest here, however, because it raises the thorny question of the validity of using evidence obtained from experimental studies of this kind to evaluate pedagogic options. There are a number of reasons for exercising caution. First, the setting of the experiment was not a classroom; there can be no guarantee, therefore, that the subjects' responses are representative of normal classroom behaviour. Second, the measures of learning (ie the sentence-copying and cued-recall tasks) do not tap the kind of language use which teachers might recognise as the goal of instruction. In other words, because the experimental conditions are so remote from normal classroom conditions, extrapolation to teaching is not warranted. This is not intended as a criticism of Hulstijn as he was concerned only with theory-testing and made no attempt to apply the results to pedagogy. It is a general warning against such application.

This review of the methodological procedures which have character-

ised research based on 'psycholinguistically valid design features' does not afford great confidence in the capacity of this approach to supply reliable and valid information upon which pedagogic decision might be based. One is left with the impression of a number of interesting and, in some cases, ingenious attempts to overcome the research problems, but it is hard to point to a single study that would satisfy the discerning teacher.

To summarise, the studies suffer from one or more of the following:

1. a failure to supply information regarding the classroom processes which are involved in instruction;
2. a failure to obtain information regarding how the classroom learners responded to the instruction;
3. a failure to employ measures of learning that adequately reflect the kind of language use teachers recognise as the goal of language teaching;
4. a failure to recognise that instructional events are not monolithic, but often mean different things to different learners;
5. a failure to control for potentially confounding variables;
6. difficulty in operationalising a design feature.

Many of these weaknesses can be redressed by better-designed studies, in particular by incorporating a process-element through classroom observation and/or introspective accounts collected from learners and teachers. However, as I shall argue later, even better-designed studies based on psycholinguistic design features will not provide a sufficient basis for pedagogic decisions. First, though, I shall consider the alternative approach.

### **The educational/pedagogic approach**

The last few years have seen a growth in research based on the educational/pedagogic approach. Long (1987) identifies a number of issues that have attracted the attention of researchers. These include:

- the effects of teacher question types on student production.
- language use in lock-step and small-group work.
- the relationship between practice and achievement.
- teachers' treatment of error.
- the effect of task type on learner production.

In the case of some of these issues there is a 'received opinion' as to what constitutes sound pedagogy. For example, there is general agreement among teacher trainers that small group-work should be encouraged. Other issues, however, are more controversial. As Larsen-Freeman (1988) observes:

. . . it is the case that for almost every theoretical principle upon which some current methodological practice is based, there exists a contrary principle underlying some other current methodological practice.

Irrespective of whether there is received opinion or disagreement, the close study of key instructional procedures is potentially rewarding. It provides a means of submitting what has become conventional practice to critical evaluation or it can provide information that may assist in resolving a controversy.

Empirical research based on the educational/pedagogic approach — like all empirical research — can take two basic forms (Grotjahn, 1987). It can be analytical-nomological and involve experimental or quasi-experimental designs, the collection of quantitative data and statistical analysis. Alternatively, it can be exploratory-investigative, in which case qualitative data collected in authentic classroom settings is submitted to interpretative analysis. Long (1988) clearly favours the analytical-nomological paradigm, arguing that classroom research should take the form of ‘true experiments’. I shall argue the case for a more balanced approach that draws on both research styles.

A good example of pedagogically-based empirical research is that which has examined to what extent practising language forms aids their acquisition. Ellis (1988b) has reviewed a number of studies in the analytical-nomological tradition which have investigated this issue. His review demonstrates that the results are surprisingly mixed and conflicting. There are studies which show a positive effect for practice (eg Naiman et al, 1978), there are studies which show that amount of practice is inversely related to acquisition (eg Ellis, 1984) and there are studies that show practice has no effect one way or the other. In other words, the research does not help teachers to decide whether to engage their learners in language practice.

There are a number of reasons why this analytical-nomological research is so inconclusive:

1. ‘Practice’ is, in fact, a vague construct. It means different things to different teachers. Not surprisingly, researchers have operationalised the construct in different ways.
2. As noted in an earlier section of this paper, ‘acquisition’ can also mean different things. Researchers measure ‘acquisition’ in different ways.
3. Most of the studies examined the relationship between overall practice and overall acquisition rather than investigating whether practice in the production of a specific linguistic feature led to the acquisition of that feature.

Doubtless it would be possible to design a ‘true experiment’ that would overcome these weaknesses. There would still be an additional problem, however:

4. The research is predicated on the assumption that practice causes acquisition. However, it is equally possible to argue that it is acquisition (what the learners know) that determines who receives opportunities for practice.

The fact is that ‘practice’ is an interactive event and is subject to a host of social and personal factors that influence how such events are enacted in the classroom. Analytical-nomological research (even if conducted as a ‘true experiment’) will always have difficulty in taking such factors into

consideration — if only because we have only a very hazy idea of what they consist of.

It is for this reason that exploratory-investigative research is essential. It serves as the means of determining what classroom behaviours occur when an instructional technique such as 'practice' takes place. It enables the researcher to uncover the factors that appear to influence these behaviours. Only if such information is provided can valid experimental studies be designed. Educational/pedagogical research calls for a combination of exploratory and experimental methods.

The testing of educational principles and pedagogical practices is an on-going one. It is part of the history of language teaching. Teachers do not *need* empirical research to evaluate how they teach. Informal evaluation occurs as part and parcel of everyday teaching; it is the means by which teachers, individually and collectively, build up experience as to what 'good teaching' consists of. Empirical research should not be seen as a substitute for informal evaluation. Rather it should seek to complement it. The advantage of the educational/pedagogic approach, whether carried out as action research by teachers-cum-researchers or by trained researchers, is that it is based on constructs (such as 'practice') which are immediately recognisable by teachers. For this reason, the gap between formal research and informal evaluation can be minimised. Teachers and researchers, so often distrustful of each other, can work collaboratively within such an approach.

## **Conclusion**

The last twenty years have seen a remarkable growth in empirical research of classroom language learning. As I have tried to show, this research has been broadly of two kinds. One kind — the linguistic/psycholinguistic — is theory-led; it seeks to test hypotheses drawn from a linguistic or psycholinguistic theory. This branch of classroom research is sometimes dismissed as irrelevant to the needs of teachers. This is, however, a mistaken viewpoint. I have shown how it can contribute to our understanding of how instruction relates to L2 acquisition. The advantage of this kind of research is that the theoretical constructs upon which it is based are relatively explicit. The other kind — educational/pedagogic research — has tried to examine to what extent instructional procedures such as practice or error correction contribute to L2 acquisition. Frequently such procedures are difficult to operationalise in terms of actually occurring classroom behaviours. For this reason exploratory research is needed to establish valid and reliable descriptors and to uncover causal factors. The advantage of this kind of research is that it has an immediate appeal to teachers because it addresses issues with which they are familiar.

What has all this research offered the teaching profession? Chaudron (1988) is doubtful whether any firm pedagogical proposals can be based on the results which have been obtained to date. He comments:

Despite the obvious increase in the amount of classroom-oriented research in recent years, few of the suggestions offered here can be made with great confidence, for the existing research is difficult to synthesize (p. 180).

It is not only a problem of synthesis, however. As we have seen, much of the *best* research raises almost as many questions as it answers. Chaudron lists four areas of weakness in the existing research:

1. the lack of consistent measures of classroom processes and products.
2. poor research design.
3. inexplicit or incomplete quantitative or qualitative analysis.
4. the failure to provide a 'theoretical specification' of the constructs to be investigated.

Chaudron's view is that as we learn to do better research we will be better able to advise decision-makers about which pedagogical procedures work best. No one would wish to dispute the need for 'good research' and I have suggested a number of ways in which the existing research can be improved. In particular, I have emphasised that researchers need to recognise that individual learners respond to the same classroom events differently and that they need to treat instruction as a social and personal phenomenon and not as a set of monolithic constructs, whether these be derived from a linguistic/psycholinguistic theory or a language teacher's handbook. Researchers need to pay attention to process, both in the sense of the unfolding of classroom interaction and in the sense of how learners react cognitively and affectively to specific instructional events. The value of classroom observation as a research tool is now firmly established (Allwright, 1988). The contribution to be made by the intro/retrospective methods is described in Faerch and Kasper (1987). Such methods are crucial as they provide the only effective means of discovering how the individual learner reacts to the instruction she experiences. Researchers concerned with language pedagogy need to consider how they can measure 'acquisition' in a way that reflects the recognised goals of language teaching. Research that does not do so cannot expect to be taken very seriously by teachers. Finally, researchers need to undertake longitudinal case studies of the kind that proved so insightful in the early work on naturalistic acquisition. Allwright (1980) noted the absence of such studies in classroom research. The position is still very much the same ten years later. Researchers who pay attention to all these factors are more likely to produce results that will assist language pedagogy.

Nevertheless, I do not share Chaudron's positivistic view of the role that research can play in language pedagogy. I remain sceptical whether 'true experiments' will produce the definitive answers that some researchers expect — even if they are designed rigorously with due attention to all the factors mentioned above. There are two principal reasons for my scepticism. First, the instruction-learning relationship is a complex one. It is a variable relationship, probably curvilinear rather

than linear (Politzer, 1970). Success, therefore, can be achieved in many ways. Experimental research will provide 'piecemeal understanding' (Allwright, 1988) and contribute to theory-building, but it will never provide the comprehensive answers upon which pedagogic decisions can be based. Second, a positivistic view of the role of research fails to recognise the intrinsic nature of educational change. Innovation in the classroom can never be just a question of implementing a recommendation derived from research. It is always a process of negotiation, involving the teacher's overall educational ideology, the learners' expectations and preferences and local constraints that determine what is feasible. There is no single pedagogical solution that is applicable to all classrooms.

Research does have an important role to play, however. The results which it provides constitute a valuable source of information which *in conjunction with information from other sources* can be used to reach pedagogic decisions. These other sources include teachers' accumulated knowledge about what is 'good teaching'. Research should never seek to supplant this, although it can stimulate teachers to question it. A pedagogical decision should be seen as a 'hypothesis' regarding what will work best in a particular context. The inputs to the decision-making process will be multiple. The information supplied by classroom research should be one of them.

## References

- Allwright, R. (1980) 'Turns, topics and tasks: patterns of participation in language learning and teaching', in Larsen-Freeman, D. (ed.), *Discourse Analysis in Second Language Research*, Rowley, Mass.: Newbury House.
- Allwright, R. (1983) 'Classroom centred research on language teaching and learning: a brief historical overview', *TESOL Quarterly* 17: 191–204.
- Allwright, R. (1984) 'The importance of interaction in classroom language learning', *Applied Linguistics*, 5: 156–71.
- Allwright, R. (1988) 'The characterization of teaching and learning environments: problems and perspectives', paper given at the conference on Empirical Research on Second Language Learning in Institutional Settings, Bellagio, Italy.
- Chaudron, C. (1988) *Second Language Classrooms: Research on Learning and Teaching*, Cambridge: Cambridge University Press.
- Chomsky, N. (1981) *Lectures on Government and Binding*, Dordrecht: Foris.
- Cohen, A. and Apeh, E. (1981) 'Easifying second language learning', *Studies in Second Language Acquisition*, 3: 221–36.
- Day, R. (1984) 'Student participation in the ESL classroom or some imperfections in practice', *Language Learning*, 34: 69–102.
- Ellis, R. (1984a) *Classroom Second Language Development*, Oxford: Pergamon.
- Ellis, R. (1984b) 'Can syntax be taught? A study of the effects of formal instruction on the acquisition of WH questions by children', *Applied Linguistics*, 5: 138–55.
- Ellis, R. (1985) 'Teacher–pupil interaction in second language development', in Gass, S. and Madden, C. (eds.) *Input in Second Language Acquisition*, Rowley, Mass.: Newbury House.
- Ellis, R. (1988a), 'Investigating language teaching: the case for an educational approach', *System*, 16: 1–11.
- Ellis, R. (1988b) 'The role of practice in classroom language learning', *Teanga*, 8: 1–25.
- Ellis, R. (forthcoming) *Learning Through Instruction: The Study of Classroom Second Language Acquisition*, Oxford: Blackwells.

- Faerch, C. and Kasper, G. (eds.) (1987) *Introspection in Second Language Research*, Clevedon: Multilingual Matters.
- Gregg, K. (1984) 'Krashen's Monitor Model and Occam's Razor', *Applied Linguistics*, 5: 79-100.
- Grotjahn, R. (1987) 'On the methodological basis of introspective methods', in Faerch, C. and Kasper, G. (eds.) *Introspection in Second Language Research*, Clevedon: Multilingual Matters.
- Harley, B., Allen, P., Cummins, J. and Swain, M., (1987) 'The Development of Bilingual Proficiency', Final Report, Volume II: *Classroom Treatment*, Toronto: The Ontario Institute for Studies in Education.
- Hatch, E. (1978) 'Discourse analysis and second language acquisition', in Hatch, E. (ed.) *Second Language Acquisition*, Rowley, Mass.: Newbury House.
- Hulstijn, J. (forthcoming), 'Implicit and incidental second language learning: experiments in the processing of natural and partly artificial input', in Dechert, H. (ed.) *Interlingual Processing*, Tübingen: Gunter Narr.
- Krashen, S. (1981) *Second Language Acquisition and Second Language Learning*, Oxford: Pergamon.
- Krashen, S. (1985) *The Input Hypothesis*, London: Longman.
- Larsen-Freeman, D. (1988) 'Research on language teaching methodologies: a review of the past and an agenda for the future', paper given at the conference on Empirical Research on Second Language Learning in Institutional Settings, Bellagio, Italy.
- Long, M. (1980) 'Inside the "black box": methodological issues in classroom research on language learning', *Language Learning*, 30: 1-42.
- Long, M. (1983a) 'Native speaker/non-native speaker conversation in the second language classroom', in Clarke, M. and Handscombe, J. (eds.) on *TESOL '82: Pacific Perspectives on Language Learning and Teaching*, Washington D.C.: TESOL.
- Long, M. (1983b) 'Does instruction make a difference? A review of the research', *TESOL Quarterly*, 17: 359-82.
- Long, M. (1985) 'Input and second language acquisition theory', in Gass, S. and Madden, C. (eds.) *Input in Second Language Acquisition*, Rowley, Mass.: Newbury House.
- Long, M. (1987) 'The experimental classroom', *The Annals of the American Academy of Political and Social Science*, 490: 97-109.
- Long, M. (1988) 'Focus on form: a design feature in language teaching', paper given at the conference on Empirical Research on Second Language Learning in Institutional Settings, Bellagio, Italy.
- McTear, M. (1975) 'Structure and categories of foreign language teaching sequences', in Allwright, R. (ed.) *Working papers: Language Teaching Classroom Research*, University of Essex, Department of Language and Linguistics.
- Naiman, N., Frohlich, M., Stern, H. and Todesco, A., (1978) 'The Good Language Learner', *Research in Education No. 7*, Toronto: Ontario Institute for Studies in Education.
- Pica, T. (1983) 'Adult acquisition of English as a second language under different conditions of exposure', *Language Learning*, 33: 465-97.
- Pienemann, M. (1984) 'Psychological constraints on the teachability of language', *Studies in Second Language Acquisition*, 6: 186-214.
- Politzer, R. (1970) 'Some reflections on 'good' and 'bad' language teaching behaviors', *Language Learning* 20: 31-43.
- Scherer, A. and Wertheimer, M. (1964) *A psycholinguistic experiment in foreign language teaching*, New York: McGraw Hill.
- Seliger, H. (1977) 'Does practice make perfect? A study of interaction patterns and L2 competence', *Language Learning*, 27: 263-75.
- Slimani, A. (1987) 'The teaching/learning relationship: learning opportunities and learning outcomes. An Algerian case study', unpublished doctoral dissertation, University of Lancaster.
- Smith, P. (1970) *A Comparison of the Cognitive and Audiolingual Approaches to Foreign Language Instruction*; The Pennsylvania Foreign Language Project, Philadelphia, PA.: Center for Curriculum Development.

- Wells, G. (1985) *Language Development in the Pre-school Years*, Cambridge: Cambridge University Press.
- White, L. (1986) 'Against comprehensible input: the input hypothesis and the development of second language competence', *Applied Linguistics*, 8: 95-110.
- Widdowson, H. (1978) *Teaching Language as Communication*, Oxford: Oxford University Press.
- Widdowson, H. (1984) *Language Use and Learning Purpose*, Oxford: Oxford University Press.



# The I-Language Approach and Classroom Observation

Vivian Cook

*University of Essex*

## **I-language and E-language approaches**

This paper starts from a distinction between two approaches to language, to linguistics, to language learning and teaching, and indeed to life. One approach is concerned with the world outside the speaker — language as the possession of society. Its most important aspect is people's relations with each other; it takes the subject matter of linguistics to be large samples of language that people have actually produced; it sees the main function of language as communication; it considers that children learn language by working out the regularities in a sample of speech and by relating to other people; language teaching means providing sufficient data for students to work out regularities and opportunities for them to relate to each other. The other approach is concerned with the world inside the speaker — language as the possession of the individual. The most important aspect is the speaker's knowledge of language; linguistics should study how language is stored in the mind; the functions of language are many, both inside the mind and out; children acquire language by applying their minds to the speech they hear; language teaching means providing evidence for the students' minds to work on. In linguistics this distinction has been called E-language (External language) versus I-language (Internal language) (Chomsky, 1986; Cook, 1988); in terms of British linguistics today a typical E-language approach might be the corpus-oriented view taken by John Sinclair (Sinclair et al, 1987) compared to the I-language approach taken by Gerald Gazdar (Gazdar et al, 1985). Most current approaches to language teaching are E-language in that they concentrate on the provision of language data and interaction with other people and with the world outside rather than on mental knowledge and the world inside the speaker. Also, in L2 learning research, 'strategies' approaches that are inherently E-language may be contrasted with Universal Grammar approaches that are inherently I-language; McLaughlin (1987) makes a parallel distinction between bottom-up inductive theories and top-down deductive theories. Going outside the language area the distinction is paralleled in the Jungian Extrovert versus Introvert personalities, or in

Classical versus Romantic literature, or in Western versus Taoist philosophy; indeed one may find overtones in society-oriented versus individual-oriented positions in politics.

### **Methodology of classroom observation**

Where does classroom research into second language acquisition fall on this dimension? We will confine the discussion here to the use of classroom observation for research; its success as a form of consciousness-raising for teachers in training or for second language students has little to do necessarily with its worth as a method of research. One issue is its attitude to data and research: an E-language approach sees its first task as collecting observations of the world, from which it deduces regularities and rules; an I-language approach seizes on whatever signs can be found of inner states without restricting itself in terms of amount or type of evidence. Classroom research has consistently taken a particular position, advertised indeed even by its name of classroom 'observation': the evidence it admits is based on observable evidence — X students and teachers observed in Y classrooms carrying out Z tasks. Hence the proliferation of schedules for observing 'behaviour', starting from a tradition of research outside the second language acquisition area represented say, by Flanders (1960) or by Sylva et al (1980), within second language acquisition ranging from FLint (Moskowitz, 1967) to FOCUS (Fanselow, 1977) to ECS (Long et al, 1976) to COLT (Allen et al, 1984), or a range of other AFCOs (Acronyms For Classroom Observation). A large proportion of such research is therefore E-language in orientation and so concerned with describing the external world. In its own terms it has to obey the methodological criteria required for E-language observational research, as discussed in Cook (1986): the sample of behaviour needs to be adequate in size and representativeness; the techniques of data collection and recording have to be spelled out explicitly; the systems of analysis have to be capable of being checked and replicated by other investigators; relevant variation between the subjects studied has to be reported. Perhaps crucially, since the results of such studies are typically statements about the proportions and significance of various features of the sample, the figures that are presented have to be treated in a numerically appropriate way and the relevant statistical tests applied. This is not the place to document individual failings in this respect; let Chaudron's summing up suffice:

much classroom research has demonstrated marked misuse, frequent underuse, and occasional unwarranted overuse of various statistical procedures  
(Chaudron, 1988, p. 183).

Suppose, however, that such an approach were totally successful in its own E-language terms, a researcher with an I-language bias may still find the results of comparatively little interest. The main reasons for this fall into two groups — methodological and theoretical. The methodologic-

al reasons derive from the standard objections to corpus-based research; it may not be possible to find a sufficient number of examples of a particular point, however large the sample; frequency of occurrence rather than importance almost inevitably dominates the results. The conclusions that can be drawn are restricted to regularities in the data; going beyond what is visible in the corpus needs other types of evidence, as argued in Chomskyan linguistics for many years. In first language acquisition it has been suggested time and again that crucial aspects of language are not in principle learnable by the child from 'positive' evidence alone, that is to say from samples of actually occurring forms, but must be learnt from other types of evidence, or be part of the mind to start with. E-language research is faced with the same problem as the child; it cannot legitimately go beyond statements of positive regularities found in the data, as argued in Cook (forthcoming, a). Research based on observation is furthermore limited to what can actually be observed; the processes of speech production, of comprehension, indeed of learning itself, are not in themselves observable; the utterances and gestures that *are* observable are only the tip of the iceberg. Basing an account of learning on observation alone means severely limiting what can be said. Similarly the observation of classroom interaction legitimately concerns itself with patterns visible in the classroom — say, the sequences of moves used by students and teachers described in Sinclair and Coulthard (1975). While such moves can be labelled in terms of actual properties of the data such as 'student initiates discussion' or 'student interrupts' (Long et al, 1976), problems arise when the label is more interpretive than descriptive and not directly linked to objective evidence in the data, say, learning strategies such as 'Checking the outcomes of one's own language learning against an internal measure of completeness and accuracy' (Chesterfield and Chesterfield, 1985); several steps beyond this are interpretations of moves such as 'Yeah' in response to '*For for to end the pollution problem?*' as 'Igor therefore shows no special interest in pursuing either the topic or the sequence of turns, given that he could have used his opportunity to develop his earlier statement, or to change topic or 'subject' again' (Allwright, 1988, p. 189, citing Allwright, 1980). This is not to say that speakers do not have such intentions or choices open to them nor that it is not possible to discuss these, but that the motivation for speech moves cannot be established only by the occurrence of certain speech forms; other evidence is necessary, whether introspective report, interview, questionnaire, or experiment. Reading such internal motivations into spoken discourse solely from observational evidence is unjustified in an E-language approach, partly because there is no checkability or reliability for such interpretations; if I insist that Igor showed his great enthusiasm for the topic by saying 'Yeah', which the teacher wilfully ignored, my interpretation is as good as Allwright's, in the absence of further evidence. Assertions about people's intentions are dangerous unless corroborated from other sources. Methodologically it amounts to covert conversion of E-language data into statements about

I-language knowledge, a delicate if not impossible task first perhaps highlighted in Chomsky (1965) as 'a general tendency . . . to assume that the determination of competence can be described from description of a corpus by some sort of sufficiently developed data-processing technique'. Like the related approach of Error Analysis, Classroom Research based on observation finds it difficult to go behind the visible facts, to look for deeper, more abstract, explanations for classroom behaviour. E-language describes language as an external object; to jump from this to internal psychological reality needs a chain of evidence and of argument. Statements about I-language need to be supported by explicit evidence and argument as much as those for E-language; but it is evidence and argument of a different type than sheer occurrence (Cook, 1988).

### **I-language Theory and Classroom Observation**

The theoretical objections to observation as a research technique again come down to the researcher's basic I or E affiliation. The I-linguist is concerned with language as a property of the individual, not as the possession of many people; one may indeed argue that learning is necessarily by individuals rather than groups. This is particularly true in L1 acquisition work where the fact that all children learn their first language successfully means they must have something in common which enables them to carry out the task; the crucial first steps in research are to look at what learners share — the properties of their minds that they all bring to language learning. Much L2 observational research has concentrated on the differences between learners — on how differences of motivation, intelligence, first language ability, etc, facilitate or detract from L2 learning. There may be more justification in this approach with L2 learning than with L1 acquisition as the final competence of L2 students varies far more than that of L1 children. Nevertheless the I-linguist will still say that the crucial first step is to establish whatever learners have in common, however great or small, before going on to their differences. An I-linguist will point to the lack of effects of situation in L1 acquisition; whether parents construct complex interactions with their children, as described for English in Bruner (1983) or whether they ignore them completely as conversational partners, as described for Kahuli by Schieffelin (1985), the children still acquire normal competences; to quote Gleitman (1984):

Under widely varying environmental circumstances, learning different languages, under different conditions of culture and child rearing, and with different motivations and talents, all non-pathological children acquire their native tongue at a high level of proficiency within a narrow developmental timeframe.

The first task with L2 acquisition is to establish the features common to all L2 situations; once these have been discovered, the research can go on

to the differences between situations, between classrooms and non-classrooms, or between teaching methods. Again it may well be that situational differences loom larger in L2 acquisition; this still means starting from the core features not subject to situational variation. Hence, to an I-linguist concerned with L2 learning, research into variation and into situation is premature; at present they are side-issues that divert attention from the crucial work into what L2 learners have in common, however intriguing the results may be.

Similarly the I-linguist will be little impressed by the investigations of the language addressed to learners — ‘teacher talk’, input, etc. The uniformity requirement for first language acquisition claims that its central aspects cannot be held to depend on some feature of language input which is not available to all learners; in L2 learning it is the central unavoidable aspects of any input that are important — just as it is impossible for any L1 or L2 learner not to hear all these phonemes of the English, so the evidence necessary to indicate the crucial aspects of syntax must be present in virtually all input. Central facts about English — the fact that it has Subject Verb Object order, the fact that it does not permit declarative sentences without overt subjects, the fact that it does not allow related Verbs and Objects to be separated (\*‘Peter likes very much London’) — are all deducible from virtually any handful of sentences and, as discussed in Cook (forthcoming, b), come in the first few lessons of any language teaching course. The well-established features of teacher language — shorter utterances (Wesche and Ready, 1985), less subordination (Ishiguro, 1985), slower speed (Mannon, 1986), and so on — are of marginal importance. I-linguists working within a UG model will indeed stress the importance of input to the learner — without it no learning will take place. But all that is required is minimal examples of the key features of the language to ‘trigger’ learning in the mind. For the essential aspects of language, precise types of sentence, frequency of occurrence or indeed comprehensibility of input in the Krashen sense (Krashen, 1981) are beside the point, the only qualification being that it is desirable for the input to be readily segmentable into grammatical constituents, as UG theory implies (Cook, forthcoming, a; Morgan, 1986).

This paper has then attempted to highlight the extreme I-language position towards L2 classroom observation research; it is an attempt to explain why such research seems peripheral to those studying second language acquisition within I-language frameworks such as the Universal Grammar paradigm. From this perspective, classroom observation is in principle unrevealing because of the deficiencies of positive evidence; it concentrates on differences between learners rather than on their similarities and on the effects of variation in situation rather than on situation-free aspects of learning. From an E-language perspective on the other hand, the research is interesting and suggestive, at least those aspects of it that use an adequately formulated E-language methodology. Needless to say, a pure I-language or E-language approach

never exists, any more than a pure Introvert or Extrovert can be found; second language learning research needs to find an overall framework within which both research paradigms can co-exist rather than denying each other's existence, to explore alternative avenues simultaneously in the Feyerabend approach to science (Feyerabend, 1975). To quote Walt Whitman, 'Do I contradict myself? Very well then, I contradict myself (I am large; I contain multitudes).'

