TABLE OF CONTENTS

TESOL QUARTERLY

Volume 25, Number 4 □ Winter 1991

A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect

Credits

Advertising arranged by Patti Olson, TESOL Central Office, Alexandria, Virginia
Typesetting, printing, and binding by Pantagraph Printing, Bloomington, Illinois
Design by Chuck Thayer Advertising, San Francisco, California

Copies of articles that appear in the *TESOL Quarterly* are available through *The Genuine Article*®, 3501 Market Street, Philadelphia, Pennsylvania 19104 U.S.A.

US ISSN 0039-8322

TESOL QUARTERLY

CONTENTS

To print, select PDF page nos. in parentheses

ARTICLES

REVIEWS

BOOK NOTICES

BRIEF REPORTS AND SUMMARIES

THE FORUM

TESOL is an international professional organization for those concerned with the teaching of English as a second or foreign language and of standard English as a second dialect. Benefits of membership include four issues of the *TESOL Quarterly* and six issues of *TESOL Matters* a year, plus a variety of other services and opportunities for professional development. Each year TESOL holds a Convention in the spring and a Summer Institute. TESOL also publishes and distributes a wide range of professional resource materials, which are available to members at reduced rates. Membership in TESOL makes one eligible to subscribe to *TESOL Journal* (four issues a year). A TESOL Membership Application is printed in every issue of the *Quarterly*. For further information, contact the TESOL Central Office, Suite 300, 1600 Cameron Street, Alexandria, Virginia 22314-2751 U.S.A., (703) 836-0774.

TESOL QUARTERLY is published in Spring, Summer, Autumn, and Winter. Publisher's representative is Helen Kornblum, Director of Communications & Marketing. *Subscriptions are not sold without membership.* Communications regarding back issues, single copies, and permission to reprint should be addressed to the TESOL Central Office, Suite 300, 1600 Cameron Street, Alexandria, Virginia 22314-2751 U.S.A., (703) 836-0774. Contributions should be sent to the Editor or the appropriate Section Editors at the addresses listed in the Information for Contributors, which is included in each issue of the *Quarterly*.

TESOL Matters is published in February, April, June, August, October, and December. *TESOL Matters* contains organizational news and announcements, Affiliate and Interest Section news and information, book reviews, conference information and reports, short refereed articles, and general information. *TESOL Matters is available only through membership in TESOL or by subscription through a regional affiliate of TESOL.* All contributions should be directed to Helen Kornblum, TESOL Central Office, Suite 300, 1600 Cameron Street, Alexandria, Virginia 22314-2751 U.S.A., (703) 836-0774.

ANNOUNCEMENTS should be sent directly to the Editor, *TESOL Matters*, two months prior to the month of publication and must be received by the first of that month (e.g., February 1 for the April issue). *The TESOL Quarterly does not publish announcements.*

ADVERTISING in all TESOL publications is arranged by Patti Olson, TESOL Central Office, Suite 300, 1600 Cameron Street, Alexandria, Virginia 22314-2751 U.S.A., (703) 836-0774.

OFFICERS AND EXECUTIVE BOARD 1991-92

TESOL QUARTERLY

Editor's Note

■ This final issue in the *TESOL Quarterly's* 25th-anniversary volume combines state-of-the-art discussions with reports of exploratory studies. This combination of retrospection and exploration is a fitting note on which to begin our second quarter century.

In this Issue

■ Articles and reviews in this issue of the *TESOL Quarterly* explore different ways of approaching and researching the teaching process. We begin with an introduction to the informed reading of statistical studies, followed by a study to which the first article is keyed. The next article takes a turn in the debate regarding whether to teach grammar, exploring a communicative task-based approach to grammar instruction. Articles on vocabulary acquisition present a case study and statistical approach respectively. Vocabulary acquisition is also the topic of the books noted and reviewed in this issue. We end with a state-of-the-art discussion of language testing.

- James Dean Brown's article, addressed to ESOL teachers who currently avoid statistical reports, explores "statistics as a foreign language." Drawing examples from the second article in this issue of the *TESOL Quarterly*, Brown outlines strategies that can help the noninitiate gain access to statistical studies.
- In the second article, James Dean Brown presents the results of a statistical study investigating whether differences exist in the writing scores assigned by ESL and English faculties to compositions written by native-speaker and international students. Raters were also asked to choose the best and worst features of each composition. Results showed no significant differences between ratings given; however, the

features analysis suggested that "the ESL and English faculties may have arrived at their scores from somewhat different perspectives."