## References

- Allen, J. P. B., Frohlich, M. and Spada, N. (1984) 'The communicative orientation of language teaching: an observation scheme', in Handscomb, J., Orem, R. A. and Taylor, B. (eds.), *On TESOL '83: the Question of Control*, Washington, D.C., TESOL, pp. 231–252.
- Allwright, D. (1988) *Observation in the Language Classroom*, Harlow: Longman.
- Bruner, J. (1983) *Child's Talk*, Oxford: Oxford University Press.
- Chaudron, C. (1988) *Second Language Classrooms: Research on Teaching and Learning*, Cambridge: Cambridge University Press.
- Chesterfield, R., and Chesterfield, K. B. (1985) 'Natural order in children's use of second language learning strategies', *Applied Linguistics*, 6, 45–59.
- Chomsky, N. (1965) 'Formal discussion: the development of grammar in child language', in Bellugi, U. and Brown, R. (eds.), *The Acquisition of Language*, Indiana: Purdue University.
- Chomsky, N. (1986) *Knowledge of Language: Its Nature, Origin and Use*, New York: Praeger.
- Cook, V. J. (1986) (ed.) *Experimental Approaches to Second Language Learning*, Oxford: Pergamon.
- Cook, V. J. (1988) *Chomsky's Universal Grammar: An Introduction*, Oxford: Blackwell.
- Cook, V. J. (forthcoming, a), 'Setting the head parameter: observational evidence and the UG theory of language acquisition', in Roca, I. (ed.), *Logical Issues in Language Acquisition*, Foris.
- Cook, V. J. (forthcoming, b), 'Universal Grammar theory and the classroom, *System*.
- Fanselow, J. F. (1977) 'Beyond 'Rashomon' — conceptualizing and describing the teaching act', *TESOL Quarterly*, 11, 17–39.
- Feyerabend, P. (1975) *Against Method*, London: Verso.
- Flanders, N. A. (1960) *Interaction Analysis in the Classroom: a Manual for Observers*, Ann Arbor: University of Michigan Press.
- Gazdar, G., Klein, E., Pullum, G., and Sag, I. (1985) *Generalised Phrase Structure Grammar*, Oxford, Blackwell.
- Gleitman, L. (1984) 'Biological predispositions to learn language', in Marler, P. and Terrace, H. (eds.) *The Biology of Learning*, Springer.
- Ishiguro, T. (1986) *Simplification and elaboration in foreign language teacher talk and its source*, Ph.D., Stanford.
- Krashen, S. (1981) 'The Fundamental Pedagogical Principle in Second Language Teaching', in Sigurd, B. and Svartvik, J. (eds.), *AILA 1981*, CWK Gleerup.
- Long, M. H., Adams, L., McLean, M. and Castanos, F. (1976) 'Doing things with words: verbal interaction in lockstep and smallgroup classroom situations', in Fanselow, J. F. and Crymes, R. (eds.), *On TESOL 1976*, 137–153, Washington, D.C.: TESOL.
- McLaughlin, B. (1987) *Theories of Second-Language Learning*, London: Edward Arnold.
- Mannon, T. M. (1986) *Teacher-talk: comparison of a teacher's speech to native and non-native speakers*, MA thesis, UCLA.
- Morgan, J. L. (1986) *From Simple Input to Complex Grammar*, MIT Press.
- Moskowitz, G. (1967) 'The FLint system: an observational tool for the foreign language classroom', in Simon, A and Boyer, E. G. (eds.) *Mirrors for Behavior: An Anthology of*

- Classroom Observation Instruments*, section 15, pp. 1–15, Philadelphia, Center for the Study of Teaching, Temple University.
- O'Malley, J. M., Chamot, A. U., Stewner-Manzares, G., Kupper, L., and Russo, R. P. (1985) 'Learning strategies used by beginning and intermediate ESL students', *Language Learning*, 35, 21–46.
- Schieffelin, B. B. (1985) 'The acquisition of Kahuli', in Slobin, D. I. (ed.) *The Crosslinguistic Study of Language Acquisition: Volume 1*, LEA, New Jersey.
- Sinclair, J. M. et al (1987) *Looking Up: an Account of the COBUILD Project in Linguistic Computing*, London: Collins.
- Sinclair, J., and Coulthard, M. (1975), *Towards an Analysis of Discourse*, OUP.
- Sylva, K., Roy, C., and Painter, M. (1980) *Childwatching at Playgroup and Nursery School*, London: Grant McIntyre.
- Wesche, M. B., and Ready, D. (1985) 'Foreigner talk in the university classroom', in Gass, S. M., and Madden, C. G. (eds.), *Input in Second Language Acquisition*, Rowley, Mass.: Newbury House.





## ***II Specific Approaches***



# A Look at the Research Process in Classroom Observation: A Case Study

Nina Spada

*McGill University*

## Introduction

This paper is about doing long-term classroom observation research in second language (L2) classrooms.<sup>1</sup> Its purpose is to describe, using a case study, some aspects of the process that researchers go through in their attempts to describe classroom behaviours and link them to learning outcomes. This kind of research is referred to as process-product research; process referring to descriptions of classroom behaviours and product to learning outcomes. Process-product research has a long tradition in first language (L1) education research (see Dunkin and Biddle, 1974; Rosenshine and Furst, 1973, and Brophy and Good, 1986, for reviews of this research). However, it is only within the last decade that process-product research has become prevalent in the field of L2 teaching and learning. Prior to this, most L2 classroom research focused on either the product, as evidenced in the global method comparison studies in the 60s (Scherer and Wertheimer, 1964; Smith, 1969) or on the process, as seen in the proliferation of L2 observation schemes in the 70s (see Allwright, 1988; Chaudron, 1988, and Long, 1980, for reviews of L2 observation schemes). Few studies focused on both.

A strong call for L2 classroom research to examine both process and product and the relationship between the two was made by several researchers in the 1980s (Long, 1980, 1984; Gaies, 1983; Allwright, 1983). Since that time, a number of process/product studies have been undertaken. Some of this research has been undertaken within the context of naturally occurring L2 classrooms over extended periods of time. These studies have typically used multi-category observation instruments to quantify different aspects of instruction in order to

---

1 The research reported here was funded by the Fonds pour la formation de chercheurs et l'aide à la recherche (FCAR) of the Quebec Ministry of Education. Patsy M. Lightbown of Concordia University has been the co-principal investigator in this project. I thank the teachers and students who cooperated so fully and the wonderful team of research assistants who made it all possible.

examine relationships between instructional behaviours and learning outcomes measured by correlational or multivariate analyses (see Allen, Carroll, Burtis and Gaudino, 1987; Mitchell, Parkinson and Johnstone, 1981; McDonald, Stone and Yates, 1977; Spada, 1987). Other process-product research has been either experimental in nature – taking place in controlled laboratory settings to investigate the effects of a specific instructional feature on learning such as error correction and input modification, or has been quasi-experimental in nature – isolating and/or manipulating particular features of instruction in naturally occurring classrooms to examine their effects on learning (see Chaudron, 1988 for a review of different methodological approaches and studies in L2 classroom-centred research).

In a widely-cited article reviewing classroom observation research in L1 education, Rosenshine and Furst (1973) have identified three elements in a model for studying teaching and learning:

1. development of procedures for describing teaching in a quantitative manner;
  2. correlational studies in which the descriptive variables are related to measures of student growth;
  3. experimental studies in which the significant variables obtained in the correlational studies are tested in a more controlled situation.
- (Rosenhine and Furst, 1973, p. 122)

Although the authors state that these three elements are a minimum requisite for research on teaching effectiveness, they emphasise that the sequence of these steps is not fixed. That is, just as it is possible to use the results from correlational studies of student achievement and instructional variables to set up experimental studies, it is also possible to use the results of an experimental study to create new observation categories or instruments. These stages are not intended to be isolated from each other either, but rather ‘research at each step should influence modifications of research at other steps’ (Rosenhine and Furst, 1973, p. 123).

In this paper, I will describe the process of carrying out a long-term process-product study in L2 classrooms with reference to the three elements proposed by Rosenshine and Furst. My intention is to discuss how we have made decisions regarding the selection and implementation of the three elements of description, correlation and experimentation, and how the interaction between them has guided approaches to data analysis, hypothesis generation and future research designs. In referring to the first element in the research loop, a brief description of the observation scheme and the results obtained through its use will be provided along with a discussion of how macro-level activities-based analyses informed subsequent micro-level linguistic analyses. In a discussion of the interdependent/dynamic nature of the three elements, a description of how product data influenced directions for further process data analysis will be provided. Finally, a discussion of how a ‘natural

experiment' concerning the effects of instructional input on learning outcomes has shaped the direction for future experimental studies within this research will be provided.

### **The research context**

In 1986, we became involved in a three-year research project designed to investigate the classroom interaction and the L2 development of children enrolled in experimental intensive English second language (ESL) programs in several French-language primary schools in the province of Quebec in Canada. These are programs in which students in grades 5 or 6 receive intensive English language instruction for five hours a day, five days a week for five months of the school year. In the remaining half of the school year, students receive their regular subject matter instruction (eg math, science, language etc) in French. These intensive ESL programs contrast strikingly with the regular ESL programs in Quebec's primary schools which usually provide less than 120 minutes a week of ESL instruction spread over the entire school year. The impetus for the intensive programs came from francophone parents who were concerned that their children were not obtaining sufficiently high levels of English language ability in the regular ESL program. This parental pressure has led to the creation of a number of experimental intensive programs in several different schoolboards in recent years. Although parents, schoolboard administrators, teachers and students have informally expressed a great deal of satisfaction with these programs, there is a clear need for systematic research to evaluate their effectiveness.

Our research project is intended to respond to this need by investigating the L2 development of children in these programs, the characteristics of classroom instruction and learners' contact with and attitudes toward English. So far, we have collected learner language (product) data from approximately 1000 students in 33 classes at the grade 5 and 6 levels in eight schoolboards.<sup>2</sup> We have also collected classroom observation (process) data from twenty classes. This paper focuses on the classroom observation results from eight classes in the first year of our research. (See Lightbown and Spada, 1987, in prep; Spada and Lightbown, 1988, for reports on various aspects of this research.)

It is important to note that all intensive programs are based on the philosophy of the Ministry of Education of Quebec (MEQ) ESL program which emphasises the development of fluency over accuracy, student-centred rather than teacher-centred interaction, function over form, and communicative rather than grammatical practice. However, because of the experimental nature of these programs, each schoolboard has developed and implemented its program independently. As a result,

---

2 We have also collected data from two comparison groups: learners in regular ESL programs at the same grade levels as well as learners who have received a comparable number of hours of instruction over a longer period of time.

while there has been some limited consultation, there has been little ongoing contact between program administrators and teachers in these separate intensive programs regarding student selection, materials development, methodological procedures, teacher training and program evaluation. There has been even less contact about the day-to-day activities of the classes. This made it essential for us to include in the study a classroom observation component which could describe with some accuracy the characteristics of classroom instruction in these various classes and to investigate whether potential differences in classroom instruction might be related to potential differences in learning outcomes.

### **Classroom observation procedures and instruments**

Eight intensive program classes were observed during the 1986-87 school year. Each class was observed four times during the five month intensive session resulting in a total of 32 observation visits. The first observation took place approximately one week after classes had begun and the remaining three were carried out at approximately four week intervals. Each observation began as students arrived in the morning and continued until the end of the school day. This usually represented about five hours of observation per visit. During these visits an observer sat in the classroom and coded for specific features on the observation scheme used for this study. This observation instrument, referred to as the COLT (Communicative Orientation of Language Teaching) scheme, is a modified version of the original scheme which was designed to capture differences in the communicative orientation of L2 instruction (Allen, Fröhlich and Spada, 1984; Fröhlich, Spada and Allen, 1985, and Spada, 1987, in press).

The scheme consists of two parts: Part A (see Appendix 1) which describes classroom activities in organisational and pedagogical terms and Part B which describes the verbal interactions which take place within activities. Part A contains five major categories: Activity Description, Participant Organisation, Content, Student Modality and Materials. These categories and their subsections are designed to describe classroom activities in order to investigate such aspects as whether the classroom organisation is student-centred or teacher-centred, whether the instructional focus is on code (ie language form) or meaning and whether students are using pedagogical materials in which extended or minimal texts predominate.

Part A of the COLT scheme is a 'real time' coding scheme and is used during the classroom observations by one observer. As the observer is coding for the Part A features, the classroom events are audio-recorded in two ways. A microphone is placed on the teacher's desk to record verbal interaction going on in the whole class. In addition, the teacher wears a small tape-recorder around her neck to pick up all teacher-talk as well as any interaction she has with individual students or a group of students.

Part B, which is a post-hoc analysis of the audio-taped recordings, is used to analyse classroom activities at the level of verbal interaction on the part of teachers and students. It measures such features as the use of the target language, the extent to which learners are given opportunities to produce language without teacher-imposed linguistic restrictions, to engage in sustained speech, to exchange known as opposed to relatively unpredictable information and the extent to which teachers and students respond and react to code/accuracy-related aspects of the L2. (For a more detailed description and rationale of the Part A and B categories as well as information on the coding procedures, see Allen, Fröhlich and Spada, 1984, and Fröhlich, Spada and Allen, 1985.)

### **Classroom interaction outcomes: Macro-level process analysis**

In carrying out the analysis of the observation data, we decided to first conduct a macro-level analysis of the activities from the Part A coded classroom data. This would enable us to see if there were any differences among the classes at the level of participant organisation, content/language focus and modality and materials use. If differences emerged, a micro-level analysis of the linguistic interaction among teachers and students would be undertaken using all or a portion of the categories on Part B of the COLT scheme. The precise nature of the micro-level analysis would be determined on the basis of the results of the macro-analysis. Another factor which we knew might also point to the need for a more detailed classroom interaction analysis was the pre/post-test results of learners' performance on the receptive and productive measures. That is, if learners in one of the classes were performing significantly better (or worse) than learners in another class on a particular measure or in a particular aspect of their language ability, this would also suggest that a closer examination of the classroom observation data might yield some explanatory information.

The macro-level analysis involved a calculation of the amount of time that teachers and students spent on the various categories and sub-categories on Part A of the scheme; for example, how much time students spent in group work or teacher-centred work, how much time was spent with an explicit focus on language (ie vocabulary, grammar), how much time was spent listening, speaking, reading etc. These features and others were calculated for each observation visit per class and the amount of time spent on each feature was then totalled for individual classes. The results of this analysis indicated that all eight classes were similar to each other in terms of most features. For example, in the participant organisation category, analysis showed that instruction was teacher-centred for about 50% of the time for all classes. The remainder of the time was spent on a variety of student-student, student-class, group-class and individual activities. Although there were

some individual classroom differences, these were quite minimal.<sup>3</sup> In terms of the student modality category, the analysis indicated similarities across classes as well. For example, all classes spent most of their time primarily listening to the teacher or to other students (approximately 45% of the time) and about 14% of the time on a combined listening/speaking focus. The rest of the time was spent on various skill combinations (eg listening, speaking and reading or reading and listening). In terms of the type of materials used, there were some differences among the classes. For example, materials in which extended texts predominated were used more often in some classes compared with others. We intend to investigate these differences in a more detailed analysis of the pedagogic materials. Finally, the results of the content/language category analysis indicated that there were both similarities and differences across classes. For example, all classes spent approximately 50% of the time focusing exclusively on meaning and comparatively less time (approximately 14%) focusing on the code-related aspects of the language. These figures are consistent with the MEQ program philosophy which emphasises a focus on meaning-based instruction rather than grammatical instruction and practice. However, there were some individual class differences in the amount of time spent on code-focused instruction which are worth noting. Table 1 presents these results for each class. As indicated, classes 3 and 7 spent the most time on code (33% and 29%) and classes 4 and 8 the least (10% and 11%). Classes 1 and 2 spent about the same amount of time on code-related instruction (22% and 25%) as did classes 5 and 6 (14% and 13%). Although these differences in code-related instruction were noteworthy, we decided not to undertake a micro-level analysis of the classroom interaction until the results of learners' performance data were complete.

*Table 1* Percentage of instructional time with a focus on code (grammar, vocabulary, phonology) by class: 1986-87 cohort

<i>Class 1</i>	25.40	<i>Class 5</i>	13.49
<i>Class 2</i>	22.56	<i>Class 6</i>	13.42
<i>Class 3</i>	32.96	<i>Class 7</i>	28.77
<i>Class 4</i>	9.61	<i>Class 8</i>	10.73

### **Learner language outcomes: Product analysis**

As indicated above, there were two factors which we anticipated would point to the value of a micro-level analysis of the classroom data: 1)

<sup>3</sup> For more details of the results of this analysis, and some preliminary results of the classroom interaction analysis from the second year of our research, see Lightbown and Spada, 1987, Spada and Lightbown, 1988 and Spada, in press.



macro-level results which indicated differences in instruction and/or 2) class differences in learners' outcomes on the language tests (ie pre/post-test listening and reading comprehension and performance on an oral production task). Since the macro-level process analysis indicated some differences in the extent to which teachers focused on grammar, this seemed a useful area to explore in micro-level analyses. Such a decision would be reinforced if there were indications in the results of the product data analysis that some classes were outperforming others on the proficiency measures or were more accurate than others in their use of particular grammatical forms. An examination of the intensive program learners' performance on the tests of listening and reading comprehension indicated no differences that were related to individual classroom instruction. That is, all intensive program learners improved significantly on these receptive measures and dramatically outperformed most learners in regular ESL programs at the same grade level (ie grades 5 and 6) as well as learners who had received a comparable amount of instruction over a longer period of time (ie grades 9 and 10). There was no indication in an analysis of variance that learners in some intensive classes were doing significantly better than learners in other classes and that this might be related to the kind of instruction they received.<sup>4</sup> However, when we began to analyse the learners' oral performance on a communication task, some interesting results emerged which necessitated a closer look at the classroom observation data.

In an analysis of the learners' performance on an oral communication task involving picture descriptions, we discovered that an overwhelming majority of learners used the verb 'have' rather than the verb 'be' in their introducer forms for the picture descriptions. For example, when describing a picture of a classroom with children in it, most of the learners would start by saying 'We have a classroom and we have five children', instead of saying 'There's a classroom and there are five children.'<sup>5</sup> However, we were surprised to discover that the learners in one of the intensive classes did not tend to use the 'have' introducer form as frequently as the others. Instead, most of these learners used 'be' forms to describe the pictures most of the time (see Lightbown, in press, for further details).

Naturally, we were curious to know why so many of the learners in this class were using the 'be' introducer form when learners from the other classes were not. Our first 'clue' was in the macro-level classroom

---

4 The only within-intensive-program difference was that the grade 6 learners significantly outperformed the grade 5 learners on the listening and reading comprehension measures. It should be noted, however, that in subsequent analyses of the data from the second year of our research which represents a much larger number of classes, significant differences have emerged within the intensive ESL group.

5 An explanation for the incorrect use of 'have' in these introducers appears to be directly related to the transfer of the French 'il y a' introducer form which uses the verb 'avoir' (ie have), although other factors may also be involved. See Lightbown, 1984, 1987, for further discussions of this.

interaction results which had indicated that these learners were from one of the classes in which the greatest amount of time was spent on code (Class 7 in Table 1). However, because the macro-level analysis did not provide us with the specific nature of the code-focused instruction (other than whether it was vocabulary, phonology, or grammar), it seemed necessary at this point to do a micro-level analysis of the code-related behaviours in this class to see what specific aspects of code were the focus of instruction and how teachers and students dealt with it. This is discussed in the next section.

### **Classroom interaction outcomes: Micro-level process analysis**

A micro-level analysis of the code-related behaviour in four of the intensive classes (ie classes 5, 6, 7, and 8 in Table 1) was undertaken at this time. This analysis focused on both teacher and student code-related behaviour for a portion of the observation time.<sup>6</sup> With regard to the speech of the teachers, we examined how often they explicitly presented code-related information and how often they reacted to code (ie through correction) in learners' speech. With regard to student code-related speech, we examined how often learners reacted to code through questions and repetitions.<sup>7</sup>

The results of this analysis confirmed that the teacher in class 7 focused on code more than the other three teachers for whom the micro-level analyses were carried out. Furthermore, the teacher's code-related behaviour in class 7 was considerably more reactive than presentational in nature. That is, she reacted to learners' errors more than she presented grammar rules or explicit information about the language system. When this teacher reacted to code, her reactions were most often related to grammar rather than vocabulary or phonology. This differed from the code-related behaviours of the other teachers who tended to react to vocabulary more than any other language feature. When we examined the precise nature of the grammatical reactions of the teacher in class 7, we discovered that she did not respond to a particular feature of grammar more than others. That is, in the observation sessions which were coded,

---

6 For classes 6 and 7 this represented approximately four hours of classroom data from the four observation sessions. For classes 5 and 8, this represented eight hours of classroom data from the last observation session. All four classes took place during the same intensive program session.

7 The micro-level analysis also examined the extent to which learners were given opportunities to engage in extended discourse. Since their performance on the oral communication task had indicated that they all had developed fairly high fluency levels we wondered whether this was the result of considerable opportunities for them to engage in sustained speech in the classroom. This was not borne out in the results of the micro-level analysis of classroom learner language. (See Spada and Beaumont, 1988 and D'Amour, 1988 for further details of these results.)

it was not the case that this teacher reacted to the 'have/be' introducer form distinction any more than she reacted to the use of the 3rd person singular, plural *s* or *ed* past tense inflection. Therefore these data did not confirm that the type of code-related instructional behaviour in this class contributed to the learners' use of the correct 'be' introducer form.

It must be recalled that only a small portion of the coded data was used in this micro-level analysis. Furthermore, the total classroom observation data base in this study is quite small when compared with the total amount of instruction provided in the intensive programs (ie four days of observation spread over approximately 100 days of teaching). This raises the important issue of how representative randomly selected observation data are and therefore, how necessary it is to supplement these data through other means. This was particularly important in the process of discovering potential relationships between classroom instruction and learning outcomes in this research.

In addition to collecting classroom observation data, we had developed a teacher-structured interview/questionnaire to obtain more detailed information about the type of instruction provided. We conducted these interviews with the teachers at the end of the school year. During the course of the interview with the teacher in class 7, it was revealed that she had a particular concern about the learners' incorrect use of the verb 'have' in introducer forms. Indeed she claimed to have reacted to this error so often that she and the students joked about it. This seems to have made a difference in the learners' use of this form. Results such as these suggest that an explicit focus on code within a communicative context may lead to higher levels of accuracy in learners' use of grammatical forms. However, this may not work for all grammatical forms. The learners in class 7 were more accurate in their use of 'be' introducer forms and were also more accurate and proficient in their use of other morphological and syntactic forms (eg *-ing* progressive and possessive pronouns), they were less accurate in their use of other forms (eg plural *s* morpheme and adjective/noun placement). (See Spada and Lightbown, 1989 for a discussion of further analyses of the intensive program learners' oral production abilities.)<sup>8</sup>

Results such as these suggest that some forms may be more amenable to instructional intervention than others and only at particular times in the developmental progression (see Pienemann and Johnston, 1985, and Lightbown, in press, for further discussions of this issue). This brings us to a discussion of the subsequent phases in our research which represent the second and third elements in the Rosenshine and Furst research paradigm – correlation and experimentation.

---

8 It is interesting to note that the learners in class 3 – the one in which the greatest amount of time was spent on grammar-related instruction and practice (see Table 1), do not appear to be any more accurate in their use of these grammatical forms. However, a detailed analysis of both the learner language data and the observation data for this class are not yet complete.

### **Process-product correlation and experimentation studies**

The focus of this paper has been on eight classes from the first year of our research. As indicated earlier, observation data has been collected from twelve additional classes in the second year of our study. Preliminary results of the macro-level analysis of these classes indicate considerable differences among the classes according to some of the features on Part A of the COLT scheme. For example, as with the results of the eight classes presented here, some teachers dealt with grammar and accuracy much more than others. Also, some teachers were more teacher-centred than others, thus providing learners with considerably more opportunities for listening to the teacher and fewer opportunities to speak and listen to each other in group work. In addition to the differences which exist between individual classes and teachers, there are also within-class differences from one observation to the next. That is, the same teacher is more (or less) teacher-centred, or more (or less) meaning-based on one day than on another.

The results of the learner product analyses for the second year of the study have also indicated considerable variation. That is, some intensive program classes are performing at significantly higher levels than others on the tests of listening and reading comprehension. The relationship between these differences in learning outcomes and instructional behaviours will be investigated in exploratory correlational analyses. These will be followed by multi-variate analyses, specifically analyses of covariance (ANCOVA) to examine the extent to which differences in instruction contribute to differences in learning outcomes. As well, other variables such as contact with English outside the classroom and attitudes toward English will be examined in the ANCOVA to determine their potential contribution to performance differences and the extent to which they might interact with differences in instruction.

In addition to these further analyses of our current data base, we plan to investigate the hypothesis that explicit code-focused instruction within a communicative context will lead to higher accuracy levels in a series of quasi-experimental studies with learners in the intensive program classes in the third year of our research. Our intention is to provide learners with explicit instruction and correction in the use of specific forms chosen on the basis of linguistic theory, other SLA research, and our further analysis of learner speech, and to investigate whether this instructional treatment leads to differences in learners' accuracy levels. The design of this study includes the development of instructional units to teach the specific forms, observation of classes while the instruction is carried out and the development of a variety of tests (both receptive and productive) to measure pre/post test performance.

### **Conclusions**

This paper has provided a case study description of the research process involved in carrying out long-term L2 classroom observation research. In

describing this process, I have referred to the three elements of a research paradigm for investigating teacher effectiveness proposed by Rosenshine and Furst. These three elements – description, correlation and experimentation – all form part of our overall research design. To date, we have completed the first major stage – the description of classroom interaction and learner language outcomes. We are now at a point in our study where we have a larger number of classes and sufficient variation among them (both in terms of learner performance and instructional behaviour) to enter the second stage of the research – that of investigating the extent to which the instructional differences contribute to the performance differences. These analyses will help to refine the design for the third and final experimental stage of our research – that of hypothesising differences in learning outcomes on the basis of specific instructional behaviours.

## References

- Allen, J. P. B., Fröhlich, M. and Spada, N. (1984). 'The communicative orientation of language teaching: an observation scheme' in Handscomb, J. Orem, R. and Taylor, B. (eds.): *On TESOL '83: The Question of Control*, Washington, D.C.: TESOL.
- Allen, J. P. B., Carroll, S. Burtis, J. and Gaudino, V. (1987) 'The core French observation study', in Harley, B. Allen, J. P. B. Cummins, J. and Swain, M. *The Development of Bilingual Proficiency*, Final Report, OISE, Toronto.
- Allwright, R. (1983) 'Classroom-centred research on language teaching and learning: a brief historical overview', *TESOL Quarterly* 17:191-204.
- Allwright, R. (1988) *Observation in the Language Classroom*, London: Longman.
- Brophy, J. E. and Good, L (1986) 'Teacher behaviour and student achievement' in Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, New York: Macmillan Publishing Company.
- Chaudron, C. (1988). *Second Language Classrooms: Research on Teaching and Learning*, New York: Cambridge.
- d'Amour, M. (1988) 'Some aspects of the L2 instruction in two elementary school intensive ESL classes', M. Ed. Monograph, McGill University, Montreal.
- Dunkin, M. and Biddle, B. (1974) *The Study of Teaching*, New York: Holt Rinehart and Winston.
- Fröhlich, M., Spada, N. and Allen, J. P. B. (1985). 'Differences in the communicative orientation of L2 classrooms', *TESOL Quarterly* 19:27-57.
- Fanselow, J. F. (1977) 'Beyond Rashomon – conceptualizing and describing the teaching act', *TESOL Quarterly* 11:17-39.
- Gaies, S. (1983) 'The investigation of language classroom processes', *TESOL Quarterly* 17:205-217.
- Lightbown, P. M. (1984) 'Input and acquisition in second language classrooms', *TESL Canada Journal* 1:55-68.
- Lightbown, P.M. (1987) 'L'enseignement de l'anglais langue seconde au Québec: bilan and prospective', paper presented at the Premier Congrès des sciences de l'éducation de langue française du Canada, Quebec, March, 1987.
- Lightbown, P. M. (in press). 'What have we here: some observations of the influence of instruction on L2 learning' in Phillipson, R. Kellerman, E. Selinker, L. Sharwood-Smith, M. and Swain, M. (eds.) *Foreign Language Pedagogy Research: A Commemorative Volume for Claus Faerch*, Clevedon: Multilingual Matters.
- Lightbown, P. M. and Spada, N. (1987) 'Learning English in intensive programs in Quebec schools: 1986-87', report on the first year of research, Concordia University, Montreal.

- Lightbown, P. M. and Spada, N. (in prep) 'Learners' developing L2 abilities in intensive, communicative ESL programs', unpublished manuscript, Concordia University, Montreal.
- Long, M. (1980) 'Inside the "black box": methodological issues in classroom research on language learning,' *Language Learning* 30:1-42.
- McDonald, F., Stone, M. and Yates, A. (1977) *The effects of classroom interaction patterns and student characteristics on the acquisition of proficiency in English as a second language*, Princeton, N. J.: Educational Testing Service.
- Mitchell, R., Parkinson, B. and Johnstone, R. (1981) *The foreign language classroom: an observational study*, Stirling, Scotland: Department of Education, University of Stirling.
- Pienemann, M. and Johnston, M. (1986). 'An acquisition-based procedure for second language assessment (ESL)', *Australian Review of Applied Linguistics* 9:92-122.
- Rosenshine, B. and Furst, N. (1973) 'The use of direct observation to study teaching' in Travers, R. (ed.) *Second Handbook of Research on Teaching*, Chicago: Rand McNally.
- Scherer, G. and Wertheimer, M. (1964) *A Psycholinguistic Experiment in Foreign Language Teaching*, New York: McGraw Hill.
- Smith, P. (1969) *A Comparison of the Cognitive and Audiolingual Approaches to Foreign Language Instruction: the Pennsylvania Foreign Language Project*, Philadelphia: Center for Curriculum Development.
- Spada, N. (1987) 'Relationships between instructional differences and learning outcomes: a process-product study of communicative language teaching', *Applied Linguistics* 8:181-199.
- Spada, N. (in press). 'Observing classroom behaviours and learning outcomes in different second language programs', to appear in Richards, J. and Nunan, D. (eds.) *Second Language Teacher Education: Content and Process*.
- Spada, N. and Lightbown, P. M. (in press). 'Intensive ESL programs in Quebec primary schools', *TESL Canada Journal*.
- Spada, N. and Lightbown, P. M. (1989) *The role of grammatical instruction in second language acquisition*, paper presented at the Ninth Second Language Research Forum, UCLA, Los Angeles, February 23-26, 1989.
- Spada, N. and Beaumont, K. (1988). 'A look at some aspects of the classroom interaction in intensive ESL programs', paper presented at the 1988 TESL Canada Summer Forum, Montreal, July, 1988.

### Appendix 1 (Modified COLT, Lightbown and Spada, 1987)

SCHOOL: \_\_\_\_\_ VISIT NO: \_\_\_\_\_  
 TEACHER: \_\_\_\_\_ DATE: \_\_\_\_\_  
 SUBJECT: \_\_\_\_\_ OBSERVER: \_\_\_\_\_ PART A

COLT	ACTIVITIES AND EPISODES	PARTIC. ORGAN.		CONTENT		STUDENT MODALITY						MATERIALS						
		Class	Group	Language	Genre/Topics	Reading aloud	Writing	Repetition	Recitation	Paralanguage	Other	Minimal	Extended	Minimal	Extended	Audio	Visual	
1		T → C		Procedure														
2		S → C		Vocab.														
3		C → C		Grammar														
4		C → C		Phonology														
5			Pair	Function														
6			3+	Familiar/Con.														
7			Individual	Novel/Etc.														
8				Listening														
9				Speaking														
10				Reading aloud														
11				Writing														
12				Repetition														
13				Recitation														
14				Paralanguage														
15				Other														
16				Minimal														
17				Extended														
18				Minimal														
19				Audio														
20				Extended														
21				Visual														
22																		
23																		
24																		
25																		
26																		
27																		
28																		

TIME

# Investigating Learners' Language

Peter S. Green  
*University of York*

Karlheinz Hecht  
*University of Munich*

Any investigation of learner's language involves:

1. obtaining access to a sufficient number of representative learners;
2. devising appropriate ways of getting the learners to produce language;
3. capturing the language produced in a convenient form for subsequent analysis, and
4. evaluating it in a way that permits some valid general conclusions to be drawn that might be helpful to other language learners, language teachers or researchers into the nature of language learning.

That all sounds obvious, but there are problems that arise at every stage. Some of them are quite unpredictable, but others can be foreseen and can therefore be anticipated rather than coped with ad hoc. The aim of this article is to consider some of the predictable problems in the light of our experience in investigating learners' language.

Our project is a joint one of the *Lehrstuhl für die Didaktik der englischen Sprache und Literatur* of the University of Munich and the *Language Teaching Centre* of the University of York. Both institutions train teachers of foreign languages for secondary schools, and it is therefore with the language of school learners that we are concerned. Reference will be made to specific aspects of our research in so far as they illuminate areas of general concern. It is not the intention to present the results of our research: that has been done elsewhere (project publications are included in the References section).

The order presented in the first paragraph is an order of procedure, but initial planning obviously begins with the sort of conclusion that one hopes to come to. The questions we wanted to provide some valid preliminary answers to were principally:

1. How well can German school learners of English perform certain communicative tasks?
2. How does their performance develop over the years of secondary schooling?



The first question was refined into:

- 1.1 How effectively, despite linguistic shortcomings, can pupils understand and convey a message in English after five years of school learning?
- 1.2 How accurate is the language they use?
- 1.3 How appropriate is the language they use?
- 1.4 What are their areas of deficiency (in grammar, vocabulary, appropriacy, spelling, and pronunciation)?
- 1.5 What strategies do they employ for coping with deficiencies?
- 1.6 To what extent do they monitor their own performance and self-correct? How successful is their self-correction?
- 1.7 What is the relationship between their linguistic knowledge and their communicative performance?
- 1.8 How do native speakers react to their performance?

Asking the question 'how well?' presupposes a model performance against which the learner's performance is measured. For foreign language learners, the model is often not specified because it is assumed to be a native speaker. Native speakers, however, come in various ages, with differing abilities and from different backgrounds, both linguistic and otherwise. So what is a model native speaker like? Probably something akin to Chomsky's 'idealized speaker-hearer' (Chomsky 1972, p.116), ie an adult, educated, intelligent, experienced speaker of the standard dialect who is rarely guilty of slips of the tongue or brain. Clearly, most real native speakers fall short of the ideal in a number of respects. In order to know to what extent our German learners fell short of the model because they were *non-native* speakers of English rather than because they were, say, *young* speakers of English, we decided to seek answers to a third question:

3. How well can native English pupils, who match the German pupils in age, background and range of ability, perform the same communicative tasks in their own language? And as further evidence, how well can the German pupils perform in *their* own language?

This is, of course, an ambitious programme, which has grown out of much humbler origins as one question has led to another. A general caveat that might be made about research projects is that a sort of Murphy's law<sup>1</sup> applies to them, which says that the scope of any project is greater than you think. It may be as well to bear this in mind at the outset and circumscribe the area of investigation tightly by making it very specific.

---

1 Murphy's law (also known less politely as "sod's law") is really three laws, applying to any human enterprise, and stating about it that:  
 it will take longer than you think,  
 it will cost more than you think, and  
 if anything can go wrong, it will.

## **Access to pupils**

As can be seen from the above, we needed access to a large number of pupils representing the full range of ability if we were to be able to claim any general as well as experimental validity for our findings. The number of pupil performances has now grown to well over 1000, each representing a different pupil.

As teacher-trainers, we have been in the fortunate position of being in contact with a network of teachers in different kinds of school, who were willing to allow us access to their pupils. That willingness arose from their knowing us well enough to be sympathetic to the objectives of our research and to be assured that confidentiality would be respected. We in turn could count on them to administer our tests and adhere faithfully to our instructions, thereby saving us the virtually impossible task of going into each of their classes ourselves at some agreed time.

Researchers who do not have a ready-made relationship with the teachers whose pupils they wish to use as subjects should probably budget time for building up a relationship before administering any tests. One way is to offer to share some teaching with the teacher, who is then more likely to see the researcher in the role of a colleague rather than some daunting external 'expert'.

Teachers not only offer access to their pupils. If they are well disposed to the research and have taken time to understand its aims, they are much more likely to dispose the pupils favourably towards it, and without the willing cooperation of pupils, little of value can be achieved. Even with teacher support, pupil cooperation may not always be forthcoming:

'Despite earnest entreaties from all sides, this pupil steadfastly declined to participate in this part of the project.' (Comment about a pupil in one of the classes involved in our project.)

Access to pupils needs to be complemented by access to their syllabus and to the textbooks they have used in order to find out what, and possibly in what way, they have been taught. We were fortunate enough to have such access as well.

## **Eliciting the language**

### *Task criteria*

In devising tasks that would elicit from our learners representative samples of their communicative performance, we attempted to satisfy a number of criteria:

1. The tasks should sample both spoken and written performance and both narrative and transactional competence.
2. They should be tasks that the pupils had been taught to perform.
3. They should be tasks that they might have to perform in 'real life' (so situation, role, partner and purpose should be clear to them).

4. They should elicit some common message content to allow comparability between different pupils' performances.
5. They should be capable of being administered by the class teacher, and normally within the confines of a single lesson.

We did not anticipate any conflict between criteria 2 and 3 since the curriculum for the Bavarian schools where our German pupils were taught stipulates that English is to be taught for purposes of practical communication. There might, however, be tension between criteria 3 and 4 – as so often in language testing between the demands of validity and reliability – since too tight a control of meaning might jeopardise the attempt to simulate real life.

Our tasks were:

1. replying to a letter from an imaginary pen-friend;
2. reporting witnessed events;
3. making a youth-hostel booking by phone.

Controlling the message (criterion 4) in task 1 consisted in making the pen-friend's letter a 'letter of elicitation', which asked for specific information and called forth certain speech functions. In task 2, we used a series of pictures showing the course of events. In task 3, the role of the hostel warden was pre-recorded and the pupils were given a flow-chart telling them what they had to ask. Since the pupils in task 3 were supposed to be ringing on behalf of a group of German pupils on holiday in England, the flow-chart instructions ('Ask him...', 'Tell him...' etc, given in German) corresponded to the real-life situation of a young German in an English phone-booth being told by friends what to say.

### *Task instructions*

Task instructions are of two kinds – those concerned with the administration and those concerned with the execution of the task. The first address the teacher, the second the pupil.

Instructions need to be carefully thought through and very explicit. If they are not, there is a considerable risk of the data being invalidated. In our spoken tasks, pupils often made an appeal for help in their native language. Sometimes, it was really only a comment or an expression of frustration, to which the pupil probably did not expect a reply, eg 'Was heißt 'verhaften'?' ('How do you say 'verhaften' [=arrest]?'). We had assumed that the teacher would not respond to such appeals: we were, after all, interested in the strategies pupils employed for coping with linguistic deficits (in the above case, the pupil after a pause used an approximation – 'stop'). However, in one case, the teacher helped pupils with whispered prompts, and a whole set of data had to be abandoned.

If the researcher plans to administer a task personally, then the temptation to neglect explicit instructions is great. They should *nevertheless be written down, as if they were addressed to someone knowing nothing of the research.* In that way, comparability between

different administration occasions and accurate reporting of procedures can be ensured.

### *Pre-testing tasks and instructions*

However good one's powers of anticipation are, problems will arise that have not been foreseen. To discover and eliminate them, some form of pre-testing is essential.

We used three kinds of pre-test:

1. A rapid, informal pre-test with colleagues (of the kind 'Could you take this test away and have a go at it when you have a moment? I want to see if there are any problems I haven't foreseen') was easy to organise and did not 'use up' valuable real subjects. It served to identify any fundamental problems with the task or the instructions.
2. A pre-test with learners comparable to our ultimate subjects served to establish whether the task actually fulfilled criteria 2, 4 and 5 above.
3. A pre-test with a peer group of native speakers was used to see whether the task appeared to fulfil criteria 3 and 4.

The learners needed for the second and third types of pre-test are potential suppliers of data and therefore a precious commodity. Care with the formulation of instructions and judicious use of the first type of pre-test may well mean that no further problems arise and that pre-tests 2 and 3 yield usable data.

## **Preparing data for analysis**

### *Written data*

As a first step in our analysis, we wanted a number of native speakers to identify what they considered to be errors in the learners' performances. For written tasks, the most convenient way of doing that was for them to mark directly the offending portions of text. Obviously, the original manuscripts could be photocopied in a sufficient number of copies, but we judged it expedient to prepare typewritten versions, because, on the one hand, they could allow more space between lines for marking and, on the other hand, we did not want the irrelevant factors of handwriting quality and neatness of presentation to influence the judges. Also, typewritten versions would be less tiring to work with. We were careful, however, to preserve the spelling and punctuation of the original handwritten versions and also any alterations the pupils had made, which might serve as important clues to self-monitoring or the use of strategies of communication (see Example A). Each transcript was given a letter and number code to identify the task, the origin of the pupil (school type, class) and the individual pupil whilst preserving the anonymity of both school and pupil.

## EXAMPLE A

## Transcriptions of written test

## Pupil's handwritten version:

It was standing very high in one of my shelves, so I had to climb up. ~~With~~ while I did so, the young men probably took a little suitcase - radio and threw it ~~so~~ fast into his bag. When I came down with the radio-recorder <sup>which</sup> he has ordered, he suddenly said that he doesn't want this piece anymore. It ~~is~~ would be too expensive, ~~so~~ he said. While ~~going out~~ leaving ~~of~~ my shop suddenly the little suitcase - radio sounded from his bag.

## Typewritten version:

IT WAS STANDING VERY HIGH IN ONE OF MY SHELVES, SO I HAD TO CLIMB UP. ~~WITH~~ WHILE I DID SO, THE YOUNG MEN PROBABLY TOOK A LITTLE SUITCASE-RADIO AND THREW IT ~~SO~~ FAST INTO HIS BAG. WHEN I CAME DOWN WITH THE RADIO-RECORDER <sup>WHICH</sup> HE HAS ORDERED, HE SUDDENLY SAID THAT HE DOESN'T WANT THIS PIECE ANYMORE. IT ~~IS~~ WOULD BE TOO EXPENSIVE, ~~SO~~ HE SAID. WHILE ~~GOING OUT~~ LEAVING ~~OF~~ MY SHOP SUDDENLY THE LITTLE SUITCASE-RADIO SOUNDED FROM HIS BAG.

*Spoken data*

Transcriptions were even more essential for the processing of spoken data, which had been recorded on cassette tape. Although tapes can easily be copied and are essential to judges assessing factors such as pronunciation, fluency and comprehensibility, it is unnecessarily time-consuming to use them to identify errors.

However, transcribing natural spoken language is no straightforward matter. It contains all manner of false starts and hesitation phenomena, such as pauses, sound lengthening and hm's, as well as non-linguistic sounds, such as coughs and laughs. Learners' language contains additionally distortions of the standard language patterns of sound, stress and intonation. Decisions have to be taken on how much of all that should be transcribed. Consider the three transcriptions in Example B of the same spoken text, produced by a pupil performing our second task.

The first of these transcriptions is the easiest to read and is perhaps the most suitable for evaluating message content. The second simplifies the

assessment of monitoring (in Krashen's sense, 1981) by its clearer identification of words rejected by the speaker (shown in [ ]). It also shows hesitation phenomena, particularly sound lengthening, much more fully and consistently. For the identification of communication strategies, that may be a valuable indication of whether or not the learner is experiencing a problem. The distinction between accepted and rejected words and the accurate timing facilitate studies of vocabulary density and speech rate. The third transcription adds a narrow phonetic transcription to the information given in the second and, for comparison, an RP version of the same text. This version permits a detailed investigation of pronunciation (Pascoe 1987).

All three transcriptions were used in conjunction with the sound recording from which they were taken. For most of our analyses (see below), we found that the second transcription gave clear, accurate and adequate information, and was a good compromise between comprehensiveness and ease of reading. Its most valuable feature was the consistent use of a set of transcription conventions (given in Example B).

For close analysis of spoken language, an accurate, detailed transcription of the third kind is an invaluable, probably essential, resource. It should be stressed, however, that transcriptions of both the second and the third types are enormously time-consuming to produce, a fact that needs to be taken into account when planning the collection of data. We were fortunate in having the help of a trained phonetician for this arduous task.

## EXAMPLE B

### Transcription of spoken text

#### *Transcription 1:*

Half an hour ago, a man came into my shop and wanted to buy – a radio. Um, he show/– he showed me a radio behind me, and so – I had to climb – up a ladder. When I cl' / – I'm / I was <sup>[θtændɪŋ]</sup> standing at the – ladder, he took the / a radio which was standing at the table and put it in his bag. Uh, while he put it / while he put it in the bag, he – he wass- / he was clicking the radio on. Th<sup>2</sup> / Then, he wanted to go out, but I heard the m<sup>2</sup> / melody which is / which came in the radio. Th<sup>2</sup> / Then I run behind him, and I wanted to stop him. At the street, a policeman – help – me to stop him.

#### *Transcription 2:*

Half an hour ago, a  $\left\{ \begin{array}{l} \text{main} \\ \text{man} \end{array} \right.$  came into my shop and wanted to: buy a radio. [ʔəm He show..] He showed.. me a radio behind me, and so.. I had to climb up a ladder. When [I cl] [I'm] I was standing at ze: ladder, he took [ze r] a radio which was standing at the table and put it in his  $\left\{ \begin{array}{l} \text{beck} \\ \text{beg} \end{array} \right.$  [ʔə While he put it] While he put it in that  $\left\{ \begin{array}{l} \text{beck} \\ \text{back} \end{array} \right.$  [he:] [he was:] he was clicking ʔə the radio on. [Z] Zen he: wanted to go out, but I heard ze: melody [which is] which came in the radio. [Ze] Zen I run behind him, and I wanted him. At ze street a policeman help.. me to stop him.

78 secs.  
 104 accepted words  
 132 accepted syllables in RP  
 3 hesitation sounds  
 18 non-accepted words  
 21 non-accepted syllables incl. hesitation sounds

*Conventions used for transcription 2:*

*Symbols*

•	voiceless sound. Thus $\text{m̥}$ = a voiceless m, ie exhalation through the nose with closed lips.
<b>Right</b>	whole word spoken in a whisper
ɿ	dental or alveolar click
ʔ	glottal stop (only indicated in hesitation sounds)
ɸ	bilabial fricative
:	preceding sound is extra-long due to hesitation (a colon is <i>never</i> used as a punctuation mark, only to indicate extra length)
•	indicates that the <i>following</i> sound is made with an ingressive airstream. Thus •h = inhalation through the mouth, •m̥ through the nose with mouth closed.
head	realisation is ambiguous; the word as pronounced might be either 'head' or 'hat' – going by pronunciation alone, the top word would be the more likely.
hat	
..	slight pause
...	longer pause
....	unusually long pause
.....	extremely long pause of several seconds
(10 secs.)	length of particularly long pauses
[but]	non-accepted words, ie words 'cancelled' by the words immediately following them. Hesitation sounds between or following non-accepted words are included in brackets, hesitation sounds between accepted words are not put in brackets.
{	syntactical break (eg when a sentence is abandoned)
[but]	non-accepted words having no syntactical connection with what precedes or follows them
[oder]	German words (by definition non-accepted)
((Yes))	words not spoken by the pupil but by teacher etc.
immediátely	stress put incorrectly on indicated syllable
ve(ry)	sounds in brackets not pronounced at all, but inferred
transistor-radio	word pronounced as a compound, with only one main stress, on first element
"Help!"	exclamation mark <i>only</i> used if appropriate 'shouting' intonation is used (very rare!)
o <sub>1</sub> kay	teacher and pupil speak simultaneously
yes	

Counting Conventions

don't words including an apostrophe = 1 word  
 help-shouts words including a hyphen = 2 words  
 hesitation sounds by definition = 1 syllable, so eg ?əm = 1 syllable, 1 hesitation sound, mhm = 2 syllables, 2 hesitation sounds  
 (h·h) various kinds of laugh, usually polysyllabic, always disregarded for counting purposes  
 in RP syllable-count not made direct from pupil's text (too much work, too indeterminate), but by reading accepted text in RP and counting the syllables.

Transcription 3:

4 [ou] für /əv/  
 4 [o] " /əv/  
 1 [o] " /ə/  
 9<sup>h</sup> asp.

		RP IL + +	RP IL + -	RP IL - +
A	1 ə art.	13	-	-
	2 ə word	5	2	1
	3 ə wkEm	-	17	-
	4 h word	6	-	-
	5 h wkEm	4	-	5
B	6 hax	-	1	/
	7 [ɪ]	7	-	-
	8 [ɪ]	/	/	-
	9 [ɪ]	7	-	2
	10 [ɪ]	-	2	-
	11 v	-	-	-
	12 w	5	5	-
	13 [ə]	/	/	5
C	14 ə	-	9	-
	15 e	5	-	18
	16 ɜ:	1	-	-
	17 u:	-	-	6
	18 əv	-	9	-
	19 dʒ	-	-	-
	1-19:	53	45	37

		RP IL + +	RP IL + -	RP IL - +
D	20 θ	-	-	-
	21 θ	4	5	-
E	22 ʒ	-	-	-
	23 ɒ	6	1	-
	24 eɪ	7	1	-
	25 eə	-	-	-
F	26 p	8	-	-
	27 b	7	-	-
	28 t	25	2	9
	29 d	12	6	3
	30 k	6	-	2
	31 g	2	2	-
G	32 z	-	4	4
	33 ɪ:	16	-	2
	34 ɑ:	1	-	-
	35 ɔ:	-	-	5
	36 ɪ	18	3	-
	37 ʌ	2	-	1
	38 v	3	-	-
	39 ɪə	-	-	-
	40 və	-	-	-
	20-40:	117	24	26

Schüler	RS 11			
	RP	IL		
H	Seg.nahl	335	365	
	Kontrakt	-	-	
	[2]wnt	-	19	
	[2]hm.	/	3	
	RP IL + +	RP IL + -	RP IL - +	
J	41 aɪ	11	-	-
	42 ɔɪ	-	-	-
	43 əv	2	-	-
	44 ə	7	-	5
	45 ʃ	2	-	-
	46 tʃ	2	-	-
	47 f	1	-	-
	48 ʊ	3	-	-
	49 j	-	-	-
	50 n	23	-	-
	51 m	13	-	-
K	52 phoneme total	234	69	68
L	53 \ HF	-	8	-
	54 \ F-R	-	2	-
	55 \ LR	2	-	20
	41-51:	64	-	5



Half an hour ago, a <sup>man</sup> came into my shop and wanted to: buy  
 RP: 'ha:f ən 'a:ʊə.əgəʊ ə <sup>man</sup> kə'm ɪntə meɪ ʃɒp | ən 'wɒntɪd tə 'baɪ  
 IL: 'ha:f ən 'a:ʊə əgəʊ | ə 'mæn kə'm ɪntə meɪ ʃɒp | ɪnd 'wɒntət tu 'baɪ

a: radio. [ɪm He shov..] He shoved.. me a radio behind me, and  
 RP: ə 'reɪdiəʊ || hi 'ʃɒvd mi ə 'reɪdiəʊ bɪ'hænd mi | ən  
 IL: zə 'reɪdiəʊ || [ɪm hi 'ʃɒv..] hi 'ʃɒvd.. mi ə 'reɪdiəʊ bɪ'hænd mi | ɪnt

so.. I had to climb.. up a ladder. When [I cl] [I'm] I was  
 RP: 'səʊ əɪ 'hæd tə 'klaɪm ʌp ə 'leɪdə || 'wen əɪ wəz  
 IL: 'sɒv. ɪz 'hæt tu 'klaɪm.. ɪp ə 'leɪdə || wen [aɪ klɪ] [ɪzɪm] ɪz wɜ:s

standing.. at so: ladder, he took [se r] a radio which was standing  
 RP: 'stændɪŋ ət ðə 'leɪdə | hi 'tʊk ə 'reɪdiəʊ wɪtʃ wəz 'stændɪŋ  
 IL: 'stændɪŋ. ɪz zə 'leɪdə | hi 'tʊk [zə ɪ] zə 'reɪdiəʊ | wɪtʃ wəz 'stændɪŋ

at the table and put it in his <sup>back</sup>. [ɪə While he put it] While  
 RP: ət ðə 'teɪbl̩ ən pʊt ɪt ɪn ɪz 'bæk || [ɪə wəɪl hi pʊt ɪt] wəɪl  
 IL: ət ðə 'deɪb | ɪnt 'pʊt ɪt 'ɪn hɪs 'bæk || [ɪə wəɪl hi pʊt ɪt] wəɪl

he put it in that <sup>back</sup> [he:] [he was] he was clicking [ɪə] the  
 RP: ɪ 'pʊt ɪt ɪn ðæt 'bæk | [he:] [he wəz] he wəz 'klɪkɪŋ [ɪə] ðə  
 IL: hi 'pʊt ɪt ɪn ðæt 'bæk | [hi:] [hi wɜ:s] hi wɜ:s 'klɪkɪŋ [ɪə] ðə

radio on. [ɪ] Sen he: wanted to go out, but.. I heard.. so: [mɪ]  
 RP: 'reɪdiəʊ ɒn || 'ðen hi 'wɒntɪd tə 'ɡəʊ 'aʊt | bʌt 'aɪ 'hɜ:d ðə  
 IL: 'reɪdiəʊ ɒn || [ɪ] 'sɛn hi: 'wɒntɪd tu 'ɡəʊ 'aʊt | bʌt 'aɪ 'hɜ:t. zə.. [mɪ]

melody [which is] which came in the radio. [ɪə] Zen I run behind  
 RP: 'melədi wɪtʃ 'keɪm ɪn ðə 'reɪdiəʊ || 'ðen əɪ 'rʌn bɪ'hænd  
 IL: 'melədi [wɪtʃ ɪz] wɪtʃ 'keɪm ɪn ðə 'reɪdiəʊ || [ɪə] 'zɛn ɪz 'rʌn bɪ'hænd

him, and.. I wanted to stop him. At so street a policeman.. help.  
 RP: ɪm | ən əɪ 'wɒntɪd tə 'stɒp ɪm || ət ðə 'stri:t | ə pə'lɪsmən 'help  
 IL: hɪm | ɪnt 'aɪ 'wɒntɪd tu 'stɒp hɪm || ɪz zə 'stri:t | zə pə'lɪsmən.. help

me to stop him.  
 RP: mi tə 'stɒp ɪm ||  
 IL: mi tu 'stɒp hɪm ||

## **Evaluation of data**

Our aim was to investigate communicative effectiveness – how well German pupils could transmit a message in English with limited linguistic means. That meant looking at the content of the message, the linguistic form in which it was encoded, and the interplay of the two.