- Sandra Fotos and Rod Ellis report on an exploratory trial of a communicative grammar-based task for promoting L2 linguistic knowledge. They argue that "providing learners with grammar problems they must solve interactively integrates grammar instruction with opportunities for meaningful communication." The results of their study suggest that the grammar task both increased knowledge of a difficult L2 grammar rule and encouraged the kind of communicative interaction widely assumed to facilitate language acquisition.
- Kate Parry presents the results of a series of longitudinal case studies exploring how language learners build their academic vocabularies. Parry concludes that a range of strategies may be used, "each involving liabilities as well as assets." Language teachers can make students aware of this range, helping them develop flexibility in their responses to unfamiliar words. Parry ends with a call to researchers for more attention to the problem of building a theory of vocabulary acquisition.
- Thomas Brown and Fred Perry, Jr. report the results of a statistical study comparing three vocabulary learning strategies for Arabic-speaking ESL students. Of the three treatment groups—keyword, semantic, and keyword-semantic—the combined strategy seemed to increase retention. Brown and Perry also call for more research, "in order to gain an overall picture of the optimal use of learning strategies for vocabulary learning."
- This volume ends with a commentary on the state-of-the-art in language testing: Lyle Bachman surveys advances in the field over the past decade, then describes a promising current approach. The interfactional model of language test performance described includes two components: language ability and test method. Two aspects of authenticity are discussed: situational and interfactional. The paper outlines practical implications of this framework for language testing.

Also in this issue:

- Reviews: Marianne Celce-Murcia and James Cody each review I. S. P. Nation's *Teaching and Learning Vocabulary;* Terry Santos reviews *Second Language Writing: Research Insights for the Classroom,* a volume edited by Barbara Kroll.
- Book Notices The role of vocabulary in language teaching is the focus of this issue's Book Notices section, for which Marianne Celce-Murcia has been guest editor.
- Brief Reports and Summaries: Frances Boyd reassesses the use of the case method in English for specific purposes for business; Mark Patkowski reports the results of a study of the predictive value of a university basic skills test; Jerry Gebhard explores issues of teacher supervision in the teaching practicum.

- The Forum: Martha McCall's commentary on Alastair Pennycook's recent *TESOL Quarterly* article, "The Concept of Method, Interested Knowledge, and the Politics of Language Teaching," is followed by a response by the author; Sandra McKay and Sarah Freedman respond to comments by Marjorie Hoye on their recent *TESOL Quarterly* article, "Language Minority Education in Great Britain: A Challenge to Current U.S. Policy." In the subsection Research Issues, Anne Lazaraton and Graham Crookes comment on power and effect size in second language research.

Sandra Silberstein

TESOL Quarterly

Spring 1992 Highlights

Articles:

Task-Based Syllabus Design: Michael H. Long and Graham Crookes on three approaches

"Spinach to Chocolate": Lise Winer on changing awareness and attitudes in ESL writing teachers

Variation and Universality in Communicative Competence: Philip Shaw on Coseriu's model

Reviews:

Thomas Ricento on ***Perspectives on Official English***

Macey Taylor and Teresa Talia on ***Breaking the Language Barrier***

The Forum:

Ulla Connor, Elizabeth Closs Taugott, Patricia Dunkel, and Patricia L. Carrell comment on Ann Raimes's "The TOEFL Test of Written English: Causes for Concern"; the author responds.

Erratum

The following references were omitted from Ann Raimes's article, "Out of the Woods: Emerging Traditions in the Teaching of Writing," in the Autumn1991 issue (Vol. 25, No. 3) of the *TESOL Quarterly*.

Tsao, Fen-Fu. (1983). Linguistics and written discourse in English and Mandarin. In R. B. Kaplan (Ed.), *Annual review of applied linguistics 1982* (pp. 99-117). Rowley, MA: Newbury House.

Weissberg, R. C. (1984). Given and new: Paragraph development models from scientific English. *TESOL Quarterly, 18* (3), 485-499.

Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Zamel, V. (1976). Teaching composition in the ESL classroom: What we can learn from research in the teaching of English. *TESOL Quarterly, 10* (1), 67-76.

Zamel, V. (1980). Re-evaluating sentence-combining practice. *TESOL Quarterly, 14* (1), 81-90.

Zamel, V. (1982). Writing: The process of discovering meaning. *TESOL Quarterly, 16* (2), 195-210.

Zamel, V. (1963). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly, 17* (2), 165-187.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19* (1), 79-101.

TESOL QUARTERLY, Vol. 25, No. 4, Winter 1991

Statistics as a Foreign Language— Part 1: What to Look for in Reading Statistical Language Studies *

JAMES DEAN BROWN
University of Hawaii at Manoa

This article is addressed to those practicing EFL/ESL teachers who currently avoid statistical studies. In particular, it is designed to provide teachers with strategies that can help them gain access to statistical studies on language learning and teaching so that they can use the information found in such articles to better serve their students. To that end, five attack strategies are advocated and discussed: (a) use the abstract to decide if the study has value for you; (b) let the conventional organization of the paper help you; (c) examine the statistical reasoning involved in the study; (d) evaluate what you have read in relation to your professional experience; and (e) learn more about statistics and research design. Each of these strategies is discussed, and examples are drawn from the article following this one in this issue of the *TESOL Quarterly*.