### *Message content*

What are the essentials of a message? Have they been successfully conveyed? Neither question can be answered simply, and both involve subjective judgment.

### *Essentials of message*

How easily the essentials of a message can be established depends on the degree of control over content in the task set. Our first task – writing a letter – was very open-ended: the control of content was limited to the questions asked in the letter of elicitation. Answers to them could, however, be considered to be the essential core of the message. The 46 English pupils who wrote replies to the letter in fact gave the sort of information in their answers that we had hoped to elicit.

The series of pictures in our second task – reporting a theft to the police – was an attempt to illustrate the main links in a chain of events that constituted the attempted theft and the arrest of the thief. However, pictures are not reality and are open to different interpretations, though, in practice, these particular ones did not prove to be too ambiguous. A further drawback to using pictures to control content is that the speaker tends, in interpreting them, to assume that they are also in front of any listener, ie that they constitute shared knowledge and that certain things, therefore, do not need to be expressed.

Despite such problems, we felt that we could establish nine ‘bits of information’<sup>2</sup> essential to an adequate report of the incident. Unfortunately, they did not all appear essential to the control group of 34 English pupils who also performed the task, since only two of them conveyed all nine. On average, seven were conveyed, though not always the same seven. One English pupil, giving free rein to his fantasy, conveyed only one!

In our third task – the telephone conversation – the essentials of the message were clear-cut, since they were laid down by the flow chart. Not surprisingly, then, the twenty ‘speech acts’ that were involved were all performed by the English control group.

### *Success of message*

Even when it is possible to break down the content of a message into discrete ‘bits of information’ or ‘speech acts’, it turns out to be a matter of subjective judgment whether or not they have been successfully

---

2 The term is not employed here in the technical sense, used in information technology, of ‘binary digit’.

conveyed. We therefore used three independent native speakers as assessors in order to achieve a majority verdict.

It seems worth taking account not simply of success or failure but also of whether a particular fragment of the message is attempted or not. We used the chart shown in Example C for recording success of message.

This sort of record not only gives individual pupil scores (the rows) but also an 'info bit' score (the columns), ie it reveals which bits of the message are most often neglected and which ones pupils have most difficulty in *expressing successfully*. A pupil who makes no attempt to convey part of the message may have failed to do so because of lack of linguistic resources or because that part of the message did not seem essential. It would be misleading to equate the absence of an attempt with a failed attempt, as the performance of the native speakers shows, who often did not attempt to convey certain parts of the message. The use of a native peer group can play a valuable role in disentangling linguistic and cognitive problems. Overall, the English pupils conveyed only 74% of the information we regarded as essential to the message in task 2. Whatever the reason for that, it was clearly not a lack of linguistic resources, as we might otherwise have assumed it to be with the German pupils, who overall attempted 72% of our message essentials.

### *Language form*

#### *Errors*

What kind of errors do German pupils commonly make in English; why do they make them; what consequences do their errors have for their communicative effectiveness? We adopted a classical approach to the analysis of linguistic error: identify the errors, describe them and explain them (cf Corder 1974).

#### *1. Identification*

Error has nearly always played a major role in assessing foreign language proficiency, presumably because deviance from native norms is the most striking characteristic of 'foreigner talk' and because error is thought to be amenable to objective measurement. But the objectivity is illusory: for the 2095 'errors' identified by five native judges in sixty pupil letters in task 1, there was agreement on only 1045, ie roughly half. By 'agreement' is meant only that at least three of the five markers identified the error: unanimous agreement was rare (Green and Hecht 1985). In fact, only a minority of errors – spelling errors and some morphological errors (like 'caught') – can be seen objectively as errors without reference to a context. The amount and kind of context to be considered for the rest often involve subjective judgment. It seems necessary, therefore, to use at least three independent judges and to treat as errors only what offends at least two of them (our practice for tasks 2 and 3).

We made use of the 'error survey chart' shown in Example D for this

process. The first column shows sufficient of the immediate context of the 'error' for it to be located in the text. The next three columns show which of the three markers identified it, whether they regarded it as slight, medium or severe, and what category they assigned it to (see next section). The summary column gives each majority error a serial number and records its majority category and gravity.

2. Description

Errors can be classified in several ways, eg as 'local', ie affecting single elements in a sentence, or 'global', ie affecting overall sentence organisation (Burt and Kiparsky 1972) or, more commonly, by assigning

EXAMPLE C

Information bit record sheet

Pupil no.	Information bit no.									Score		
	1	2	3	4	5	6	7	8	9	✓	(✓)	-
xyz 1	✓	✓	(✓)	(✓)	(✓)	-	✓	-	✓	4	3	2
xyz 2	✓	-	(✓)	(✓)	(✓)	✓	✓	✓	✓	5	3	1
xyz 3	✓	✓	(✓)	✓	(✓)	-	✓	-	✓	5	2	2
xyz 19	✓	-	(✓)	(✓)	(✓)	-	✓	-	✓	3	3	3
Totals												
✓	19	12	0	2	4	4	14	4	12	71		
(✓)	0	1	13	5	14	1	3	1	1		39	
-	0	6	6	12	1	14	2	14	6			16
Attempt rate %	100	68	68	37	95	26	90	26	68	64		
Success rate %	100	92	0	29	22	80	82	80	92	65		

✓ = succeeded  
 (✓) = attempted  
 - = omitted

EXAMPLE D  
Error Survey chart

Pupil No: xyz 1	MARKER 1				MARKER 2				MARKER 3				Summary		
	1	2	3	cat. no.	1	2	3	cat. no.	1	2	3	cat. no.	majority error no.	no. accepted words	majority severity
ERROR IN CONTEXT	~	-	=		~	-	=		~	-	=				
What cost...		x		1	x			1		x		1	1	1	2
I call...		x		3		x		3		x		3	2	3	2
lunch...			x	5			x	5					3	5	3
We are...					x			6					-	-	-
Please tell me... etc...		x		8						x		8	4	8	2
air (= R)...			x	5		x		10			x	10	24	10	3

GPE  
(general pron. errors)

1. No. of majority errors 24
  2. No. of accepted words 97
  3. Error rate (correct to one decimal place) 24.7
- $[\frac{1}{2} \times 100]$

them to grammatical categories. Our classification was a combination of these two types of category in that it included both grammatical categories and a category called 'discourse organisation' (Example E).

Our purpose in carrying out an error analysis was largely a pedagogic one – feedback to teachers on the incidence of error, the areas of language most affected and error gravity (impact on message and degree of irritation). It seemed therefore appropriate to combine classification and evaluation in order to detect patterns of interplay, eg which category of error is most or least likely to cause irritation or lead to a breakdown of meaning. It is interesting that discourse errors, which are a major source of misunderstanding and irritation, were errors to which non-native teachers of the language were least sensitive. They made their judgments of error gravity much more at the level of sentence and form than at the level of discourse and meaning. Native English pupils, on the other hand, were very sensitive to discourse errors made by German pupils (Green 1985).

The definition of error categories is somewhat arbitrary and there is bound to be a certain amount of fuzziness at the category boundaries. Some subjectivity in the assignment of errors to categories is therefore inevitable. Gravity of error is entirely a matter for subject judgment, though there is often good agreement between judges. Again, therefore, at least three assessors are necessary, and categories should be carefully defined.

## EXAMPLE E

### Error category/gravity chart

*category*

Error category/gravity chart (Summary)		total	no gravity agreement	gravity		
				1 ~	2 _	3 _
grammar	1	Syntax				
	2	Morphology				
	3	Tense				
	4	Preposition				
vocab.	5	Concept				
	6	Collocation				
	7	Non-existent word				
style	8	Choice/Appropriacy				
	9	Discourse organisation				
10	Phonetics					
no category agreement						
total number of errors						

### 3. *Explanation*

Of all the stages in a traditional error analysis, the attempt to explain errors, to find their source or cause, is perhaps the most speculative. On the other hand, since errors are the most obvious characteristic of learners' language and are generally considered to be the product of an interim, deviant language system, they can offer important insights into the process of language acquisition, particularly if they are studied over a period of time rather than at a single point in time.

That different research projects have come up with such different results when errors have been assigned to causes (cf. the widely varying percentages of errors attributed to L1 interference, Ellis 1986, p. 29) is perhaps due less to arbitrariness of assignment than to differing interpretations of terms such as 'L1 interference' and 'overgeneralisation', which lack any universally accepted definition. It may also be due to a failure to distinguish between error behaviour and error source. (See Example F, the chart on which we recorded error causes.) Our experience suggests that judges achieve an acceptable measure of agreement in explaining error once the terms are defined.

### *Communication strategies*

Much has been written in recent years about 'communication strategies', how learners cope when their linguistic resources are inadequate or do not readily enable them to communicate the meanings they wish to express (eg Faerch and Kasper 1983). So far, much of the discussion has been theoretical, without the support of any extensive research evidence. One of the aims of our project is to look at the use of strategies in all our productions and to relate them to different levels of proficiency and different areas of language.

There is a considerable problem of identification with communication strategies. For example, one pupil said:

and he pointed [on some] on a radio that əm stands behind me [on a] [was heißt "Regal"<sup>3</sup>] on [a:] a board.

That pupil was undoubtedly experiencing a problem with expressing the concept 'shelf' and using a strategy of approximation. Another said:

he pointed to the one [behind it..] behind me.

Was that pupil using an avoidance strategy, because the word 'shelf' was not available, or was that simply the way the pupil chose to express the meaning conveyed by the pictures? Probably the latter, since the pupil in question was a native speaker of English and could certainly have produced the word 'shelf' if it had seemed appropriate. But what if the pupil had been German? In all cases where we believed we could identify a communication strategy, we looked for evidence of a problem before

---

3 German for 'How do you say "shelf"?'

**EXAMPLE F**

**Error cause chart**

*Error cause chart/School type: Pupil no: Task:*

	<i>Error source</i>	<i>Error behaviour</i>	
L1	<b>LANGUAGE</b> 1. Structural difference ----- 2. Divergent learning structure in L2	PUPIL language transfer	-----
L2	3. Offers major clear pattern or rule, e.g. -s plural, "mark past tense"	overgeneralisation of rule or pattern overextension	
	4. Lack of contrast between L2 items	Ranschburg confusion	
L1/L2	5. between L2-L1 items		
P U P I L	6a. Lack of concept refinement	overextension of lexical item or grammatical category	
	6b. Awareness of breakdown in linguistic resources	e.g. coinage or simplification strategies ----- language switching	-----
	7. Slip of the brain	unmonitored behaviour	
	8. TASK	unsystematic use of tense	
	9. TEACHER/TEXTBOOK/SYLLABUS	'transfer of training'	
	10. Ambiguous	erroneous behaviour	

accepting it as such. Was the imagined strategy preceded by a pause, a lengthening of a phoneme, a false start or an appeal for help? Once again, the careful transcription exemplified in Example B greatly facilitated the search for such evidence. Thus, the German pupil who said:

ʔə m it was .. ʔə .. behind me and [it ə stands..] [it sta] [it stood ʔə.. ɪʔ] it stood above

was assumed to have employed an avoidance strategy. Useful evidence was also provided by what those pupils said who performed the task in their own language – both German pupils in German and English pupils in English.



## EXAMPLE G

Pupil No.

**Production Strategies***Resource expansion strategies:*

a)	foreignising (L <sub>1</sub> /L <sub>3</sub> )	
b)	coinage/loan-translation (L <sub>1</sub> based strat.)	
	coinage (L <sub>2</sub> based strat.)	
c)	language switching (L <sub>1</sub> )	
d)	appeal for help/tolerance	
e)	lexical approximation (L <sub>2</sub> based strat.)	
	1. too specific	
	2. too general	
	3. co-hyponym	
	4. overextension	
f)	circumlocution (L <sub>2</sub> based strat.)	
g)	bypassing/grammatical approx. (L <sub>2</sub> based strat.)	

*Avoidance strategies:*

a)	abandonment of m.	
b)	modality reduction	

*fillers:*

	1. first word(s) of planned utterance repeated because of	
	2. last word(s) of utterance repeated because of	
	3. lengthening of sound because of	
	4. er/um because of	

*self-correction:*

	1. phonetic	
	2. lexical	
	3. morphological	
	4. syntactical	
	5. semantic	

*prefabricated pattern:*

### *Monitoring*

Monitoring, or self-correction, is another interesting aspect of learners' language that is greatly facilitated by an accurate transcription. The square brackets in the examples above readily pick out points at which learners may be correcting their production or changing their plans. (They may, of course, also be coping with a production problem.)

There are many interesting aspects of monitoring, eg to what extent is it task-related, how often does it result in an improvement of the production, are different areas of language – pronunciation, vocabulary, grammar – differently affected, what variability is there between learners? Furthermore, what is the effect of monitoring on listeners: is the greater accuracy, which monitoring does seem to achieve, outweighed at some point by the irritation caused by the interruption of the speech flow, and if so, where is that point?

The charts on which we recorded communication strategies and self-correction are shown in Example G. The nature of our tasks meant that we were concerned only with production and not with comprehension.

### *Mechanical processing*

Not all evaluation of learners' language requires the painstaking, time-consuming scrutiny of the kind described above. Some of it can usefully, and probably more accurately, be undertaken by mechanical means. A computer program like the Oxford Concordance Program (Hockey and Marriott 1980) enables a very accurate assessment to be made of aspects of learner vocabulary such as word frequency, overall vocabulary, context, collocation, type-token relationships – all for the simple effort of typing the texts to be studied into the computer. Yet again, as spoken texts have first to be represented in writing, the type of transcription that clearly distinguishes between accepted and rejected words is invaluable.

### *Interplay of message and medium in communicative effectiveness*

Communicative effectiveness in a foreign language is more than the successful transmission of a given message content: it also involves not diverting the listener/reader's attention from the message to the form of the language by excessive deviancy of pronunciation or grammar, by inappropriacy of vocabulary, or by a speech flow that is constantly interrupted or fragmented by pauses for thought, monitoring, false starts or poor cohesion. What then is the interplay of message and medium in communicative effectiveness? Both a qualitative and a quantitative approach can be adopted in seeking answers to this question.

### *Qualitative*

English pupils were asked to rank-order three pen-friend letters written by German pupils from the point of view of their willingness to take up a correspondence with their authors. They were also asked to explain their preferences. They showed a very good measure of agreement both in their choices and in their reasons (Green and Hecht 1987). The

open-endedness of this kind of procedure has the advantage of allowing any potentially useful but unexpected insights to emerge. The data, however, are very difficult to summarise fairly.

### *Quantitative*

Attempts to break down communicative proficiency into separate factors such as pronunciation, accuracy, appropriacy, message content (eg Carroll 1980) raise many contentious questions such as whether such proficiency is 'unitary' or 'divisible' (Vollmer 1981).

We asked a large number of 'reactors' – some seventy native English speakers and a roughly equal number of German teachers or trainee-teachers of English – to listen to seven recorded performances of our second task (the picture report) and assess them on a five-point Likert-type scale for ease of understanding, appropriacy of language, pronunciation, grammatical accuracy, fluency, pleasantness, message content and overall effectiveness. The reactors were not given the pictures which the pupils had had.

A number of interesting questions can be asked of the statistical data yielded by such a procedure, eg are judges capable of assessing so many different performance factors independently or do they, as has been suggested (eg Walker 1983), tend to assess the same thing over and over again? Is their judgment of overall effectiveness based on an amalgam of factors or is it strongly influenced by one or two of them? Does, for instance, a judgment that a speaker was very good overall really mean no more than that the speaker had, say, a very good pronunciation?

A further interesting area of investigation is the matching of such subjective judgment with more objective information. How, for example, does reactors' assessment of message content relate to the number of information bits recorded for the speakers concerned or their assessment of grammatical accuracy to the actual number of errors recorded or their perception of fluency to the measured speech rate?

The difficulty for many linguists is coping with the statistics involved in processing such data – not so much in handling the data at the computer level (that is relatively easy with the aid of modern statistical packages) but in knowing what statistical procedures to invoke and how to interpret the results when they are applied. Statistical primers (such as Greene and D'Oliveira 1982, Kenny 1982) are helpful for understanding basic procedures. When it comes to deciding on appropriate procedures for seeking answers to questions such as those raised above, one really needs expert advice, and that, after the event, is not always easy to come by. It is much better, if possible, to have a statistical consultant associated with any larger scale investigation from the outset, so that advice can be given not only on the actual statistical processing but also on the most convenient and appropriate way of collecting and codifying data.

### *Supplementary investigations*

When the data from any project are examined, questions are usually

thrown up which were not originally foreseen. For example, our learners made most grammatical errors in the area of tense usage (and to a lesser extent tense formation) despite the fact that tense occupies a central part in the teaching they had had. The question then arises whether the pupils had failed to learn what they had been taught or whether they had been given a competence which they were unable to convert into performance in communicative tasks, where the focus was on meaning. Such questions have led us to conduct many supplementary investigations with a limited specific aim, in the latter case, for instance, to administer both recognition (multiple-choice) and productive (gap-filling) tests of tense usage in the areas where there was most breakdown in actual use.

This kind of supplementary investigation is one of the areas where the law of project proliferation mentioned at the outset most easily applies. A technical difficulty that this creates is the large number of data tables and the ease of getting them confused if, as is often the case, they have to be expanded or revised. It is a wise precaution, we have decided, to date each table as it is produced.

## Conclusion

The outcome of any investigation of learners' language must first and foremost accurately represent the language of the actual learners who were investigated. The investigation must have *internal validity*. However, if the results are to be of value outside their original context, they need to have *external validity* as well. That is, they must be broad-based and representative enough to be generalisable to other, similar learners.

Where that is achieved, classroom-based research may help us to a better understanding of how pupils learn and specifically

- what learning and communication strategies they use;
- how those strategies develop over the years;
- how long it takes them to apply in free communication what they have learnt in controlled practice in the classroom (the 'incubation period').

Presenting results may, however, pose something of a dilemma. Many of those who would benefit from the research could easily be put off by the rather technical language that is needed to address questions of internal and external validity (Campbell and Stanley 1963). It may therefore be advisable to present results in two reports, the first permitting proper scrutiny of research procedures and the second giving direct access to practical outcomes without the technical details (but with reference to where they may be found). Unfortunately, if researchers do not themselves suggest legitimate applications for their research results, there is a risk that others may suggest less legitimate ones.

One legitimate application of the results of classroom-based research should be to offer feedback on how efficient the teaching of skills and language systems has been, to syllabus designers, course-book writers

and teachers. Ideally, they should not merely be the passive recipients of research but active participants in it, drawing on and contributing to a centrally-based corpus and thus continually increasing its representativeness. A very suitable base for such a corpus is a teacher-training establishment, where graduate students can become involved in classroom-based research from the outset, advancing it in the short-term whilst laying foundations for future research.

## References

- Burt, M. and Kiparsky, C. (1972) *The Gooficon: A Repair Manual for English*, Rowley, Ma.: Newbury House.
- Campbell, D. T. and Stanley, J. C. (1963) 'Experimental and Quasi-Experimental Designs for Research on Teaching' in Gage, N. L. (ed.) *Handbook of Research on Teaching*, Chicago: Rand McNally.
- Carroll, B. J. (1980) *Testing Communicative Performance*, Oxford: Pergamon.
- Chomsky, N. (1972) *Language and Mind*, New York: Harcourt Brace Jovanovich.
- Corder, S. P. (1974) 'Error analysis' in Allen, J. and Corder, S. P. (eds) *The Edinburgh Course in Applied Linguistics*, Vol. 3, Oxford: Oxford University Press.
- Ellis, R. (1986) *Understanding Second Language Acquisition*, Oxford: Oxford University Press.
- Faerch, C. and Kasper, G. (eds) (1983) *Strategies in Interlanguage Communication*, London: Longman.
- Green, P. S. (1984) 'Marking written work,' *Inside English 4* (Cassette magazine for English teachers), Hamburg: Petersen-Macmillan.
- Green, P. S. (1985) 'Teachers and pupils as markers' in Green, P. S. (ed.) *York Papers in Language Teaching*, University of York, England: Language Teaching Centre.
- Green, P. S. and Hecht, Kh. (1984) 'Reliability in the assessment of written communicative skills', *Finlance* 3: 1-23.
- Green, P. S. and Hecht, Kh. (1985) 'Native and non-native evaluation of learners' errors in written discourse', *System*, Vol. 13, No. 2: 77-97.
- Green, P. S. and Hecht, Kh. (1987) 'The influence of accuracy on communicative effectiveness', *British Journal of Language Teaching*, Vol. 25, No 2: 79-84.
- Green, P. S. and Hecht, Kh. (1988) 'The sympathetic native speaker – a GCSE role-play for the teacher', *Modern Languages*, Vol. 69, No 1: 3-10.
- Green, J. and d'Oliveira, M. (1982) *Learning to use Statistical Tests in Psychology*, Milton Keynes: Open University Press.
- Hecht, Kh. (1984) 'Die Fehler von deutschen Schülern bei der schriftlichen Sprachproduktion und ihre Bedeutung für den Englischunterricht der Sekundarstufe I', *Inside English 4* (Cassette magazine for English teachers), Hamburg: Petersen-Macmillan.
- Hecht, Kh. (1986) 'Das Interview. Karlheinz Hecht im Gespräch mit der Praxis des neusprachlichen Unterrichts zum Thema "Fremdsprachliche Fehler und ihre Benotung"', *Praxis des neusprachlichen Unterrichts* 1: 79-86.
- Hecht, Kh. (1987) 'Talking about the Past: Bypassing Strategies', *Englisch* 3: 105-107.
- Hecht, Kh. (1987) 'Lexikalische Kommunikationsstrategien', *Englisch* 3: 107-109.
- Hecht, Kh. and Green, P. S. (1983) *Fehleranalyse und Leistungsbewertung im Englischunterricht der Sekundarstufe I*, Donauwörth: Ludwig Auer.
- Hecht, Kh. and Green, P. S. (1987) 'Analyse und Bewertung von mündlichen Schülerproduktionen', *Praxis des neusprachlichen Unterrichts* 1: 3-11.
- Hecht, Kh. and Green, P. S. (1988) 'Verständlich oder sprachlich korrekt?' *Pädagogische Welt* 2: 84-87.
- Hecht, Kh. and Green, P. S. (1988) 'Kommunikative Wirksamkeit von Schülerbriefen – ein Produkt von Sprachrichtigkeit?', *Englisch* 1: 1-8.

- Hecht, Kh. and Green, P.S. (1989) 'Zur kommunikativen Wirksamkeit von fehlerhaften Schülerauserungen' *Praxis des neusprachlichen Unterrichts* 1:3-9.
- Hecht, Kh. and Green, P.S. (1989) 'Grammatische Kompetenz und Performanz' *Die neueren Sprachen* 6:573-590.
- Hecht, Kh. and Green, P.S. (1989) 'Fehler: leicht – mittel – schwer? Zum Problem der Fehlergewichtung und des Bewertens der interpersonellen und referentiellen Funktionen der Sprache' *Englisch* 3:81-87.
- Hockey, S. and I. Marriott (1980) *Oxford Concordance Program, Users' Manual*, Oxford: Oxford University Computing Service.
- Kenny, A. (1982) *The Computation of Style*, Oxford: Pergamon.
- Kieweg, W. (1988) 'Die Signalwortgrammatik – funktioniert sie wirklich?', *Pädagogische Welt* 2: 81-83.
- Krashen, S. D. (1981) *Second Language Acquisition and Second Language Learning*, Oxford: Pergamon.
- Pascoe, G. (1987) *Die Aussprache des Englischen an Bayrischen Schulen*, Munich: Profil.
- Vollmer, H. J. (1981) 'Why are we interested in "General Language Proficiency"?' in Klein-Braley, C. and Stevenson, D. K. (eds), *Practice and Problems in Language Testing* 1, Frankfurt: Verlag Peter D. Lang.
- Walker, C. R. (1983) 'The whole is more than the sum of the parts: evaluation criteria in oral tests', *British Journal of Language Teaching*, Vol. 21, No 1: 41-44.

# Researching Teachers: Behaviour and Belief

Tony Lynch

*Institute for Applied Language Studies, University of Edinburgh*

## Teacher roles

Research into the process of teaching foreign languages has tended to focus on tangible classroom events – teaching as action, teacher as actor or agent. Moreover, attention has often been concentrated on one specific aspect of the teacher's classroom performance, in order to flesh out our picture of how teachers enact their various classroom roles. The roles that have been studied in most detail so far include the following:

- the teacher as *producer of language* (eg FL Teacher Talk)
- the teacher as *eliciter/encourager* of learner language (eg questioning strategies)
- the teacher *arbiter/corrector* of learner language (error-handling)
- the teacher as *explainer of language* (metatalk, reformulation)

The teacher's role as *language producer* has been studied in work on Teacher Talk, which initially involved teachers' modifications of input (eg Henzl 1974), then turned to investigations of discourse adjustments (eg Hatch 1978), leading on to recent research into the differential effects of the two on learners' comprehension (Pica, Young and Doughty 1987). It is worth stressing Chaudron's (1988) point that foreign language teacher talk is not a specific linguistic genre, but lies at the intersection of three areas of language use – general (ie native/native) communication strategies, classroom talk and foreigner talk. This serves as a general reminder of the permeability of categories of behaviour that the analyst may wish to set up.

The way that teachers *elicit or encourage* target language production by their students has formed the focus for studies of teacher questioning strategies, associated in particular with the work of Long and Sato (1983). They compared the questions occurring in informal native/non-native conversation and teacher/learner interaction in the L2 classroom, and summarised their findings in this way:

whereas display questions predominate in EFL instruction. . . they are virtually *unknown* in informal NS-NNS conversation. Conversely, referen-

tial (information-seeking) questions, which predominate in NS-NNS conversation outside classrooms (76% of all questions asked) made up a mere 14% of questions asked by teachers. This result suggests that, contrary to the recommendations of many writers on L2 teaching methodology, communicative use of the target language makes up only a minor part of typical classroom activities. 'Is the clock on the wall?' and 'Are you a student?' are still the staple diet, at least for beginners, (Long and Sato 1983:280).

Until recently, following their main conclusion, it has been claimed that teachers whose questions to learners demanded genuinely informative answers (that is, answers unknown to the teacher) were doing more to assist their students' communicative competence than teachers who relied on 'display questions' to which they already knew the answer. However, van Lier (1988) challenges this assumption – what we might term the authenticity argument – that what matters about teachers' elicitation is that they should lead to informative interaction. We will return to his alternative proposal shortly.

The other two aspects of teachers' classroom performance (in their roles of *explainer* and *arbiter/corrector*) both involve the ways in which they respond to potential crisis points – points in the discourse where the learners either fail to understand the target language (whether produced by the teacher or another learner) or fail to produce the target language acceptably.

The evidence available so far allows us only limited and tentative conclusions. Mitchell, Parkinson and Johnstone (1981) offer some support for the view that an increased quantity of metatalk (grammatical explanation) is associated with relative success in target language learning. However, their data provides evidence of correlation, not necessarily of cause and effect. It is quite possible that teachers feel more able to use metatalk with learners who seem to be making relatively good progress, rather than that it is the use of such explanation that leads to better learning.

Similarly, on the issue of error handling, there is little firm evidence for what must surely be the view of the majority of language teachers, namely that corrections lead to learners' progress – or, in its weaker version, that without correction by the teacher there will be no improvement. Such evidence as there is suggests that teachers' corrections have a greater effect on syntax than on pronunciation (Ramirez and Stromquist 1979).

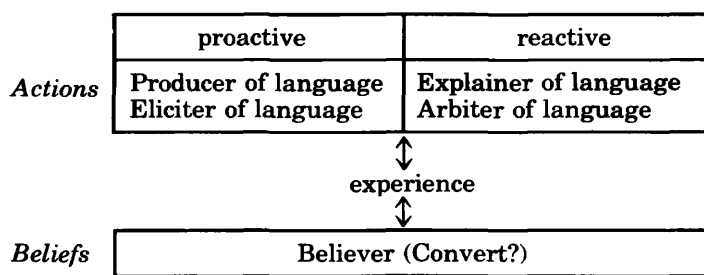
One of the things that these studies have in common, from the point of the practising teacher, is that they show the importance of the individual's underlying assumptions about how people best learn languages in the classroom. Teachers' options within the four roles investigated so far could be encapsulated by asking the following questions (chosen from many possible):

1. *Teacher as language producer*: Is it better to say things once slowly or to repeat everything you have said?



2. *Teacher as eliciter*: Is it better to ask Wh or yes/no questions?
3. *Teacher as explainer*: Is it better to use the learners' mother tongue or the target language to explain vocabulary that is causing them difficulty?
4. *Teacher as arbiter/corrector*: Is it better to give the learner a second chance (ie to self-correct) rather than to supply the correct form yourself?

Individuals' answers to these sample questions could well be on the lines of 'Well, it depends on . . .' and the variables would presumably include such factors as learners' level, age and aptitude. In coming to decisions as to which of the wide range of factors to take into account in selecting options for action, the teacher draws on experienced-based beliefs. The way in which such beliefs underpin classroom behaviour might be expressed in the following diagram:



### Behaviour and belief: a sample study

Having briefly summarised the roles that have been studied so far, I propose now to illustrate the value of classroom research data in showing the effect of individual teachers' beliefs about the teaching/learning process – in other words not only in establishing what teachers do but also in analysing (or allowing them to analyse) why they do it. Although teachers may be *actors* in that one part of their professional activity is to bring to life teaching texts written by others, they are in an important way also *agents*, taking their own decisions and responding to their individual mental scripts – sets of principles about what is advisable or necessary in the classroom.

The study I will be referring to is one that set out to investigate the common belief that teachers' decisions to adopt specific classroom procedural options depend on the learners' level. It was designed to compare what happened when different teachers used the same language teaching material with EFL classes at different levels of proficiency: post-elementary, intermediate and advanced.

The materials were in fact originally developed for use in a mother tongue context, as part of a project conducting basic research into the

listening competence of secondary school pupils in Scotland<sup>1</sup>. The researchers found evidence confirming findings from an earlier study of younger native listeners (Robinson 1981), which showed that children lacked judgment about the adequacy of messages and tended to blame themselves as listeners when they failed to understand what were in fact unsatisfactorily vague or ambiguous texts.

In the Scottish study, training materials were scripted to contain some of the inconsistencies and ambiguities that naturally occur in real life discourse. The classroom procedure was that the pupils were played an audiotape and they had to respond to the message in various ways. In each case they had to complete some sort of tasksheet; in order to do so, they were required to recognise when they were experiencing problems of comprehension and to take appropriate action. When any individual listener had a problem following what had just been said, they were to ask the teacher to stop the tape. As a group, the learners then had four possible options:

1. to have the tape replayed;
2. to play on and see if the next part of the text solved their problem;
3. to ask the teacher for additional, clarifying information;
4. to discuss the difficulty with the other listeners and see if any of them had a plausible interpretation.

This classroom procedure was intended to result in a re-alignment of classroom roles, so that learners would be encouraged to take the initiative in getting the help they required to understand the message. They were made responsible both for the timing and for the content of questions that might enable them to solve their current problem. The teacher was to provide only the information that the learners specifically asked for.

Such a change of teacher and learner roles is supported from two different perspectives on the foreign language classroom – one emphasising the linguistic benefits of that change, the other the social/interactional. In comparing different forms of discourse modification by teachers, Pica, Young and Doughty conclude:

Facilitating input comprehension in classroom settings . . . requires a teacher-student relationship and patterns of classroom interaction that are radically different from the pattern of teacher elicitation, student response, and teacher feedback . . . It also requires contributions from learners that are geared toward understanding input and not simply toward providing formally correct speech. (Pica, Young and Doughty 1987:754)

More recently, van Lier (1988) has addressed the wider issue of the classroom as a setting for social interaction. He considers the ways in

---

1 The project (JHH/190/1) was funded by the Scottish Education Department and was directed by Professor Gillian Brown at the University of Edinburgh in 1982-85. The main research is summarised in Brown, Anderson, Shadbolt and Lynch (1987). For details and analysis of the L2 classroom study, see Anderson and Lynch (1988).

which questions (and the questioner) exert social control and argues that whether teachers ask display or real questions may well turn out to be unimportant. What *is* important is the way in which control over discourse is maintained by the teacher or made available to the learners:

Discourse studies . . . of questioning strategies in practical everyday contexts shed considerable light on the controlling power of questions. It is important to point this out since the practice of questioning in L2 classrooms, pervasive though it is, has so far received only superficial treatment . . . An analysis must go beyond simple distinctions such as display and referential questions, yes/no and open-ended questions, and so on, to investigate what different tasks questions set, and the different commitments they place on the answerer. (van Lier 1988:224)

I would like to focus on one particular aspect of my study of the classroom performance of students and teachers in the EFL setting<sup>2</sup>: the way they approached problems caused by unfamiliar lexis. The reason for singling this out for attention here is quite simple. I assume that most language teachers would predict that the principal difference between learners at post-elementary, intermediate and advanced levels of proficiency would be their ability to cope with the (unmodified) vocabulary used in the cassette. Similarly, if asked to forecast differences in teachers' tactics, they would probably say that the teacher with the weakest group would have to devote time to explanation of new lexical items.

In fact, the data support neither of these assumptions. Below are extracts from transcripts, showing all the questions asked by each group on the same activity (which involved marking in a route on a city map):

**Route-drawing task: Learners' questions (to each other and to the class teacher)**

<i>Group A: Post-elementary</i>	<i>Group B Intermediate</i>	<i>Group C: Advanced</i>
What to do?	What is Fragrant Lotus pagoda?	Replay?
Are you agree to repeat?	The first left?	Which way do you go back?
Right or left?	May I listen to the tape again?	What's the national monument?
Go down?	Where is the national monument?	First left? (x2)
Ibrahim, what you understand?	Market?	Which road?
Turn left this one?	What is the national monument?	Like this?
This one?	Where we are going now?	Do you want me to replay?
Where is pagoda?	Which direction – right or left?	Silk mill?
Where is the door?	Can we listen to the last part?	Silk mill – where is it?
How will you know?	Which back?	Is the mill the factory?
Which one? (x3)	Near the beach?	Which factory?
Which one pagoda? (x5)	Is it to the right side?	Come back which way?
You didn't understand?	Silk factory?	Nations Road?
Do you want to ask teacher?	But how?	Which door we use?
But how can you turn left?	Go back?	
When did you turn right?	You mean Nations Road?	
How do you sure?	/sukunel/ – what does it mean?	

2 For further discussion, see Lynch (1988).

Is this explain? (x2)	What is 'milk'?
Why you need to listen and make it?	Silk milk?
National what?	Did you hear 'tower'?
Where down?	Did you hear?
Where is monument?	Is there anybody to hear 'tower'?
From which way?	How about we listen again?
National palace this one?	After the museum or after the tower?
Where is the national monument?	<i>What's the meaning of a silk mill?</i>
How you know is it?	Which kind of information can we ask?
What's the name of the road? (x4)	Why don't we ask a question?
Which one?	
What is Vat Road?	
You hear 'by bus'?	
That's the national monument?	
'National' – may you write?	
<i>What does 'statue' mean?</i>	
Which one market?	
This one?	
Clear?	
Is it museum or. . . ?	
Where is first left?	
Where is the national museum?	
From which door you come out?	
From this side?	
Silk mill? (x6)	
Do you agree to listen again?	
It is in the tower? (x4)	
Near factory is here?	
<i>Is factory for silk, silk mill?</i>	
Did you hear 'turn right'?	

---

Total: 47 questions

27 questions

14 questions

---

It will be clear that hardly any of the questions that the post-elementary learners asked were about vocabulary. Of a total of 47 questions, only two (in italics in the table) involved questions of lexical meaning; this was no more than the number asked by the more advanced Group B. So a basic finding of the study is a surprising one: that lexical unfamiliarity was less of a problem than might have been expected, given the range of language proficiency across the three learner groups. Against this background, let us examine the tactics deployed by one of the teachers, in two different situations, in relation to lexis-based difficulties.

### *Situation 1*

In the first case, the reaction of Group A's teacher to one particular comprehension problem suggests that her own perception of what was going on (and consequently how best to deal with the difficulty) was rather different from that of the students. In the following extract, the post-elementary group is trying to identify a location on their map which might correspond to the 'silk mill' mentioned on their tape. There are three possible candidates on their map: one item marked 'tower' and two marked 'factory'.

*Extract 1*

- Cassette: . . . The last stop on the tour is the silk mill.  
Students: Silk mill?  
Student O: It is the tower or . . . ?  
Student Y: It's better to — uh — we need more information.  
Student O: The silk mill in the tower or not?  
Teacher: Do you know the meaning of mill?  
Student K: Milk?  
Student I: Mill.  
Student M: Mill? It's the postman.  
Student Y: Mail.  
Teacher: Yes, that's one kind.  
Student I: Air or wind mill.  
Teacher: But this mill is for making silk — do you know silk? — cloth.  
Student Y: It's cloth.  
Teacher: A kind of cloth.  
Student M: Yeah — elegant.  
Student K: Can you write?  
Teacher: Silk mill (*writes on board*).  
Student M: Ah, I think he go to the factory.  
Student Y: To factory, but which factory?  
Student O: You have two factory.  
Student Z: Yes, near factory is there.  
Student Y: If we go on maybe we will know.

We seem to have three different strategies here. Student O's attention is focused on solving the task and he wants to use the fact that the possible range of destinations is, by this stage of the task, limited to three. Student Y believes that they may be able to sort it out by listening to the next part of the tape. We could say that both have assessed the task as a *listening* problem. However, their approaches to the problem contrast with the teacher's perception of the source of difficulty; she regards the basic problem as one of vocabulary, ie a *language* issue, and nudges the students in the direction of a 'teaching episode'. ('Do you know the meaning . . .?') in which she offers help, in her view, by providing or eliciting the meaning of 'mill' and 'silk' in turn.

As we can see, in this case the teacher's intervention as *explainer* of language does not in fact result in forward movement. The group have effectively completed a loop or digression as a result of the teacher's effort to assist them. We may assume that she was following some underlying principle such as 'Failure to understand can be due to lack of vocabulary' — a principle which may well be valid for most circumstances, but not for this particular configuration of input, learner and task.

*Situation 2*

The second extract is interactionally more complex. Indeed, it serves to illustrate the fact that some recordings and transcripts of classroom discourse are hard to interpret if one adheres only to the linear sequence

of utterances as they occur on tape or page. In the extract below the speaking turns have been numbered for ease of reference.

Here, the same teacher appears to adopt a rather different tactic, when a lexical difficulty arises, from that displayed in the case of 'silk mill'. This time Group A are working on a task that requires them to complete and label a geometric figure. They have in front of them a square ABCD and they have just heard the taped instruction 'the first thing you have to do is to put a dot on the line of the left-hand side of the square — and you put the dot halfway down'.

*Extract 2*

- 1 Student A: What?
- 2 Student B: Stop.
- 3 Student I: Dot?
- 4 Student J: Dot?
- 5 Student O: Point.
- 6 Student Y: Dot?
- 7 Students: Point/point, yeah.
- 8 Student A: Point?
- 9 Student O: Small point.
- 10 Student I: Dot.
- 11 Student L: What means?
- 12 Student M: Can you write dot?
- 13 Student A: If my partner don't know, can I ask another friend?
- 14 Teacher: Yes.
- 15 Student B: Dot — point.
- 16 Student K: Point? You put it.
- 17 Student O: What the place where you put it?
- 18 Student E: Is a point?
- 19 Student L: This — dot?
- 20 Teacher: All right?
- 21 Student A: All right.
- 22 Student M: But dot, could you write? — spell it?
- 23 Student E: What is the spelling of dot?
- 24 Teacher: How do you spell dot?
- 25 Student O: D-O.
- 26 Teacher: Can anybody tell me?
- 27 Student I: D-U.
- 28 Student O: D-O-P-T.
- 29 Student I: B-T.
- 30 Student J: Cross.
- 31 Student I: Dubt.
- 32 Student K: Only D-O-T.
- 33 Students: Dot.
- 34 Teacher: OK?

There are four basic topics within the extract: the identity of the word 'dot' (turns 1, 3, 4, 6, 10, 15); its spelling (12, 22-29, 31-33); its meaning (5, 7-9, 11, 15, 18, 19); and the appropriate location on the square (16, 17). There is also a procedural question to the teacher (13, 14). The purpose of

the remaining turn (student J's contribution at 30) remains – to me, at least – impenetrable.

If we concentrate on the teacher's speaking turns, we see that her participation involves sanctioning/confirming (turn 14), checking students' comprehension or readiness to proceed (20, 34), and elicitation (24, 26). In this particular situation, then, in contrast to the first one we analysed, she neither initiates any clarification of the meaning of 'dot' nor responds directly when student L requests lexical explanation (turn 11).

Apart from reflecting the intricacy of discourse where several parallel student-teacher and student-student interactions are being achieved or attempted, this short but hectic extract gives rise to this question: Why did the teacher decide to leave the learners to negotiate their own explanation of the problematic term 'dot' in this situation, when she did not in the map extract?

The most straightforward reason for the apparent difference in approach to what the learners signalled as lexical gaps in the two extracts under study here is that, for whatever reason, she believed that 'silk mill' represented a greater obstacle than 'dot'. She may normally operate with an underlying pedagogic principle that if less than a certain proportion of the members of a class are likely to know (or to be able to work out) the meaning of a word, she provides the explanation herself. If, on the other hand, she has grounds for thinking that more than that proportion has access to its meaning, she attempts first to elicit that meaning from the class.

As to the teacher's precise criteria for this sort of assessment, they might be of various sorts. Perhaps she thought that, given the task and materials involved, 'dot' would be more easily recoverable from context than 'silk mill' in the map task. Perhaps she believed that, since the learner group included several graduates in technical and scientific fields, it was likely that they would have encountered the term before, despite their low general proficiency in English.

Of course, it could be that this particular teacher employs a quite different operating principle for dealing with lexical problems; her method might be deliberately to alternate teacher explanation with student explanation. What matters, it seems to me, is that as observers we *cannot know* what the motivation is for particular classroom behaviours. Only the individual teacher has access to that 'privileged' information – although, if my observations of my own teaching are any guide, even the teacher may not always make entirely conscious choices among alternative actions.

### *Applications*

The starting point for this particular study was a concern with the applicability of a single piece of teaching material with varying levels of learner. It was for this reason that I asked a number of colleagues both to try out the material with their classes and to allow me to record their lessons. What I found has led to my wider interest in why teachers take

the actions they do and in ways of enabling teachers to analyse those actions (and what gives rise to them) for themselves.

I envisage a number of ways in which teachers could engage in the sort of 'why' questioning that I have outlined here, either individually or in groups. One would, of course, be for the teacher simply to reflect on lessons. No doubt we all do that to some extent, but probably in a rather piecemeal way, relying on the salience of specific incidents and on the general vagaries of memory.

A more reliable method would be for teachers to record some of our own lessons on audiotape and to listen to them in their spare time. This would enable us, for example, to pinpoint particular moments when – as the recording helps us recall – we made a decision to do X rather than Y (or Z etc). Can we remember the reason for taking that alternative? (I would advocate the use of audio- rather than video-recordings for reasons of 'portability' and access.)

A third method would be to adopt a cooperative peer group approach. Colleagues might agree to meet weekly to view a videotaped lesson by each of them in turn – perhaps on the same skill area. (Here I would suggest the use of video because a sound recording would convey insufficient information for those who had not been present at the lesson to get an idea of what was going on in the classroom.) The viewers would be able to stop the tape at any point and explore the motivation behind specific procedures; equally, the recorded teacher would have the chance to ask his or her colleagues what decision they would have made – or believe they do make – in situations such as that depicted. Such an approach would have costs as well as benefits: the discussion could easily become adversarial – my experience is that we can be oversensitive and overprotective about what we do in the classroom; but it would be an opportunity to see our own actions through others' eyes.

The methods outlined here are possible routes to the exploration of the classroom by and for practising teachers. If the aim was to conduct 'external' research (ie investigations about teaching in general by researchers from outwith the institution in question), then the various methods would need extension, such as through retrospective interview. As should be clear, my own main interest is in auto-research.

## Conclusion

The brief samples I have discussed are, I believe, indicative of typical minute-to-minute decisions we take as teachers. They illustrate the way in which any tactical decision to adopt a particular classroom procedure may be entwined with our beliefs about language learning. For me, one of the values of research into *what teachers do* is its potential for raising our awareness by obliging us to consider *why we do it*. At times the reasons will be local and particular. At others, such as in the case of the 'silk mill' episode, we may be able to detect the influence of underlying assumptions about and attitudes to teaching and learning, or different perceptions of



the task in hand. Current technology makes it easy for us to capture what is said and done; however, it is essential that we look beyond the data as external behaviour, in order to work out the beliefs underlying those actions. It is surely through reflecting on those beliefs that we can foster individual change and development.

## References

- Anderson, A. and Lynch, T. (1988) *Listening*, Oxford: O.U.P.
- Brown, G. Anderson, A. Shadbolt, N. and Lynch, T. (1987) 'Listening comprehension skills in secondary pupils', summary of Scottish Education Department report, SED/University of Edinburgh.
- Chaudron, C. (1988) *Second Language Classrooms*, Cambridge: C.U.P.
- Hatch, E. (ed.) (1978) *Discourse Analysis and Second Language Acquisition*, Rowley, Mass: Newbury House.
- Henzl, V. (1974) 'Linguistic register of foreign language instruction', *Language Learning* 23/2: 207/222.
- Long, M. and Sato, C. (1983) 'Classroom foreigner talk discourse: forms and functions of teachers' questions' in Seliger, H. and Long, M. (eds) *Classroom Oriented Second Language Research*, Rowley, Mass: Newbury House (268-286).
- Lynch, T. (1988) 'Getting learners to ask the questions', paper presented at the Joint IATEFL/TESOL Scotland Conference, Edinburgh, April 1988.
- Mitchell, R. Parkinson, B. and Johnstone, R. (1981) *The foreign language classroom: an observational study*, Stirling Educational Monographs no. 9: Department of Education, University of Stirling, Stirling, Scotland.
- Pica, T. Young, R. and Doughty, C. (1987) 'The impact of interaction on comprehension', *TESOL Quarterly* 21/4: 737-758.
- Ramirez, A. and Stromquist, N. (1979) 'ESL methodology and student language learning in bilingual elementary schools', *TESOL Quarterly* 13/1: 145-158.
- Robinson, E. (1981) 'The child's understanding of inadequate messages and communication failure: a problem of ignorance or egocentrism?' in Dickson, W. (ed.) *Children's Oral Communication Skills*, New York: Academic Press.
- van Lier, L. (1988) *The Classroom and the Language Learner*, London: Longman.
- White, L. (1987) 'Against comprehensible input: the Input Hypothesis and the development of second language competence', *Applied Linguistics* 8/2: 95-110.

# Learner Diaries

Brian Parkinson and Christina Howell-Richardson

*University of Edinburgh*

## 1 Introduction

The interest of the language teaching community in learner diaries is very recent, having been aroused by publications such as Bailey 1983 (and the earlier works mentioned in that article) which showed how different the learner's view of classroom processes and out-of-class learning could be from that of teachers and researchers. There had, of course, always been some awareness of such differences, but the diary evidence helped to bring them home in a new way, convergent with the 'focus on the learner' gaining ground in other areas and encapsulated in one of the rare poems to appear in a collection of academic articles (Farrington 1986).

Bailey's study and those like it are, however, atypical in an important way: the diarists are linguists, usually in both senses of the word, going 'back to school' after (and usually during) work as teachers/researchers/teacher educators, and their perceptions are inevitably affected by this: they sometimes find their theoretical sophistication surprisingly unhelpful, but it cannot fail to colour their report.

At the Institute for Applied Language Studies we have experimented with learner diaries on a variety of courses: on the *MSc in English Language Teaching* (also atypical learners!), but more extensively in the two studies reported here, both within the framework of our 'Learner Variables' research project – the first on our *General English* course, the second on *Modern Language* courses.

## 2 The 'General English' study

### 2.0 Introduction

'General English' is a full-time (20 contact hours per week) course which runs throughout the year, with students arriving and leaving at irregular intervals, and studying for periods from two weeks to two years. There are usually between 30 and 40 students in four or five classes from elementary to advanced. The rate of improvement of students during

their stay, as measured by reliable internal tests, varies widely. The purpose of the 'Learner Variables' project was to explore reasons for such differences, and the learner diaries were one of several means used to do so. After an exploratory study with 23 students in 1985-6, the main research was carried out with 51 students in 1986-7. These are divided into 'cohort 1' (29 Ss, Autumn 1986) and 'cohort 2' (22 Ss, Spring 1987), who for practical reasons were treated slightly differently.

### 2.1 Diaries – format and administration

The diaries consisted of one A4 sheet (two sides) for each day, and were to be filled in over a period of seven days (cohort 1) or ten days (cohort 2). They consisted mainly of short headings/questions, with plenty of blank space for the answers.

For cohort 1, side 1 of the diary contained the headings 'in-class activities', 'out-of-class activities', 'my problems' and 'what have I learnt'; side 2 was a grid with the headings 'activity', 'time', 'skills' and 'how much'. For cohort 2, these questions were rephrased and slightly reduced, and a new section, 'your own ideas', was added.

One of the two researchers visited each class at the beginning of the seven/ten day period to explain the projects, and a sheet of guidelines was also given out. Students were asked to fill in one sheet every day (including weekends, for which a slightly different format was used). Sheets were collected daily; in cohort 1 they were collected by the class teacher and open to inspection by other teachers, in cohort 2 they were collected and seen only by the two researchers. (The change was made because some cohort 1 students had apparently been inhibited by knowing that their teacher would see their comments; the alternative, however, was not fully satisfactory either, as some students felt their criticisms were being ignored.)

### 2.2 Diary analysis

The original plan was to quantify diary responses on two dimensions, *informativity* and *use of English outside class*, as in the 1985-86 study (Parkinson and Higham 1986).

Preliminary inspection of the diaries revealed a number of entries apparently indicative of learner *anxiety*, and as this is often mentioned in the literature (eg Bailey 1983) as a variable affecting learning, we decided to make it a third category. Each of the categories was sub-divided, as indicated in the following sections.

#### 2.2.1 Informativity

##### 2.2.1.1 Identification of the data

The analysis table (not reproduced in this paper, but summarised in Section 2.3) gives a quantitative record of the informativity rates of the learner journals. A unit of informativity is registered for each separate,

identifiable item of information *concerning newly-acquired knowledge* within or beyond the context of formal instruction. Only specific comments, describing in concrete terms what has been taught or learned, or specific evaluative comments are counted as relevant.

Under this criterion general comments such as –

‘We did grammar’

‘We read a story’

‘I learnt some new words’

are not considered ‘informative,’ whereas comments containing an informativity value equivalent to –

‘We studied the contrast between the present perfect and the simple past’ qualify as a single unit.

#### 2.2.1.2. *The table and categories*

*Column 1* The first column records the following basic personal clerical details – name, country of origin and sex.

*Column 2* *Informativity rates on ‘What was covered in class’*

The column records units of informativity on the discrete targets of the lesson and the discrete areas of linguistic input covered, as perceived by the student. In order to qualify as valid, any entry must include specific identifiable details.

*Column 3* *Informativity rates on ‘What was learned in class’*

This column records the number of linguistic items which the student mentions as having been learned within the context of formal instruction.

*Column 4* *Informativity rates on ‘Evaluative comments’*

As some of the students made diary entries evaluating the contents of individual lessons or commenting on the perceived value of the syllabus/course content as a whole, it was necessary to include a column in conjunction with and relating directly to columns 2 and 3. This category may be further sub-divided into four kinds of comment:

1.(a) To signal the desire to produce a change in established classroom procedures by making suggestions or criticisms which would result in procedures to date untried within this particular classroom setting.

*Example 1* ‘We have been doing exercises in class what I really think is a waste of time.’

*Example 2* ‘I think there are too many people in the class.’

1.(b) As 1.(a) – the global aim is identifiable as a desire to produce change. In contrast to 1.(a) however, the method employed is to praise significantly a newly-introduced method or class activity and to call for this to become established practice.

*Example* ‘... (the teacher) ask(ed) students about the new words.  
... I think that the good way for learning new words and  
we must do it much.’

2. Praise or criticism of an *individual* teacher or lesson, distinguishable and separate from 1.(a) and 1.(b) as there is no evidence of the wish to initiate structural changes.

*Example 1* ‘... (the teacher) led the discussion very well.’

*Example 2* ‘... (the teacher) is interested in our progress...’

3. To express dissatisfaction with the mismatch between the course content and the student’s own conceptualised needs. *No attempt is made to offer alternative procedures.*

*Example* ‘I am very “weak” in grammar, what I think we don’t pay much attention on it.’

*Column 5 Informativity rates on “What has been learned outside class”*

The column records the informativity rates on specifically identified items of knowledge, acquired (or learned) beyond the context of formal instruction. The scope of the items recorded was broad, including such diverse items as idioms and colloquialisms, cultural insight (eg the Scottish/English cultural conflict), practical transactional know-how, eg the format of a British Telecom bill, etc.

## 2.2.2 Anxiety levels

### 2.2.2.1 Identification of the data

The units recorded in the table each reflect one discrete item of information which is judged to reflect anxiety on the part of the language learner. Despite the expectation that difficulties would arise in clearly identifying units of anxiety, ‘covert’ expressions were rare. Accordingly, any entry which did not overtly reflect anxiety – given the full linguistic and non-linguistic context – was not included.

Subsequently, the following sub-types of entry were identified:

(a) *Self-derogatory comments*

‘I am slow[ly] to understand.’

‘I’m stupid.’

‘Other students can listen and read[ing] better than me.’

(b) *Entries including lexical items drawn from the semantic field of anxiety*

Examples of these lexical items are:

stressed    ashamed

unhappy    (I) worry

nervous    depressed

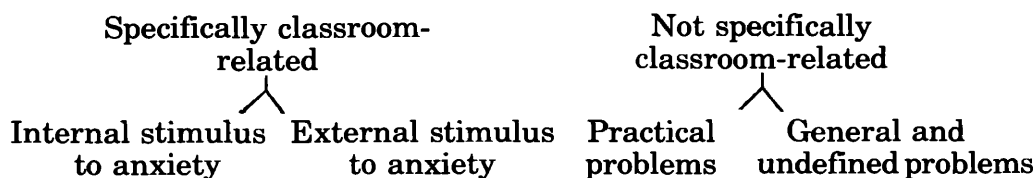
confused    frustrated

- (c) *Pre- or post-modification of a factual statement*  
 ‘I don’t understand the teacher, *when she talks too quickly.*’  
 ‘I feel angry at myself, when I stick at easy situations.’
- (d) *Wishes or desires concerning a change in the present state of affairs*  
 ‘If I knew more Scots people, I’d feel as a person and not as a language learner.’  
 ‘I have to try to go out.’
- (e) *Comments expressing dissatisfaction with formal instruction*  
 ‘The lessons don’t match my needs. I need grammar.’
- (f) *Rhetorical questions*  
 Some diaries include entries which at face value appear to be rhetorical questions, but in fact seem most reasonably interpreted as a message or question to the researcher. This phenomenon occurred in the case of two students, both of whom were taught by one of the researchers.  
 ‘Am I studying wrong?’  
 ‘How can I find accommodation for less than £200? Somebody must help me with accommodation.’
- (g) *Descriptions of reactions to L2 situations*  
 The students commented on L2 situations which they experienced as stressful.  
 ‘When it’s difficult to speak, I isolate myself.’
- (h) *Practical problems*  
 An item measures an entry in which the student is able to identify one particular practical problem giving rise to anxiety, such as insomnia, accommodation difficulties or bad news. In all cases, the student has both identified and labelled the source of anxiety.

#### 2.2.2.2. *The table and categories*

Application of the criteria described above yielded data which is broadly divisible into two groups — ‘classroom-related anxiety’ and anxiety which is not specifically attributable to the context of formal instruction; in so far as no specific reference is made to the learning situation within the classroom.

These two broad groups may be further subdivided as follows:



#### *Specifically classroom-related*

##### *Sub-Category I — Internal stimulus*

The column measures units of informativity reflecting a derogatory image of oneself as a language learner or units which overtly express

anxiety or distress in relation to a specific task or activity undertaken in the classroom. All the entries are characterised by their introspective or emotive content and the *absence* of the identification of any person or object as the actual or potential source of anxiety.

*Sub-category 2 — External stimulus*

The data consists of entries which specifically identify a person (or persons) or situation within the classroom context as a source of anxiety, but exclude any reference to a derogatory self-image.

These entries are of the type:

‘I feel no incentive from the class.’

‘It is useless to speak in English with others who are of the same nationality.’

Only two examples in the sample proved ambiguous in interpretation:

(a) ‘I don’t understand the teacher when she talks too quickly.’

(b) ‘I haven’t yet integrated in the class.’

Both units were eventually included as valid data. The post-modification in example (a) implies that blame is placed on the teacher for failing to adjust her speech patterns. Example (b) is also creditable as the student concerned expresses her reaction to a new group, and later returned to the original class.

*Not specifically classroom-related*

*Sub-category 1 – Practical problems*

The key-word is ‘concrete’. Difficulties which are well-identified practical problems unrelated to the learning situation, such as insomnia, receipt of bad news, or accommodation difficulties, form the data.

In one case entries concerning accommodation problems are repeated daily and are specifically identified by the student as debilitating and directly detrimental to his capacity as a language learner.

‘Accommodation, I think about accommodation more than about my lessons.’

It is impossible to draw similar conclusions in regard to the data received from other students, as this requires inference beyond what is justified by the data.

*Sub-category 2 – General and undefined problems*

The category was intended to be very general: to cover data which was too diffuse and undefined to merit separate categories.

However, it would appear that the majority of the entries forming the data for this category express anxiety in the form of the student’s lack of confidence in his/her linguistic competence or frustration at the difficulties involved in obtaining social contact with native speakers:

‘I can’t understand the Scottish accent.’

‘. . . not learning. . .’

‘. . . don’t meet people. . .’

### 2.2.3. *Out-of-class activities in English*

The table gives the amount of time each individual student assessed him/herself as being engaged in 'linguistically relevant' activities, over and above the hours of formal instruction, during the period in which the data was collected.

The figures recorded in the diaries are clearly subjective approximations to the actual length of time spent engaged in any one activity. A further variable to be taken into account is the value placed on various activities by the student himself. Clearly what a teacher or researcher may regard as 'linguistically relevant' is not always valued as such by the student diarist and was consequently omitted.

The sub-divisions of 'out-of-class' language use were as follows:

- (i) Private study, including homework.
- (ii) Social interaction in a group including a native speaker of English.
- (iii) Social interaction in a group not including a native speaker of English.
- (iv) 'Transactional language use', eg shopping, booking a holiday.
- (v) 'Receptive event', eg TV, cinema, attending a lecture.

## 2.3 Results

The number of entries for each of the categories of informativity, anxiety and out-of-class activity was as follows:

*Table 1*  
**Informativity**

	Number of Units									
	0	1	2	3	4	5	6	7	8	9+
'covered'	2	6	9	9	5	5	6	7	5	1
'learned'	11	17	9	8	4	1	0	0	0	1
'evaluative'	15	21	12	1	0	0	0	0	1	1
'outside class'	36	12	1	1	0	0	0	0	0	1

*Table 2*  
**Anxiety**

	Number of Units									
	0	1	2	3	4	5	6	7	8	9+
'internal'	35	7	2	2	1	0	1	0	1	2
'external'	19	17	6	4	3	1	0	0	0	1
'practical'	28	11	6	4	0	0	0	0	1	1
'general'	30	12	4	4	0	0	0	0	0	1



**Table 3**  
**Out-of-class activity**

	Number of Units									
	0	1	2	3	4	5	6	7	8	9+
'private'	7	4	6	9	6	2	3	2	9	3
'social (+native)'	6	4	10	5	6	4	3	4	7	2
'social (-native)'	9	6	7	3	7	5	2	5	5	2
'receptive'	8	4	1	8	7	8	0	0	2	13
'transactional'	Data not available									

## 2.4 Conclusions

Comparison of the results with test data showed a high correlation (+0.33, sig 0.02) between rate of improvement and the amount of time which students spent outside class in social interaction with native speakers of English. No other diary variables correlated significantly. This confirms the 1985-86 finding, and refines it by showing which kind of out-of-class L2 use appears to be most effective. The absence of clear effects of anxiety is surprising, although there may have been a non-linear relationship (cf. Hatch and Farhady 1982 p. 204) which we failed to detect.

We felt that the diaries gave us a lot of valuable information about our learners, later pursued in interviews, but that their usefulness in this study was somewhat compromised by a lack of clarity concerning aims, and the conflicting requirements of research uses of diaries and other possible uses, including course evaluation, student counselling and language practice. The multiplicity of diary uses can sometimes be a handicap rather than a benefit, an issue further discussed in Howell-Richardson and Parkinson 1988.

## 3 The 'modern languages' study

### 3.1 Introduction

We now turned our attention to a very different group of learners, the local Edinburgh people attending Modern Language classes at our Institute. In this case, the diaries were designed mainly to collect information in two areas:

- (a) *Use of FL outside class* The EFL study had shown this to be an important factor in rate of improvement, so we wanted to establish what possibilities for this existed in the apparently less favourable environment of once-a-week L2 learning by Britons in Britain. For how many hours a week were they exposed to the L2 outside class, in what kinds of activity, and involving which of the 'four skills'?

- (b) *Use of 'strategies'* In earlier studies and in our everyday life as teachers we had been struck by the greater variation among learners in the number and nature of 'techniques' or 'strategies' which they used, successfully or otherwise, to facilitate learning. We thought that this area, which we had not so far systematically explored, might account for some of the as yet unexplained variability between learners, but before attempting a process-product study we wished to get an overview of the range and kind of strategies reported.

We were aware of a recent tradition of research, by Rubin and others, on learner strategies, leading to the learner handbook Rubin and Thompson 1982, with its list of fourteen strategies and numerous sub-strategies (see Appendix). The 'Rubin tradition', like the 'Bailey tradition', is however skewed towards more educated and successful learners, and we wanted to see which strategies were used by apparently more 'normal' learners, either spontaneously or after prompting.

Unlike our earlier work, this study had a 'before and after' aspect. We prepared a handout summarising the strategies recommended by Rubin and Thompson, and after learners had submitted several weeks of diaries they were given the handout with a covering page saying, *inter alia*, 'You can try out as many or as few of the strategies as you like.' 'Some may be things you do already, others may not be right for you, because everybody learns in different ways.' 'We would like you to continue with the diaries until the end of term, in order to let us know if the strategies have influenced your learning.'

The 'pre-handout' and 'post-handout' entries were examined for any effects. (As the list of strategies was so wide ranging, we expected effects on L2 use and other dimensions and not merely on 'strategies' in our narrower definition.)

### *3.2 Interim results*

We have so far analysed returns from learners of only two languages, French and Spanish, in only one of the above areas, use of 'strategies.'

#### *3.2.1. Pre-handout results*

*Findings:* Tables 4 and 5 give a quantitative account of the range of strategies recorded over a pre-handout period of six weeks. The activities recorded exclude exercises or assignments set by the teacher. The basis for coding differs from that used in tables 1-3 in that one unit represents the overall participation of one student, irrespective of the number of individual entries recorded. This change was on practical grounds: as the subjects concerned, in contrast with 'short-term' EFL learners, lead relatively stable lives, they tended to adopt strategies which were incorporated into long-lasting language-learning routines. In this respect they seemed clearly different from the EFL diarists, who were more influenced by variable external factors.

**Table 4 Preliminary findings: Pre-handout: Learners of French**  
Total no. of entries

Strategy	Advanced	Higher Intermediate	Lower Intermediate
TV Film	7	2	2
Novels in TL	7	–	–
Newspapers in TL	5	–	2
Cinema – TL	7	3	–
Radio in TL	1	–	–
Media course	5	–	–
Role-play	1	–	–
Grammar exercises	2	1	–
Extensive dictionary use	1	–	–
Vocabulary exercises	1	–	1
Taped songs	–	–	–
Performing everyday activities in TL	2	–	–
Daily practice	2	–	–
Interaction with native speakers	7	3	–

### 3.2.2. Post-handout results

We have not included tables on post-handout strategies: surprisingly, there was little change in overall quantitative terms, but there was a significant difference in the narrative style and in individual subjects' language learning attitudes. The most striking changes were as follows:

1. The narrative tone of the diary entries was sometimes replaced by evaluation of the validity of the strategies suggested in the light of the diarists' own circumstances and learning-objectives, or comparison of these strategies to approaches already in use.
2. The narrative tones of the entries was sometimes replaced by introspective accounts of the learner's experiences of using one or more of the strategies recommended in the handout.
3. Perceived success or failure in the performance of tasks in the TL was analysed with reference to the strategies proposed.  
'I'm much better at matching [French] idiomatic phrases with their meanings. I'd attribute this to being good at guesswork.'  
(cf. Strategy 10: 'learn to make intelligent guesses.')
4. A number of diarists assumed a more personal relationship with the researchers than previously, addressing them directly as in:  
'Your handout'  
'I can see how you might find this useful for some students...'
5. The diary entries in general were much shorter than previously, and many diarists transferred to note-form.

**Table 5 Preliminary findings: Pre-handout: Learners of Spanish**  
Total no. of entries

Strategy	Higher Intermediate	Lower Intermediate	Elementary	Beginners
Recording self	–	–	1	–
Media language course	–	3	23	4
Reading books in TL	1	–	–	–
Vocabulary exercises	1	–	5	–
Setting and evaluating own goals and needs	–	–	4	1
Grammar exercises	3	–	8	1
Role-play	–	–	1	–
Extensive dictionary use	–	1	2	1
Drawing on present resources eg Latin, French, English grammar	–	–	1	–
Reading newspapers in TL	–	–	–	1
Daily practice	–	–	1	1
Interaction with native speakers	–	–	3	–
Re-organisation of personal notes	–	–	1	2
Performing everyday activities in TL	–	–	1	1

6. Certain diarists, without reference to previous strategies used, or attitudes recorded, appear to have altered their approach to language-learning. Generally, as in the following example, there is no indication of whether this is a conscious shift.

Wk.3 – ‘When I’m watching a film on television by scanning the subtitles I can usually recognise part of the dialogue.’

Wk.6 – ‘I covered the subtitles. It’s amazing how you can concentrate harder if you need to.’

7. A number of diarists attempted to adapt Rubin and Thompson’s suggestions to their own circumstances. This, however, was usually by a process of trial and error. One learner writes in response to Strategy 4, ‘make your own opportunities’:

Wk.7 – ‘It is difficult to get face-to-face practice which allows for question-and-answer feedback.’

Wk.8 – ‘I advocate the use of rehearsal aloud with as critical an ear as possible.’

### 3.3 Conclusions

Although the MI study is still at a preliminary stage and has gathered a small amount of data — subjective, self-reported data — we already have an impression of certain noteworthy tendencies among the ML group.

Rubin and Thompson (1982) include both cognitive factors and personality factors as being equally relevant and important to successful second-language acquisition. Our findings suggest that while the majority of the group initially showed a marked preference for traditional cognitive strategies [Tables 4 and 5], enhanced performance was later reported by those who adopted or used strategies of the type 2, 4, 9 and 10.

The change in the style and format of the post-handout entries towards a more analytic approach is also of interest. This shift tended to be accompanied by greater emphasis on the learner's approach to the FL material, in contrast to the narrative descriptions of the material which were a feature of pre-handout entries.

We have clearly established that diaries can be a rich source of information about learners. The main problems lie in refining research techniques so that this information becomes more fully interpretable and reliable, in integrating diaries with other research and teaching tools (questionnaires, interviews, tests, observation, class discussion) and in picking our way through the ethical minefield so that research can be of maximum value to present and future learners.

### References

- Bailey, K. M. (1983). 'Competitiveness and anxiety in adult FL learning: looking at and through the diary studies', in Seliger, H. W. and Long, M. H. (eds) *Classroom Oriented Research in Second Language Acquisition*, Rowley, Ma.: Newbury House.
- Farrington, O. (1986) 'As teacher, as learner', in Meara, P. (ed.), *Spoken Language. British Studies in Applied Linguistics 1*. London: CILT.
- Hatch, E. and Farhady, H. (1982) *Research Design and Statistics for Applied Linguistics*, Rowley, Ma.: Newbury House.
- Howell-Richardson, C. and Parkinson, B (1988) 'Learner diaries: possibilities and pitfalls', in Grunwell, P. (ed.) *Language in Society. British Studies in Applied Linguistics 3*, London: CILT.
- Parkinson, B and Higham, J. (1986) 'Are there discrete learner types?', paper presented at Annual Meeting of British Association for Applied Linguistics.
- Rubin, J. and Thompson, I. (1982). *How to be a more successful language learner*, Boston: Heinle and Heinle.

### Appendix: The list of strategies

(from Rubin and Thompson 1982)

1. Find your own way.
2. Organise (information about the language, your own programme of study).
3. Be creative.
4. Make your own opportunities.
5. Learn to live with uncertainty.
6. Use mnemonics.

7. **Make errors work.**
8. **Use your linguistic knowledge.**
9. **Let context help you.**
10. **Learn to make intelligent guesses.**
11. **Learn some lines as wholes.**
12. **Learn formalised routines.**
13. **Learn production techniques.**
14. **Use different styles of speech.**

**Each of these strategies is expanded to 2-4 pages in the source text, and about a quarter of this, modified for the local context, in our handout.**

# Classrooms and Display Questions

Lynn Banbrook and Peter Skehan

*Institute of Education, London University*

## **The nature of research**

As this volume demonstrates, there are many ways of conducting classroom research. One approach is to use a coding scheme (Allen et al 1984). An alternative is to take a more ethnographic perspective (van Lier 1988). These two approaches, the coding and the ethnographic, differ in that the former provides the researcher with provisional hypotheses about the important categories that should be focused on in classrooms, ie the events which are coded (Long 1980), while the latter does not narrow and focus the data collection in the same way. But the two approaches are alike in that they both attempt to gather data without the clear formulation of specific hypotheses. In the ethnographic case, the data is collected and generalisations, if they are made, will emerge from what is salient from the observations. In the coding scheme approach, while the categories of the scheme are accepted prior to the investigation, there is a sense in which all the categories are equally valued, and it is the purpose of the research to explore which of these wide range of categories are the most important, the most causal, the most predictive etc. The coding and ethnographic approaches are alike, then, in that if one relates them to the traditional distinction between theory-then-research and research-then-theory perspectives (Larsen-Freeman and Long, in press), they both need to be located nearer the research-then-theory end of any such continuum. In either case, the attempt is made to avoid or at least minimise theoretical bias during research and to enable relevant data (in the case of ethnography) or patterns of classroom events (in the case of the coding schemes) to emerge. Causal statements, predictions, and abstract categories are not prominent, and it is the data itself which is emphasised.

Recently, however, a contrasting paradigm of research has emerged. The developments that have taken place in second language acquisition research enable more theory-driven approaches to analysing second language classrooms to be explored. In particular it is proposed that

*teacher questions* have a particularly important role. They can be used to allow the learner to keep participating in the discourse and even modify it so that the language used becomes more comprehensible and personally relevant. Several teacher-question studies have now been completed. Long and Sato (1983) compared the numbers of display (teacher knows the answer) questions and referential (teacher does not know the answer) questions in 'naturalistic' and classroom discourse, and found that while in naturalistic discourse referential questions far outnumbered display questions, the reverse is true for lockstep, IRF (Initiate-Respond-Feedback) dominated language teaching classrooms (Sinclair and Coulthard 1975). Pica and Long (1986) confirmed the findings of Long and Sato (1983) and also compared the performance, with respect to display questions, of experienced *vs.* inexperienced teachers. They found that both types of teacher used far more display than referential questions. Drawing on these initial research studies, Brock (1986) examined the consequences of teachers changing the numbers of referential questions that are asked. She reported that a higher incidence of teacher referential questioning was associated with greater student length of utterance, and utterance complexity, as well as a greater number of teacher confirmation checks and clarification requests. White and Lightbown (1984) also report typical classrooms as being those in which the teacher asks all the questions, and produces a lockstep approach to classroom organisation as a result.

The implications of the teacher-question research findings are considerable. They suggest that there is a divergence between what theorists would consider to be good practice and what is actually going on in classrooms. The purpose of the present chapter is not to take (serious) issue with the findings of the teacher-question studies but to examine to what extent such a theory-driven approach to classroom investigation is currently achievable. It will examine, that is, whether the approach is able to demonstrate adequate internal as well as external validity. In the former case the main emphasis will be on conceptual issues and particularly whether the categorisation of questions which underlies the approach can be sustained in practice, or whether an excessive degree of idealisation is necessary. In the latter case, the discussion will centre on how generalisations are made in classroom studies, and how the findings from any particular piece of research can be linked to more general contexts of teaching. In this respect, the issue will be whether we are dealing here with a sustainable claim to generality typical of a theory-then-research orientation or whether we need to take more seriously the research-then-theory perspective, (ie generalisation is achieved only through more extensive replication with systematic variation in many contextual factors) (Larsen-Freeman and Long, in press; Skehan 1989). The argument will be illustrated at various points by data drawn from Banbrook (1987), which was an empirical replication of many of the claims made by the studies briefly described earlier, and in which more extensive discussion of many of the issues is provided.



### Conceptual issues

There are two major conceptual difficulties to confront. The first of these is to ask what constitutes a question. One wonders how exchanges such as the following should be handled:

(1)

S 7: Don't losing weight.

T: OK. (to the others) *Can you help him?*

... Not 'don't'. Don't say 'don't'. Use the gerund. OK. So.  
(Banbrook (1987, p. 40), underlining added.)

The issue is whether the *Can you help him?* functions as a question, or is meant to be, or interpreted as, an imperative. Without knowing the discourse context, it is difficult to decide. We also have the problem of whether to count as questions utterances which are not interrogatively marked, as the following extract makes clear:

(2)

T: OK. What's this word?

Sev.Ss: (Unclear)

T: What's this one?

Sev.Ss: I'd.

T: OK. Together.

All: I'd.

T: Good. This one?

Sev.Ss: Known.

T: Yeah?

(Banbrook 1987, p. 41)

The problems are located towards the end of this excerpt. 'This one?' connects discursively with the very first teacher utterance, and is meant to cue a word. 'Yeah?' is used to indicate disbelief at the students' answer of 'known'. Both of the problem utterances were used with rising intonation, and were interpreted as questions, although not overtly marked as such. How would one treat them? There are other examples in Banbrook (1987) which reveal the use of paralinguistic cues by the teacher to indicate that a question is intended (and a student response expected). Teacher and students are adept at interpreting the rules of classroom interaction to work out what is being orchestrated in each case.

The problem of counting questions spills over into decisions concerning the *range* of functions served by display questions. In the next example, the teacher uses successive questions which are all aiming at the same answer but yet the additional questions seem to be of more value in the discourse than mere repetitions. The students are looking at a page where there is a diagrammatic representation of the path of someone's life including his missed opportunities.

(3)

T: (. . .) Yeah. Any more? What about the end? Top? Any at the top?

The potential value as input seems more than just 'one' question. Yet paradoxically coding this utterance as four display questions does even less service to its value as input.

The second major conceptual issue concerns the clarity with which questions can be assigned to display or referential categories. In the next example, referential questions seem to have 'display' features. The teacher is talking about keeping fit, and has elicited the word 'jogging' with a display question. But now he asks the student a more personal question.

(4)

- T: Good. Jogging. Yeah? OK.  
Do you jog?  
S 10: No.  
T: Good. Are you fit?  
S 10: No.  
T: All right. OK.

You could say that 'Do you jog?' is a referential question because the teacher does not know the answer and the student can choose. But two things about it are definitely 'display': first, the 'Good' is more of an evaluation of the student having coped with the question than of the fact that he does not jog (and this puts the exchange firmly into a display IRF mould); and secondly, the overall aim is probably to check that the student has registered the new lexical item or to give him a chance to relate it to his own life. So it could be maintained that questions of this type are both display *and* referential.

'Are you fit?' is similar but shows an increased 'display-ness' because the teacher, and probably also the students, see it as a 'logical' follow on from the 'jogging' question since it was previously established that jogging and fitness go together. However, it is not inconceivable that a student *could* still break the mould here and say, for example, that yes, he was fit because he went swimming every day.

There are also occasions when display questions can become, through interpretation, more referential in nature. We looked a little earlier at some ways of eliciting examples of missed opportunities in the 'path of life' exercise. This example from the same activity shows that although the eliciting questions aim at a display of the third conditional, it also works as a real (referential) question:

(5)

- T: Yes. OK. Very good. Any more?  
Maria, can you give me one or not?  
S: No.  
T: Or somebody else?

since while the structure must be paraded, it must also be related to a personal context.

Another perspective on the problem of categorising questions comes from comparing the different sorts of questions that were used in two of the classes which were investigated by Banbrook (1987). Lesson 3 and Lesson 6 both contained a variety of display questions, and in generally similar proportions to the total. However, Lesson 3 had more questions that focused narrowly on word meaning, while lesson 6 had more display questions of a broader sort.

It is not that the narrow questions in Lesson 3 produced no interaction. They often did, eg:

(6)

- T: What sort of industrial places produce a glow? What is a glow?  
 S 4: (brief inaudible mutter)  
 T: Sorry? Glow?  
 S 7: It has something to do with air pollution.  
 T: No, not necessarily. . . No? Nobody knows it? Do you know what a glow worm is? . . No? A fire produces a glow. It's a warm light, a warm red light. G-L-O-W. If you're in a room and all the lights are off but there's a wood fire or a coal fire (S1 mutter) then the room is lit by the glow of the fire. A warm light. (. . .)

It is rather that the questions in Lesson 6 produced *more* interaction, with more coming from students. Further, because the answer was less predictable in that the student might choose one of a range of things to say, there was more scope for negotiation of meaning (although this is not evident in the conversational adjustment figures). For example:

(7)

- T: (. . .) Why is the military in this particular park?  
 S 2: Why?  
 S 3: Because the military needs, er, need for a big area to do 'practick'.  
 S 5: Oh, and they use the park?  
 S 3: Yes, they have to do 'practick' and it's the (unclear) and the jungles, big areas they need. But at that time it's not useful for tourists and all the tourists they can't go there and the owner of restaurant, owner of campsite, caravan and campsite they are very angry because they can't get more tourists there to sell their produce and use from accommodation.  
 T: Yeah. OK. Right. (. . .)

In this way the interaction in the two lessons varied due, at least partly, to the different kinds of display questions asked.

It can be argued that display questions may be seen as varying along two separate but parallel clines:

N		B
A	Content	R
R	_____	O
R		A
O	Form	D
W	_____	

One cline relates to the predictability of the content, ranging from where there is only one possible answer (eg Is he inside or outside?) to questions where a range of possibilities exists, (eg How do you think people in National Parks make a living?). In neither case is the answer likely to be terribly informative for the teacher, but the student in the latter case does have to make decisions about what to select to communicate. The 'form' cline relates to the restriction of form in the response.

A question can be at different positions on the two clines. For example, one question may have a range of possible answers with respect to content yet might be understood to demand a particular syntactic structure; whereas the response to another question might have a totally predictable content but might be expressed in a variety of syntactic ways. It is also noteworthy that the 'form' cline is influenced not only by the question itself but by the kind of activity it occurs in and the students' interpretation of the teacher's intentions. (Such a view of a lesson as an event with its own sociolinguistic norms is endorsed for example by Stubbs 1975, Sinclair and Brazil 1982, Brown and Edmondson 1984 and Mehan 1979.)

The examples of display questions in Long and Sato (1983) and Pica and Long (1986) are narrow on both clines. This may be partly because this *is* the kind of question which occurred in their elementary classes, although Brock's (1986) examples for upper intermediate classes are also of the narrow kind. It is possible, then, that the variation of discourse as a result of variation in display questions did not arise in these studies. We would argue, however, that the term 'display question', when applied to the present data, turns out to be a bit of a blanket term and is less useful than expected in depicting the kind of teacher-student interaction which results. A finer sub-categorisation of this variable might lead to more interesting findings and might capture ways in which some teachers and some students are more able to breathe some life into an unpromising question type.

### **Issues of generalisability**

So far the focus has been on the micro-analysis of the categories used in the question studies and the need for contextually-bound interpretation of utterances. But there are also issues connected with the significance of these studies and with their generalisability. The issues here group under three main headings – the representativeness of the lesson samples for any particular teacher, the representativeness of the teachers who have been researched in relation to other teachers, and the nature of the generalisations that are made across learners.

With the first type of problem, one has to focus on how representative the sample of behaviour is that is used to investigate teacher questioning behaviour (or any other variable of interest, in fact). One class of decisions concerns how long should be sampled from a teacher's performance to obtain 'typical' data, ie how many minutes of observation,

how many lessons, how many lessons with different groups. But one must also consider how teaching activities cause differences in teacher performance as a function of their different communicative demands.

In general, classroom research has not addressed this issue of how one could justify one's sampling base and there is an urgent need for guidelines to enable the robustness of reported studies to be assessed. Several aspects of the question studies are implicated by this discussion. Pica and Long (1986), for example, base part of their results on a ten-minute extract from a lesson, with the ten minutes concerned being chosen by the teacher being researched, (who chose when to turn the tape recorder on to enable the data gathering, since no observer was present). One wonders also whether the nature of the activity that was the basis for the transcribed and analysed lesson data may also have exerted an influence. Long and Sato (1983) simply required that the teacher teach a normal lesson. Brock (1986) based her work on lessons taught by six teachers which each contained sections based on a reading passage and a vocabulary list. The problem we have with these findings is knowing to what extent the findings reported were a function not of these teachers in general but of the teachers concerned doing whatever activities they were doing that generated the experimental data.

Banbrook (1987, pp95-106), in developing the method of lesson profiling, (ie the lesson is tracked over time both quantitatively and qualitatively) demonstrates that there are clear differences in the number of epistemic (display) questions used at different phases, with this seeming to be linked principally to the nature of the activity being engaged in. A role play, and also a controlled exercise based on O'Neill (1971) generated the lowest frequency of display questions. In contrast, an elicitation exercise and a substitution exercise generated much more display-type activity. It is crucial, therefore, in studies of this type, to pay considerable attention to (and report) the nature of the instruction that is being provided.

There is also the issue of the extent to which teachers differ from one another. Long and Sato (1983) report an overall proportion of display to total questions of 79%. They also report (in their Appendix 2) figures for the six individual lessons that they studied. If the comparable proportion is calculated for each individual lesson, one obtains the values 82%, 45%, 83%, 83%, 69%, and 84%. This suggests little variation between four teachers (in the 82-84% range), but two teachers who are somewhat different (69%) and very different (45%). The other studies, and Banbrook (1987), show similar variation in proportions of display questions or overall number of questions. One can conclude from this data therefore that a teacher is not a teacher is not a teacher. In other words, any attempt to establish generality must take account of the extent of the variation between subjects. This is particularly important in the case of application, of using the results of research studies beyond original sources, and of advising what the impact of such studies should be on the profession in general, and teacher training, in particular. Gaies

(1983, p. 208) writes, for example, when covering the Long and Sato (1983) study, 'In the six ESL classrooms investigated' (*italics added*) when, as we have seen, one (or even two) of the classes were clearly different with respect to the use of display questions. Similarly, Brock (1986, p. 47) writes, 'Long and Sato (1983) found that *teachers* ask significantly more display questions. . .' (*italics added*). What is most interesting is how the variation which was clearly present in the original sources disappears in the retelling and is idealised into a near-uniformity and lack of variation when the original sources are being commented on.

The final 'generalisability' issue that needs to be discussed concerns learners. The approach taken in the teacher question studies is to use a variable whose importance is derived from SLA research, and then study its incidence in the ESL classroom. A move is being made, however, from the study of the individual learning process, with which SLA concerns itself, to the study of group (ie class) functioning, *with inferences being made about the impact this functioning has on the individual learner*.

There are certainly examples in the literature of individual students reacting differently to essentially the same classroom. Allwright's Igor (1980) exploited the language classroom for all it was worth, display questions or no, and focused a disproportionate amount of global quantity of classroom activity on himself. There are also students who influence the *quality* of classroom input and interaction. In the empirical work on which the present article draws (Banbrook 1987) there were many examples of students going beyond a basic thread of discourse in the classroom and investing it with more meaning. Banbrook (1987, pp. 86-95), this time using the technique of *learner* profiling, shows how two students whose quantitative indices of classroom performance (eg number of interactions, number of display questions answered, number of referential questions answered, number of clarification requests), were very similar, differed markedly when one examined what they were doing with the classroom interactions they engaged in. In a collaborative story-writing task the first student, Angelo, makes contributions which respond to the teacher's elicitations, and are very much in line with what the teacher wants with regard to content and form, and so the exchanges follow an IRF pattern. For example:

(Angelo is S 1)

(8)

T: (. . .) Stumbled and . . . ?

S 1: Fell down.

S 7: Fell down.

T: All right.

Sabine, on the other hand, is moved more than once to question the content of the story. In this example, Sabine begins by trying to clarify why the heroine became so flustered simply because the man behind her in the queue at the window was impatient.

(Sabine is S 3)

(9)

S 3: Because she di. . ., I don't understand.

S 8: The man behind her shouted.

S 1: She dropped her purse.

S 8: (Unclear)

S 3: So why didn't she move?

T: (to S 1) Angelo, what did you say? It's a good idea.

S 1: And she dropped her purse.

T: All right. Any other possibilities?

S 3: Why didn't she move? She should move there! (laugh)

T: Sabine, you're much stronger. You wouldn't have become nervous. Maybe, what type of person is Marth, do you think? Is she. . .?

S 4: Shy.

S 6: Shy.

S 3: Stupid!

T: Maybe very shy, yeah. OK.

Another example of differences in students occurs with Osman, a student in the same class as Sabine and Angelo. The class was doing a mechanical pattern practice exercise on the third conditional (O'Neill 1971, p. 151). The textbook has given the situation: A man has a bad headache this morning because he drank too much at a party last night. The response is meant to be on the pattern of 'I wouldn't have. . . if I'd been you.'

(10)

T: (. . .) What about C, Osman?

S 6: If I'd been you. . .

T: Yeah

S6: Er, I would have, er, I would have, er. . .

T: Positive or negative?

Other Ss: Negative.

S 6: I know, I know! I would have drunk a little bit more!

(All laugh)

T: OK. All right.

Osman is a risk-taker rarely participating in a straightforward predictable manner. As well as making little jokes out of display responses, he prefers to test out new ways of answering referential questions, for instance:

(11)

T: What time is it, Osman?

S 6: We have five minutes . . . is it right?

(It was 10.45 and the lesson had to end at 10.50 sharp.)

So Osman does not allow himself to be completely straightjacketed by a lockstep, display-question approach. He is able to raise the level of meaning and both explore language form as well as announce himself as

a real person. Sabine, similarly, was determined to resist the steam-rolling effect of display questions and to judge caricature EFL characters by more realistic standards.

## **Conclusions**

The display question studies have been taken as representative of the experimental and theory-generated approach to classroom research. As regards theory and conceptual issues, we have seen that there are problems in terms of what constitutes a question. We have also seen that it is by no means easy to categorise questions into display and referential. In general these problems derive from the way teachers and students are skilful participants in classrooms and transcend apparently limiting conventions. As a result, attempts at quantifying data into discrete and directly observable categories are hazardous. We have also examined how intra- and inter-teacher generalisations need to be established in the field of classroom research. The first sort of generalisation is affected by (a) general teacher variation, (b) variation that takes place over the phases of a lesson, and (c) variation in question asking (or virtually any other variable) that is the consequence of the teaching tasks or activities engaged in. In addition, there is the variation that exists between teachers. The issues we have here are how to decide how many teachers to include in any particular study to be able to claim representativeness, and the background variables about teachers that might enable a more principled sampling of such teachers. At the moment, we have variation in teacher performance (reviewed earlier), but no clear understanding of its parameters. There is also the issue of variation between learners in how they are able to invest classrooms with meaning. As a result, there will be a larger-than-necessary question mark against classroom studies in general (not simply those focusing on questioning behaviour) since the results obtained may have been a function of the particular teachers and learners sampled, rather than teachers and learners in general.

Given these points, two general suggestions are offered here for future classroom research. First of all, what is needed is a framework within which both quantitative and qualitative data can be synthesised. Such an approach to classroom research has been frequently advocated (see eg Long 1980) but is rarely put into practice effectively. Most published research, for example, seems to align itself clearly with a polarised position on the quantitative-qualitative dimension. This is to be regretted since such approaches risk either ignoring important aspects of classroom interaction, or oversimplifying them. Banbrook (1987) proposes one solution to this difficulty in terms of lesson and student profiling. Quantitative data is collected, but it is supplemented by tracking the lesson qualitatively over time. This allows the relative crudeness of the numerical data to be compensated for by relating it to transcripts of the classroom events as they unfold in time. It was lesson profiling which enabled Banbrook (1987) to focus on the distinction



between form-based display questions and content-based display questions, and so distinguish between two classes whose overall quantitative indices of display to referential questions were quite similar, thus revealing how one needs to develop interpretations of data based on both sources of information. Similarly, it was student profiling, once again based on lesson transcripts, which allowed Banbrook (1987) to demonstrate the differences in interaction pattern between students like Sabine and Osman, on the one hand, and Angelo, on the other. The profiling technique seems a fruitful one. At the very least it allows interesting multiple perspectives to emerge. More optimistically, it will generate suggestions for refinements in the coding categories that are used, as was the case with the different types of display question.

The second suggestion concerns the way in which generalisations are made in classroom research. At the beginning of this chapter two contrasting approaches to research were discussed. Each approach to research has its strengths and weaknesses. The theory-then-research approach will help the investigator eliminate the inessential and superfluous *if the right theory is chosen*. The research-then-theory approach may enable interesting generalisations to emerge from the data *if the situation is sufficiently simple for any obvious patterns to manifest themselves* (Larsen-Freeman and Long, in press). The teacher-question studies, being of the theory-then-research orientation, choose data to make prominent, and hope to reveal fundamental, and therefore generalisable, aspects of classroom activities. We have seen that, conceptually, there are problems with the operationalisability of the categories that are employed (and hence the need for qualitative data to complement the selected and categorised data highlighted by theory). But there are also problems with generalisability. We have seen significant intra- and inter-teacher variability that limits the way in which these theory-generated studies can be taken to be representative. Essentially, it is argued, there seems scope here to relate the research-then-theory requirement of establishing the robustness of generalisations through systematic study of a range of related situations to the teacher-question studies. At present, the insights concerning display and referential questions are of limited utility because we already know that individual teachers vary over time, and that different teachers may generate different question-use patterns. This would suggest that the most productive approach for classroom work at the present time would be to try to uncover variables which might constrain the generalisations made in classroom studies. In this way, a combination of both perspectives may be what is currently most appropriate to classroom research, given the complexity of the situation we are faced with.

## References

- Allen, J. P. B., Frohlich, M. and Spada, N. (1984) 'The communicative orientation of language teaching: an observation scheme', in Handscombe, J., Orem, R. and Taylor, B. (eds.) *On TESOL '83: The question of control*, 231-252. Washington D.C.
- Allwright, R. L. (1980) 'Turns, topics, and tasks: patterns of participation in language teaching and learning', in Larsen-Freeman D. (ed.), *Discourse Analysis in Second Language Acquisition Research*, Rowley, Mass.: Newbury House.
- Banbrook, L. (1987) *Questions about questions: an inquiry into the study of teachers' questioning behaviour in ESL classrooms*, *TESOL Quarterly*, 20: 47-59.
- Brown, G. A. and Edmondson, R. (1984) 'Asking questions', in Wragg E. (ed.), *Classroom Teaching Skills*, London: Croom Helm.
- Gaies, S. (1983) 'The investigation of language classroom processes', *TESOL Quarterly*, 17: 205-217.
- Larsen-Freeman, D. and Long, M. (in press), *An Introduction to Second Language Acquisition Research*, London: Longman.
- Long, M. (1980) 'Inside the 'black box': methodological issues in classroom research on language learning', *Language Learning*, 30:1-42.
- Long, M. and Sato, C. (1983) 'Classroom foreigner talk discourse: forms and functions of teachers' questions', in Seliger and Long (eds.)
- Mehan, D. (1979) *Learning Lessons*, Cambridge: Harvard University Press.
- O'Neill, R. (1971) *Kernel Lessons Intermediate*, London: Longman.
- Pica, T. and Long, M. (1986) 'The linguistic and conversational performance of experienced and inexperienced teachers', in Day R. (ed.), *Talking to Learn*, Rowley, Mass.: Newbury House.
- Seliger, H. and Long, M. (1983) (eds.) *Classroom Oriented Research in Second Language Acquisition*, Rowley, Mass.: Newbury House.
- Sinclair, J. McH. and Brazil, D. (1982) *Teacher Talk*, Oxford: OUP.
- Sinclair, J. McH. and Coulthard, R. M. (1975) *Towards an Analysis of Discourse*, Oxford: OUP.
- Skehan, P. (1989) *Individual Differences in Second Language Learning*, London: Edward Arnold.
- Spada, N. (1987) 'Relationships between instructional differences and learning outcomes: a process-product study of communicative language teaching', *Applied Linguistics*, 8:137-161.
- Stubbs, M. (1975) 'Teaching and talking: a sociolinguistic approach to classroom interaction', reprinted in *Language Development: Supplementary Readings for Block 5*, Open University 1979.
- van Lier, L. (1988) *The Classroom and the Language Learner: Ethnography and Second Language Classroom Research*, London: Longman.
- White, J. and Lightbown, P. (1984) 'Asking and answering in ESL classes', *Canadian Modern Language Review*, 40: 228-244.

### ***III Special Issues***



# Who Should Evaluate L2 Programs?<sup>1</sup>

Alan Beretta

*Michigan State University*

## Introduction

I have often wondered why program evaluation has been so neglected in our field. I used to think that the effects of the Colorado project (Scherer and Wertheimer 1964) and the Pennsylvania project (Smith 1970) were so traumatic that a generation of L2 researchers has been barely able to contemplate the subject without acute anxiety.

The burden of the past may be part of the explanation, but it is not the whole story. Similar disappointments were experienced by our colleagues in education (the nationwide evaluations of Follow Through, for example), and they now have a whole industry devoted to evaluation. So perhaps it is just that we have neglected educational research, and evaluation is merely one of many casualties. Certainly, the idea that *language* teaching is different from *content* teaching has been preached to death by wild curates. Recently, however, L2 researchers have started to build bridges to education (eg Lange's (in press) strongly-worded attack on the solipsism of ESL in the area of teacher development), and as cross-fertilisation occurs, it may be that program evaluation assumes a greater prominence in our literature.

Not that we have ever stopped evaluating language programs. On the contrary, The British Council, along with other major bodies, are forever sending off L2 faculty to far flung locations for that very purpose. But who are they sending?

Our literature treats evaluation as a minor issue, conveniently tacked onto the end of articles and books about curriculum development (eg Richards and Rodgers 1986). When articles do focus on evaluation, they can be out of touch (eg Long's (1984) insistence on true experiment, or the countless product comparisons of immersion studies published mainly in the *Canadian Modern Language Review*).

In education, by contrast, standards have been drawn up for evaluation (Joint Committee 1981), a number of journals have

---

1 Following the convention adopted by Cronbach and Associates (1980), I use the masculine pronoun to refer to the educational evaluator, and the feminine pronoun to refer to other protagonists.

flourished, notably *New Directions for Program Evaluation*, *Evaluation Practice*, and *Educational Evaluation and Policy Analysis*. The American Evaluation Society (AEA) has been formed (there are two members from ESL/Applied Linguistics). Clearly, if our literature and our involvement in professional associations is anything to go by, we probably lack the knowledge that would enable us to offer those who commission evaluations an adequate service, whereas our education colleagues may approach the task with greater confidence.

This state of affairs leads us to a particular evaluation issue which I wish to focus on in this paper: the issue of who should conduct L2 program evaluations. This involves two questions: (1) is it preferable for the evaluator to be an L2 expert or an expert in educational evaluation? and (2) should the evaluator be an insider, part of the program team, or an outsider who owes no allegiance to the program?

### **Educational Evaluator or L2 expert?**

The educational evaluator has expertise in evaluation but may have little or no knowledge of second language pedagogy. The L2 expert has expertise in some areas of second language pedagogy but may have little or no knowledge of evaluation. Who would provide the more useful service? Is it preferable to offer the L2 expert on-the-spot training in evaluation or to provide the evaluator with a crash course in TESL?

The issue would be unimportant if it could be demonstrated that there are many L2 specialists who are well-trained as evaluators. A cursory glance at our literature, as discussed above, argues against this. The issue would also be unimportant if it were possible for expertise to be combined in an evaluation team, but many evaluations are carried out by solo operators. Thus, the issue is real.

An argument favoring the L2 specialist is that project directors and other stakeholders would recognise her credentials. To the evaluator, they may say, as Worthen and Sanders suggest:

I respect your training as an evaluator, but I don't see how you can evaluate my language program adequately since you don't know much about linguistics (1984:2).

The evaluation standards (Joint Committee 1981) support this view, stating that substantive knowledge of a discipline promotes credibility (and credibility promotes use of evaluation findings). Cronbach and Associates (1980) envisage an ideal training program for evaluators which includes disciplinary preparation. Stufflebeam et al (1971) agree but do not see the ideal as practicable.

From another perspective, it might be asked just how tough L2 theory and practice can get. If it is true that the concepts are relatively easy to grasp then there should be little concern that the evaluator will be able to come to terms with them. Worthen and Sanders take a bold line on this as far as educational research is concerned:

Most theories in education are essentially primitive and most educational practices can be easily comprehended if they are clearly described. It is probably neither arrogance or criticism which has led some social scientists to privately aver that they can digest even the most complex educational theories in an afternoon . . . it must be fear of having our theories and practices demeaned as simple-minded ( . . . ) which causes many educational specialists to contend that their work is too complex and filled with subtle nuances to be readily understood by outsiders (1984: 8).

There can be no doubt that many of the phenomena we deal with in L2 theory are exceedingly complex, that while some of the ways we investigate these phenomena may be very complex too, many are relatively simple (which is no criticism) to understand. The point is that for most evaluations, such L2 theory as is necessary for the evaluator can probably be digested.

Furthermore, no matter what program he looks into, the evaluator has many activities to engage in which do not require much familiarity with content. He has to identify evaluation questions, determine relevant sources of information, design the study, collect the data, analyse, interpret and finally report.

Finally, *because* the evaluator is not familiar with L2 theory, he is less likely than the L2 specialist to have attitudes to the principles behind a given program. For the L2 specialist, her familiarity may mean that personal values will intrude and that objectivity will be compromised.

Worthen and Sanders, discussing the relative merits of evaluators and content specialists conclude that:

the person trained as a professional evaluator would be the best choice to evaluate most educational enterprises (1984:28).

Similarly, the arguments that I have presented indicate that in the present circumstances an educational evaluator is probably the appropriate choice for most L2 programs. However, there is a major problem which I have already alluded to and that is the matter of credibility. It is accepted in the evaluation literature that the quality of an evaluation is to be substantially judged by the extent to which its findings are utilised (Cronbach 1982, McLaughlin, Weber, Covert and Ingle 1988, Patton 1988, Weiss 1988, and many others). If it is true that a non-L2 evaluator would lack credibility, then his findings will be ignored. Thus, it would appear that the only hope is for more L2 specialists to take up the challenge of evaluation.

### **Program insider or outsider?**

Irrespective of whether the evaluator is an L2 specialist or not, should the evaluation be carried out by a member of the program team or by an outsider? This is an equally crucial question.

Patton (1988: 91) points out that 'the largest growth among evaluation personnel has probably been internal'. This view is shared by Sonnichsen

(1987) and House (1988), who note that the literature has not yet taken account of this development. Does it make sense to go for this option?

After all, the internal evaluator costs less (Patton 1986). He promotes, as Cronbach and Associates state, the 'close coupling of research consumer to research producer' which 'permits quick correction, whenever the activity is unproductive' (1980: 321), and he is familiar with the ways his institution works (Huberty 1988, Muscatello 1988, Weiss 1972). Thus he can more readily foster utilisation (Beer and Bloomer 1986, Cummings et al 1988, Dickman 1981).

On the other hand, internal evaluators are likely to have less objectivity, independence and credibility than outsiders (House 1986, 1988, Johnston 1988, Nevo 1983, Patton 1988). Surprisingly, this fails to concern some proponents of in-house inquiry. For example, Sonnichsen (1987), the deputy assistant director for the FBI, makes a case for internal evaluation and an abandonment of neutrality. This position is also taken up by Kirkup (1986); evaluating a women's studies curriculum for the Open University (in Britain), she judges that the most appropriate investigator is a feminist evaluator who is ideologically committed.

In spite of his abandonment of neutrality, Sonnichsen believes that research procedures and ethics will ensure objectivity. Cummings et al also argue that 'objectivity, as for any evaluator, rests in the evaluation process used and data gathered' (1988: 65). But, as House points out, it is clear that methods and codes of ethics are not enough to promote objectivity; he speculates that an FBI evaluation of the feminist movement and a feminist evaluation of the FBI could both follow acceptable research procedures and yet be radically different evaluations with quite different appeals to credibility (House 1988: 44). Stressing the threats to independence and objectivity that internal evaluators face, House adds that:

Virtually every organisation deals with sensitive issues that are dangerous for the internal evaluator to touch. For example, feminist evaluators might eventually have to consider the possibility that women compete for jobs with minorities – not a pleasant or popular idea for such evaluators to raise (1988:46).

It is the relative independence of the external evaluator that enhances credibility (Mowbray 1988). For example, Johnston has found that the US General Accounting Office (GAO) has a high rate of acceptance of its recommendations (this is frequently considered a measure of utilisation), and partly attributes this success to its role as an outside agency. He notes that although the evaluation literature sometimes makes a case for the value of the internal evaluator, the organisation literature takes the opposite tack (1988: 79).

Perhaps the greatest loss brought about by a decision to seek internal evaluation is that an insider is so much involved in the institution's way of thinking that there is no fresh perspective. When academics ask colleagues to read their work and comment on it, they do so because they



know they have become so wrapped up in it that they cannot easily achieve the distance necessary for its improvement. Likewise, an internal evaluation seems to suffer from the probability that it will be partial, not only in the sense of being biased, but also in the sense of being incomplete.

## Discussion

The issue of who evaluates L2 programs is, needless to say, far more complex than this brief examination of some of the central concerns has been able to do justice to. Lack of data on personnel, the scope of the work they do, its utilisation, and so on, have necessitated meandering into the speculative byways. Nevertheless, although the questions this paper has addressed merely scratch the surface, they at least serve to focus attention on an area of continuing neglect.

Credibility and competence have emerged as crucial, sometimes competing, areas of interest. At the risk of over-simplification, it appears that credibility is higher for an external person who is also an L2 specialist. However, scruples may be raised about the competence of this same L2 specialist. The competence of the educational evaluator, by contrast, is not seriously in doubt, but I have suggested that his credibility may remain under a cloud.

Perhaps the only person who could satisfy demands of both credibility and competence is the L2 specialist who devotes her energies to external evaluation. If this does not happen, it seems unlikely that L2 program evaluation will prosper. However, if, as appears to be the case, our truancy from educational research proves to have been only temporary, there is reason to expect that it *will* happen.

## References

- Beer, V. and Bloomer, A. C. (1986) 'Levels of evaluation', in *Educational Evaluation and Policy Analysis* 8: 335-345.
- Cronbach, L. J. (1982) *Designing evaluations of educational and social programs*, San Francisco: Jossey-Bass.
- Cronbach, L. J. and Associates (1980) *Toward reform of program evaluation*, San Francisco: Jossey-Bass.
- Cummings, O. W., Nowakowski, J. R., Schwandt, T. A., Eichelberger, T., Kleist, K. C., Larson, C. L., and Knott, T. (1988) 'Business perspectives on internal/external evaluation', *New Directions for Program Evaluation* 39: 59-74.
- Dickman, F. B. (1981) 'Work activities, settings, methodologies, and perceptions: correlates of evaluation research utilization', *Knowledge* 2: 375-387.
- House, E. R. (1986) 'Internal evaluation', *Evaluation Practice* 7: 63-84.
- House, E. R. (1988) 'Evaluating the FBI: a response to Sonnichsen', *Evaluation Practice* 9 (3): 43-46.
- Huberty, C. J. (1988) 'Another perspective on the role of an internal evaluator', *Evaluation Practice* 9 (4): 25-32.
- Johnston, W. P., Jr. (1988) 'Increasing evaluation use: some observations based on the results at the US GAO', *New Directions for Program Evaluation* 39 (3): 75-84.

- Joint Committee on Standards for Educational Evaluation. (1981) *Standards for evaluations of educational programs, projects, and materials*, New York: McGraw-Hill.
- Kirkup, G. (1986) 'The feminist evaluator', in House, E. R. (ed.), *New Directions in education evaluation*, pp. 68-84. Philadelphia: Falmer.
- Lange, D. (in press) 'A blueprint for the design of a teacher development program in second language education', in Richards, J. C. and Nunan, D. (eds.), *Second language teacher education*, Cambridge: Cambridge University Press.
- Long, M. H. (1984) 'Process and product in ESL program evaluation', *TESOL Quarterly* 18: 409-425.
- McLaughlin, J. A., Weber, L. J., Covert, R. W., and Ingle, R. B. (eds.) (1988) 'Evaluation Utilization', *New Directions for Program Evaluation* 39 (3).
- Mowbray, C. T. (1988) 'Getting the system to respond to evaluation findings', *New Directions for Program Evaluation* 39 (3): 47-58.
- Muscattello, D. B. (1988) 'Developing an agenda that works: the right choice at the right time', *New Directions for Program Evaluation* 39 (3): 21-31.
- Nevo, D. (1983). 'The conceptualization of educational evaluation: an analytical review of the literature', *Review of Educational Research* 53: 117-128.
- Patton, M. Q. (1986) *Utilization-focused evaluation* (2nd ed.), Beverly Hills, CA: Sage.
- Patton, M. Q. (1988) 'The evaluator's responsibility for evaluation', *Evaluation Practice* 9 (2): 5-24.
- Richards, J. C. and Rodgers, T. (1986) *Approaches and methods in language teaching: a description and analysis*, New York: Cambridge University Press.
- Scherer, G. A. C. and Wertheimer, M. (1964) *A psycholinguistic experiment in foreign language teaching*, New York: McGraw-Hill.
- Smith, P.D., Jr. (1970) *A comparison of the cognitive and audiolingual approaches to foreign language instruction: the Pennsylvania foreign language project*, Philadelphia, PA: The Center for Curriculum Development, Inc.
- Sonnichsen, R.C. (1987) 'An internal evaluator, responds to Ernest House's views on internal evaluation', *Evaluation Practice* 8 (4): 34-36.
- Stufflebeam, D. L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R. I., Merriman, H.O., and Provus, M. M. (1971). *Educational evaluation and decision-making*, Itasca, IL: Peacock.
- Weiss, C. H. (1972) *Evaluation research: methods for assessing program effectiveness*, Englewood Cliffs, NJ: Prentice-Hall.
- Weiss, C. H. (1988) 'Evaluating for decisions: is anybody there? does anybody care?', *Evaluation Practice* 9 (1): 5-19.
- Worthen, B. R. and Sanders, J. R. (1984) 'Content specialization and educational evaluation: a necessary marriage?', paper #14 in Occasional Paper Series, Evaluation Center, Western Michigan University, (ERIC Document Reproduction Service No. ED 253 543).

# Second Language Classroom Research and Teacher Education

Michael H. Long

*University of Hawaii at Manoa*

Current writing about the preparation of language teachers is like the literature on language teaching pedagogy of fifteen years ago: assertions and prescriptions abound, but there is very little hard evidence about what works. This has serious consequences for those interested in language teacher education (or in language teaching itself) as a profession, since one of several requirements for professional status is that practitioners share a common technical language reflecting specialized knowledge unknown to the lay person and use research information to guide practice (Billups and Rauth, 1987). The lack of a research basis for training is also bad for the consumer, for if the assertions are unfounded and the prescriptions conflict, as is often the case, then at least some trainee teachers and, through them, many more language learners must be in incompetent hands.

This situation is not unique to language teacher education or even to education in general, of course. Glaser (1984, cited in Billups and Rauth, 1987, 637) noted that education as a whole is 'seriously undernourished by modern knowledge' and one of the least research-supported professions. Even medical practice, looked to by some as a potential model for professionalizing language teaching, lacked established standards and procedures until physicians launched their own 'reform movement' in the mid-nineteenth century (Billups and Rauth, *op. cit.*, 624).

Language teaching is in a particularly poor state, however. Johnson (n.d.) observes that while this specialization, unlike some subject areas, has its own academic discipline in applied linguistics, which should contribute to a high level of professionalism, just the opposite appears to be the case in practice. He suggests that the lack of a required common body of knowledge for entry into the field is an important source of the problem:

In other subject areas, professional status demands a solid grounding in the relevant academic disciplines, not because a mathematics teacher for

example will necessarily use that knowledge directly in the classroom, but because effective mathematics teaching needs to be informed by an understanding of the principles upon which mathematics is based. Being 'good at figures' is not enough, and no-one suggests that it is. For language teachers, but particularly ESL teachers, being 'good at the language' is not only 'good enough', it is frequently the sole criterion. (Johnson, n.d., 1)

Johnson goes on to point out that while much language teacher education nominally takes place in postgraduate certificate and masters courses, those programs typically constitute the trainees' first and only exposure to the field, and so often do not involve true postgraduate work at all:

If there are doubts about the level of professionalism in language teaching, the reason may be simple. Far too many language teachers are not in fact professionals in the generally accepted academic sense. Courses in language teaching methods cannot change this situation, which relates to the academic prerequisites for admission to methods courses, and indeed to the language teaching profession as a whole. (Johnson, n.d., 2)

In fact, the situation is far worse than even this gloomy picture suggests, since many teachers, perhaps the vast majority, both native and non-native speakers, are neither 'good at the language' nor the recipients of *any* formal training. Such criticisms imply, however, that things could be improved. More specifically, they assume the existence of a body of knowledge which *ought* to be common to teachers were the field so structured that the mechanism existed for its transmission.

It is my contention that while far from complete, the findings of second language classroom research (SLCR) constitute an important part of such knowledge. The main purpose of SLCR is precisely to inform future classroom practice, and teacher education, both preservice and inservice (along with materials design and learner training) is the principal means available to achieve that goal. If that is agreed, the remaining questions are (1) whether SLCR has yet gathered information worth imparting, ie knowledge ready to be applied, and if so, (2) how best to communicate that information, how best to translate research findings into improved classroom practice.

### **The applicability of SLCR findings**

Few people familiar with the literature would deny that SLCR has discovered a great deal about language learning and teaching, and done so in a relatively short time. Several hundred descriptive studies have been completed (Long, 1985a), along with a small amount of experimental work. (For recent book-length reviews of methodological issues and findings, see Allwright, 1988; Chaudron, 1988; and van Lier, 1988.) Such progress notwithstanding, there are several reasons why it might still be argued that the findings are not yet ready to be passed along to teachers.

First, due to the labor-intensive nature of SLCR and the lack of

external funding for most of it, studies of classroom processes have generally been *small-scale*, involving limited numbers of teachers, learners and classrooms, and have often employed convenience samples. Both characteristics make generalization of findings hazardous at best. Second, for the same reasons and also because the primary aim has been to procure detailed descriptions of teaching and learning, studies have tended to be *short-term*, often cross-sectional, with few attempts to relate classroom behaviours to learning outcomes over time, ie to link process to product. As a result, while a good deal is now known about what goes on in language lessons, little is certain as yet about the long-term effects of potential changes. Third, findings have almost always been *partial*, or fragmented, in the sense that most have focused on only one or two aspects of classroom life, such as turn-taking, question patterns, error correction or group work, and/or on only one dimension of those phenomena, with consequently less than complete understanding of why things are the way they are or of how the different findings are related. Finally, many studies are not just limited in these and other ways, but methodologically *flawed*. For example, reports are published with inadequate data (or no data) on the reliability of analyses (eg on levels of inter-observer agreement when rating scales and coding schemes are utilized), on the validity of those analyses or on the statistical significance of findings. (For a review of some appropriate procedures in the first two areas, see Chaudron, Crookes and Long, 1988.)

Another recurring but less often noted problem with the dissemination of SLCR findings in teacher education is the absence of *theoretical motivation* for the work which produced them, making interpretation and generalization difficult or impossible (Long, 1987a). Showing, for example, that a particular grammatical construction could or could not be taught to a particular group of learners using a particular set of materials and procedures is of virtually no interest, although some journals still publish such results. Before the materials or procedures can be utilized by teachers with other grammatical items and/or learners, it is necessary to know what it was about the original item, learners, materials, procedures or combinations thereof that produced the reported outcome. To answer that question means knowing (or thinking one knows) which *classes* of grammatical items, learners, materials and procedures were being sampled. That amounts, in turn, to claiming to *understand* the results (or thinking one does), ie to having an interim *explanation* for them, or a *theory*. Needless to say, the theory may eventually turn out to be wrong, but until that is shown, it offers a principled means of applying the original findings to new situations.

The alternative, no theory, means that research is unproductive. Results are accumulated at random, with no way of knowing, for example, with which other structures or learners instruction should be successful and with which it should not. To be generalizable, as the term suggests, a finding needs to be framed as some kind of generalization, eg 'Learners at stage A (but not stage B) of interlanguage development can

be taught structures of type X (but not type Y)', where A, B, X and Y are defined in terms of some understood categories, eg processing constraints, stages in interlanguage development and types of linguistic complexity or markedness.

To take another example, it is of little use to know that pedagogic task A serves better than pedagogic task B for a particular kind of language work. To be useful, it is necessary to be able to predict which other tasks will and will not have the same function. In other words, we need to know again, which classes of task, or *task-types* (one-way/two-way, open/closed, planned/unplanned, etc.) tasks A and B supposedly represent. It is unhelpful to tell teachers that, say, 'Buried Treasure' on page 39 of a specific textbook elicits less topic-recycling or interlanguage destabilization than 'Spot the Difference' on page 42. Rather, teachers need such generalizations as 'Closed two-way tasks produce more topic-recycling than open two-way tasks' and 'Tasks performed after planning elicit more target-like use and more syntactically complex language than the same tasks without prior planning.' (The categories also need to be self-explanatory or easily defined, or course.)

The problem is that to substantiate statements (claims) of this kind, it is insufficient simply to describe what happened when particular structures were taught or particular pedagogic tasks were used, as so many studies have done. It is necessary to theorize beforehand about (in these cases) the principles governing teachability and learnability (see eg Pienemann, 1984) or the relationships pertaining between task-types and language performance (see, eg Long, 1989) and then to test the ideas by manipulating different values of those variables hypothesized to be important in a controlled experiment. Given, as noted earlier, that most SLCR to date has been descriptive, *not* experimental, another potential problem in applying classroom research findings in language teacher education becomes apparent.

These concerns are legitimate, and researchers need to address them if their work is to receive more serious attention by teachers and teacher educators alike. They do not, however, invalidate the use of existing SLCR findings in teacher education. There are several reasons why this is so.

First, while it is true that most individual studies have been small-scale, several topics have now received sufficient attention to warrant cautious but defensible generalizations across studies. For example, there have been a number of studies of teacher feedback on learner error, the pooled findings of which allow reviewers to provide useful information to teacher educators about (1) what forms teacher error 'corrections' typically take (and about students' preferences in this area), (2) when to correct, (3) which errors to correct, (4) how to correct, and (5) who, teacher or student, should do the correcting. (For review, see Chaudron, 1988, p. 132–153.) Similarly, the existence of over a dozen studies of group work permitted reviewers to formulate generalizations about the quantity, variety and accuracy of unsupervised language

practice in group work compared with teacher-led lockstep work, as well as about students' ability to correct each other (and the effects of learner training in this regard) and to resist incorporating errors by other group members, about negotiation work and (very tentatively) about some relationships between group work and language practice. (For review, see Long and Porter, 1985. Additional key studies have since been reported on some of these issues.) Further examples of topics on which a series of small-scale studies have now been accumulated can be found in the recent book-length reviews of SLCR findings cited earlier. In each case, while few findings were alone strong enough to support firm conclusions, generalizations can be drawn on the basis of secondary analyses of series of studies. Literature reviews (of which many more are still urgently needed) incidentally make for manageable amounts of reading for trainee teachers who usually have neither the time nor the need to go through all the original studies.

The short-term nature of most studies, the second potential argument against utilizing SLCR findings in teacher education at this stage, does indeed mean that long-term learning outcomes are largely unknown. Again, however, that does not make what has already been established irrelevant. First, detailed descriptions of what really goes on in classrooms are in most respects more informative than data-free prescriptions about what methods writers think should go on. Also, while it is true that few process-product links have yet been established, findings are becoming available about the immediate, short-term effects of process changes on other processes, themselves believed to be beneficial to language learning. Certain types of teacher questions, for example, appear to increase the average length and syntactic complexity of student utterances, the percentage of communicative language use students engage in, and (tentatively) students' post-test scores on mastery of lesson content (Brock, 1986; Long and Crookes, 1987; Tollefson, 1988). Certain types of modifications of spoken and written discourse improve student listening and reading comprehension as much as traditional linguistic 'simplification' of the same texts (as in the basal reader 'See Spot run' approach), but without denying learners exposure to the lexical and morphosyntactic items they currently have difficulty with (eg Brown, 1987; Long, 1985b; Pica, Doughty and Young, 1986; and for review, Parker and Chaudron, 1987). The SLCR literature yields many other examples.

The third problem mentioned, the partial, or fragmentary, nature of the available descriptions, may not be a problem at all. It is obvious that research which focuses exclusively on describing (say) teacher feedback on learner error, group work, teacher questions or relationships between task-types and interlanguage use may not be able to *explain* the findings or to show how they relate to findings on another topic. (Such is not the intent of narrowly focused descriptive studies, after all.) This limitation is only a weakness, however, if it can be shown that the findings are invalid or misleading *because* they are focused and 'isolating' — a logical

possibility, of course, but one for which, to my knowledge, no evidence exists as yet. In other words, there is an important difference between claiming that findings are a methodological artefact and suspecting that they are limited (but valid) because of the methodology employed and/or the scope of a study.

Far from appearing invalid or misleading, as noted earlier, there is increasing evidence that at least some of the processes (uses of certain types of questions, tasks, organizational groupings, etc.) are responsible for interesting variance in other phenomena believed (by some) to be important for classroom language learning, eg the quantity, complexity and accuracy of student production. No claim is made, of course, that these early studies of limited dimensions of single processes do justice to the complexity of classroom language learning and use. However, it would be as absurd to reject what appear to be quite useful, if limited, findings just because they are only part of the picture. In addition, it turns out that, partial though they may be, such phenomena are closer to the units of analysis (activities or tasks) that studies have repeatedly found teachers employ when they plan, implement and recall lessons than the vague, more global pronouncements about 'methods' so popular among arm-chair prescriptivists.

Needless to say, it would be pointless to try to defend methodologically-flawed studies. It should be noted, however, that not all flaws invalidate all findings, and that by no means all studies have been flawed. The use of convenience samples, similarly, limits the generalizability of findings, but does not make studies employing such samples useless. Finally, while the lack of theoretical motivation for studies greatly reduces the value of their findings in various ways, it does not necessarily rule out *post hoc* interpretation in the light of other findings and theories, cautious though that interpretation must be.

In sum, SLCR does not have all the answers, nor pretends to. It does offer a growing body of tangible evidence about language teaching classroom practice, however, and is even beginning to produce explanations of some aspects of teaching and learning. Hard information about the activity one is training to perform must not only be useful, but more useful than other people's prejudices and suppositions about that activity. Even when current research information gives out, as it currently quickly does, alerting teachers to the extent of professional ignorance is both honest and potentially valuable, if only because it helps protect them from 'experts' with well-packaged legislation about 'good teaching' and 'what we know works'.

### **The communication of SLCR findings**

Having concluded that SLCR has gathered information worth imparting, it remains to establish how best to communicate it in a teacher education program. We know very little about that, so I will make some brief observations and suggest we get on with some research on the topic, with



microteaching and training in systematic classroom observation appearing to be promising candidates for researchers' attention.

A survey of syllabuses from language teaching methodology courses in US and Canadian masters programs in ESL (Long, 1983) suggested that most programs were giving very little coverage to SLCR findings in the early 1980s, at least where explicit discussion topics and assigned readings were concerned. There was a serious shortage of pedagogically accessible reviews of such work at the time, however, a situation that still prevails but which, as noted earlier, has since been much ameliorated by the appearance of a number of journal articles and three books on the subject. It is to be hoped that at least some of this literature now figures among the required reading in such courses, but given what is known about curriculum change, it is safe to assume that simply reading about classroom processes will often fail to translate into change in classrooms (for review, see Nunan, 1988). Consequently, innovations in communicating SLCR findings in teacher education are needed, plus *controlled evaluations of their effectiveness*.

Combined with a traditional diet of guided reading and discussion, two ideas look promising: microteaching and training in systematic classroom observation. Neither is original, which may be an indication that they have some merit and means they can be briefly stated here. They are often used in combination.

Microteaching has a long history in teacher education of all sorts and has often been the focus of research (for review, see, eg McIntyre, MacLeod and Griffiths, 1977; Wragg, 1983). There has also been considerable interest in its application to language teacher training (see, eg Carver and Wallace, 1981; Wallace, 1981, 1982), but virtually no formal evaluation of its effectiveness there. This is unfortunate since, while the unmarked hypothesis would presumably be that a procedure that has been shown to be effective in one kind of teacher education will work in another, this may be overly optimistic. When the medium of instruction is also the object of instruction, it may turn out that the traditional formulae used in microteaching to isolate, break down and reconstitute teaching practices will need modification. Depending on their methodological persuasion, some language teachers, after all, *may* be as much concerned with the form as with the function of 'clarifications of explanations', 'feedback moves', 'higher order questions' and so on. On the other hand, the 'local' nature of many phenomena described in SLCR, such as question patterns and feedback on error, should help make them amenable to systematic change through microteaching. Thus, at least three studies (Brock, 1986; Long *et al*, 1984; and Tollefson, 1988) have found that even very simple in-service training modules can be effective in producing substantial changes in some aspects of native and non-native speaking teachers' questioning patterns (and thereby, in the quality of student responses to them), as assessed by analyses of their subsequent classroom teaching.

As with microteaching, regrettably little work has been done in

language teacher education to evaluate the effectiveness of the second potential procedure for disseminating SLCR findings: training in systematic classroom observation. Once again, there is some reason for optimism, however. First, observation systems developed specifically for SL classrooms are now available, eg Foci for Observing Communications Used in Settings, or FOCUS (Fanselow, 1977) and the Communicative Orientation of Language Teaching, or COLT (Allen, Frohlich and Spada, 1984). Second, research with content teachers has long shown that training in the use of such systems designed for coding talk in their classrooms can be effective in changing subsequent teaching behaviour (see eg papers in Amidon and Hough, 1967). Third, smaller units of analysis like those typically utilized in SLCR have generally been found to be more susceptible to systematic change than global constructs like 'method' or 'communicative language teaching' (Amidon and Hough, 1967; Long, 1987b; Nunan, 1987, 1988). Fourth, interesting proposals for step-by-step procedures using such systems with language teachers have begun to be put forward (eg Ramani, 1987). Finally, if use of complete observation systems seems overly complex, the SLCR literature offers numerous well-trying and conceptually simple categories, quite small subsets of which can be used by teachers to monitor innovations they make in materials and in different aspects of their own and colleagues' teaching (eg one-way and two-way tasks, general and personal solicits and open and closed referential and display questions). There are also a few studies documenting the effectiveness of their implementation in a variety of programs (eg Tollefson, 1988; Nunan, 1988).

To conclude, there are some grounds for optimism in these areas and regarding the undeniable progress made in SLCR. However, it would be misleading to suggest anything other than that while SLCR findings show great promise for application in language teacher education, there is a shameful lack of research on the effectiveness of alternative dissemination mechanisms. Accountability and professional status for language teachers and language teacher educators will both require that to change.

## References

- Allen, J. P. B., Frohlich, M. and Spada, N. (1984) 'The communicative orientation of language teaching: an observation scheme' in Handscombe, J. Orem, R. A. and Taylor, B. (eds.), *On TESOL '83: The Question of Control*: 231-252. Washington, D.C.: TESOL.
- Allwright, R. L. (1988) *Observation in the Language Classroom*, New York: Longman.
- Amidon, E. J. and Hough, J. B. (eds.) (1967) *Interaction Analysis: Theory, Research, and Application*, Reading, Mass.: Addison-Wesley.
- Billups, L. H. and Rauth, M. (1987) 'Teachers and research' in Richardson-Koehler, V. (ed.) *Educators' Handbook: A Research Perspective*, 624-639, New York: Longman.
- Brock, C. (1986) 'The effects of referential questions on ESL classroom discourse', *TESOL Quarterly* 20, 1: 47-59.
- Brown, R. (1987) 'A comparison of the comprehensibility of modified and unmodified reading materials for ESL', *University of Hawaii Working Papers in ESL* 6, 1: 49-79.

- Carver, D. and Wallace, M. J. (eds.) (1981) *SCEO Microteaching Papers*, Edinburgh: Moray House College.
- Chaudron, C. (1988) *Second Language Classrooms. Research on Teaching and Learning*, Cambridge: Cambridge University Press.
- Chaudron, C., Crookes, G. and Long, M. H. (1988) *Reliability and validity in second language classroom research, Technical Report No. 8*, Honolulu: Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii at Manoa.
- Fanselow, J. F. (1977) 'Beyond "Rashomon" – conceptualizing and describing the teaching act', *TESOL Quarterly* 11, 1: 17-39.
- Glazer, R. (1984) 'Improving education: perspectives on educational research' in National Academy of Education (ed.), *The Necessity for Research in Education: An Introduction*, 4-8. Washington, D.C.: National Academy of Education.
- Johnson, R. K. (n.d) 'ESL teacher-training: the case for the prosecution', Ms. Faculty of Education, Hong Kong University.
- Long, M. H. (1983) 'Training the second language teacher as classroom researcher' in Alatis, J. E., Stern, H. H. and Strevens, P. (eds.), *GURT '83 Applied linguistics and the preparation of second language teachers: Toward a rationale*: 281-297, Washington, D.C.: Georgetown University Press.
- Long, M. H. (1985a) *Bibliography of research on second language classroom processes and classroom second language acquisition, Technical Report No. 2*, Honolulu: Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii at Manoa.
- Long, M. H. (1985b) 'Input and second language acquisition theory' in Gass, S. M. and Madden, C. G. (eds.), *Input in Second Language Acquisition*, 377-393, Rowley, Mass.: Newbury House.
- Long, M. H. (1987a) 'The experimental classroom', *Annals of the American Journal of Political and Social Science* 490: 97-109.
- Long, M. H. (1987b) "'Focus on form": a design feature in language teaching methodology', paper presented at the National Foreign Language Center/European Cultural Foundation Conference on Empirical Research on Second Language Learning in Institutional Settings, Bellagio, Italy: Rockefeller Centre, June 20-24.
- Long, M. H. (1989) 'Task, group and task-group interactions', paper presented at the RELC seminar, Language Teaching Methodology for the Nineties, Singapore: RELC.
- Long, M. H., Brock, C., Crookes, G. Deike, C. Potter, L. and Zhang, S. (1984) *The effect of teachers' questioning patterns and wait-time on pupil participation in public high school classes in Hawaii for students of limited English proficiency, Technical Report No. 1*, Honolulu: Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii at Manoa.
- Long, M. H. and Crookes, G. (1987). 'Intervention points in second language classroom processes' in Das, B. K. (ed.), *Patterns of Classroom Interaction in Southeast Asia*, 177-203.
- Long, M. H. and Porter, P. A. (1985) 'Group work, interlanguage talk, and second language acquisition', *TESOL Quarterly* 19, 2: 207-228.
- Nunan, D. (1987) 'Communicative language teaching: making it work', *English Language Teaching Journal* 41, 2: 136-185.
- Nunan, D. (1988) *The Learner-Centered Curriculum*, Cambridge: Cambridge University Press.
- Parker, K. and Chaudron, C. (1987) 'The effects of linguistic simplification and elaborative modifications on L2 comprehension', *University of Hawaii Working Papers in ESL* 6, 2: 107-133.
- Pica, T., Doughty, C. and Young, R. (1986) 'Making input comprehensible: do interactional modifications help?', *ITL Review of Applied Linguistics* 72: 1-25.
- Pienemann, M. P. (1984) 'Psychological constraints on the teachability of languages', *Studies in Second Language Acquisition* 6, 2: 186-214.
- Ramani, E. (1987) 'Theorizing from the classroom', *ELT Journal* 41, 1: 3-11.

- Tollefson, J. W. (1988) 'Measuring communication in ESL/EFL classes, *Cross Currents* 15, 1: 37-46.
- van Lier, L. (1988) *The Classroom and the Language Learner*, New York: Longman.
- Wallace, M. J. (1981) 'Toward a skills-based analysis of EFL teaching skills', *TESOL Quarterly* 15, 2: 151-157.
- Wallace, M. J. (1982) *Micro-Teaching and the Teaching of English as a Second or Foreign Language, Year Three: 1980-1981*, Edinburgh: Moray House College.
- Wragg, E. C. (1983) *A Review of Research in Teacher Education*, Windsor, Berks.: NFER-Nelson.

# Notes on Contributors

*Lyn Banbrook* has taught in Australia, Indonesia, Greece and the UK. An Australian, she has a Master's degree from London University Institute of Education.

*Alan Beretta* is an assistant professor at Michigan State University. He has published several articles on program evaluation in *Language Learning*, *TESOL Quarterly*, and *Applied Linguistics*.

*Christopher Brumfit* is Professor of Education with reference to Language and Linguistics, and Head of the School of Education at the University of Southampton. He is also Director of the Centre for Language in Education, and editor of the *ELT Documents* series.

*Vivian Cook* lectures in applied linguistics at the University of Essex. He is the author of many textbooks and research papers. His most recent book is *Chomsky's Universal Grammar* (Blackwells).

*Rod Ellis* was Head of the Department of English Language Teaching at Ealing College of Higher Education and is now at Temple University, Tokyo. He worked as a teacher and teacher trainer in Zambia for ten years. He has also worked in Morocco, Thailand, Indonesia, India, Cameroon, Venezuela, Sri Lanka, Tanzania, Japan and Canada. His main publications are in the fields of second language acquisition and teacher training.

*Peter Green* was born in 1929, studied in Cambridge and Lund, Sweden, and taught German and French in British schools. From 1965 he was Lecturer, then Senior Lecturer at the Language Teaching Centre, University of York (Director 1979-84). He has been responsible for research and publications in the effectiveness of language laboratories, aptitude testing, and interlanguage analysis. He is also a textbook author.

*Karlheinz Hecht* was born in 1928, studied in Munich and the State University of Iowa, and taught English in German schools. From 1964-74 he was lecturer at Munich College of Education. From 1975 he held the Chair of English Language and Literature Teaching at the University of Munich. He has carried out research in error analysis, performance analysis of written/oral communication, and longitudinal study of interlanguage. He has also produced publications in methodology and syllabus planning. He is a textbook author and contributor to educational TV and radio.

*Christina Howell-Richardson* has been a tutor in the Institute of Applied Language Studies, University of Edinburgh, since 1986. She is concerned with phonology and with writing skills for advanced learners, as well as learner diaries and learner strategies. She previously taught English as a first and foreign language in the UK and Sweden.

*Michael H. Long* teaches on the MA in ESL and Ph.D. in SLA programs in the Department of ESL at the University of Hawaii at Manoa, where he specialises in second language acquisition, research methods, classroom research and language teaching methodology. He is co-editor of the *Cambridge Applied Linguistics Series*.

*Tony Lynch* has taught EFL in a number of European countries and in Brazil. Since 1975 he has worked principally in English for Academic Purposes, applied linguistics and teacher education. Currently a lecturer at the Institute for Applied Language Studies, Edinburgh University, his main areas of interest are native/non-native interaction, self-directed learning and listening comprehension. He is the author of *Study Listening*

(Cambridge University Press, 1983) and the co-author, with Anne Anderson, of *Listening* (Oxford University Press, 1988).

*Rosamond Mitchell* is a lecturer in the School of Education, University of Southampton. For ten years she was a researcher, mainly into language teaching and bilingual education, at the University of Stirling, before moving to the Linguistics Department, University of Edinburgh from 1984-6. She is co-editor of *Applied Linguistics* from January 1990.

*David Nunan* was Director, National Curriculum Resource Centre for the Australian Adult Migrant Education Program, and now works at Macquarie University. He has an M. Ed. in curriculum design and a Ph.D. in applied linguistics. He has worked as a teacher, researcher and consultant in Australia, Singapore, Thailand, the Middle East and the UK and has published books on discourse analysis, teacher education, curriculum development and syllabus design.

*Brian Parkinson* is a lecturer at the Institute for Applied Language Studies, University of Edinburgh, where he has worked since 1982. He is concerned with both EFL and other Modern Languages, and with teacher education. His special interests are classroom processes, criterion-referenced testing, text analysis and translation and language for literature, as well as learner diaries and variables affecting learning. He was previously a schoolteacher in England and Germany and Research Fellow in the Department of Education, University of Stirling.

*Peter Skehan* is a lecturer in the ESOL Department at the Institute of Education, University of London. With Paul Meara, he has been editor of the series 'Second Language Acquisition', in which he published *Individual Differences in Second-Language Learning* in 1989.

*Nina Spada* is an assistant professor in the Department of Education in Second Languages at McGill University. She teaches graduate and undergraduate courses in second language acquisition, language teaching methodology and classroom-centred research. She is currently co-directing (with Patsy Lightbown) a research project investigating relationships between teaching and learning in experimental intensive ESL programs for children in Montreal, Quebec.

*Leo van Lier* obtained a Ph.D. in Linguistics from the University of Lancaster in 1982. From 1982 to 1984 he worked on a Bilingual Education Project in the Peruvian Andes. He has taught at the University of Northern Iowa, and currently teaches at the Monterey Institute of International Studies. He has written extensively on classroom research.

English Language Teaching Documents  
General Editor: C J Brumfit

## Research in the Language Classroom

A survey of what research can and cannot do for teachers, together with information about how to start examining your own and others' classroom practices. Among the topics covered are the roles of systematic and unsystematic observation; research into learner behaviour, teacher behaviour, classroom discourse, and programme evaluation; relationships

between research and teacher education, teaching theory, and classroom practice; how to write up and disseminate research. The research described always relates directly to advice on how to approach practical problems, and the contributors are experienced researchers and teachers from many parts of the world.