The *TESOL Quarterly* is currently the research journal of the organization, Teachers of English to Speakers of Other Languages. Ironically, many of the statistical studies on language learning and teaching that are found in the *Quarterly* may be incomprehensible to the very EFL/ESL teachers who makeup the intended audience. Rather than bemoaning this situation (either by berating teachers for not knowing more about statistics or by criticizing researchers for producing articles that are frequently not accessible to teachers), this article will begin by accepting statistical language studies for what they are: legitimate investigations into phenomena in human language learning/teaching which include the use and systematic manipulation of numbers as part of their argument.

Notice that I purposely avoid terms such as *empirical* and *experimental* in referring to these statistical language studies. I am doing

* Part 2, scheduled to appear in Volume 26, discusses more advanced statistical procedures.

so for several reasons. First, there are other, nonstatistical studies that could be called empirical (e.g., ethnographies, case studies, etc.) since, by definition, empirical studies are those based on data (but not necessarily quantitative data). Second, there are statistical studies that are not exactly experimental in the technical sense of that word (e.g., quasi-experimental studies, posttest-only designs, etc.). Third, there are statistical studies that have little or nothing to do with experimentation (e.g., demonstrations, survey research, etc.).

Regardless of what studies are called, when confronted with statistics, many readers will either skip an article entirely, or take a rather cursory route through the paper. Such a route might include skimming the abstract and the Introduction section, then skipping over the Method and Results sections (with their tables, figures, and statistics) to the Conclusions (and/or Discussion) section where they look to find out what the study was all about. If this strategy sounds similar to one that you use, you may be missing an opportunity. Statistical reasoning is just a form of argumentation; by skipping the Method and Results sections, readers not only miss the heart of the study, but also buy the authors' argument without critical evaluation. Most of us would not surrender so easily if the form of argument were expressed in words rather than numbers. We would read a prose article carefully and critically. We should not have to surrender professional skepticism just because the form of argument may be a bit alien, that is, numerical.

The purpose of this article is to provide some attack strategies for teachers to use in gaining access to statistical studies and in understanding them better. In the process, examples will be drawn from a study reported in the second article in this issue of the *TESOL Quarterly*.

ATTACK STRATEGIES FOR STATISTICAL STUDIES

With the following strategies in hand, you may not understand every word of a statistical study, but you will be able to gain access to such studies and will have a purposeful way of grappling with the content.

Use the Abstract to Decide if the Study has Value for You

Let us begin with the familiar and work toward the less familiar. The portion of a statistical report that is probably most often read is the abstract. An abstract typically contains about 150 words in the

TESOL Quarterly. Other journals may have somewhat longer or shorter abstracts. Regardless of their length, these handy summaries should contain enough information for the reader to know what the study was about, how it was conducted, and the general trend of the results. In other words, an abstract should tell the reader in a nutshell what is presented in the study and allow you to determine if an article is pertinent enough to your professional life and teaching situation to be interesting and worthy of your time.

Indeed, there is an overwhelming and increasing amount of information competing for our professional attention. Along with the *TESOL Journal* and the *TESOL Quarterly*, EFL/ESL teachers choose among other journals, such as the *ELT Journal*, the *ESP Journal*, *Language*, *Language Learning*, *Language Testing*, *Modern Language Journal*, *Studies in Second Language Acquisition*, and *TESL Canada Journal/Revue TESL du Canada*, to name just a few in the United Kingdom and North America, as well as *Cross Currents* and the *JALT Journal* (Japan), *Prospect* (Australia), *RELC Journal* (Singapore), *System* (Sweden), and many others.

Because of this plethora of journals, it is essential to use the abstracts to advantage. Consider the abstract associated with the example article that follows this one. Is there sufficient information in that abstract for you to decide whether the article is of interest to you?

Let the Conventional Organization of the Paper Help You

The *TESOL Quarterly* and many other journals in our field generally follow the format and organization described in the *Publication Manual of the American Psychological Association* (APA) (American Psychological Association, 1983). That manual advocates using the following general sections and subsections in reporting a statistical study:

- Introduction
  - Introduction to the problem
  - Background
  - Statement of purpose
- Method
  - Subjects
  - Materials (or Apparatus)
  - Procedures
- Results
- Discussion (and/or Conclusions)
- References

Typically in our journals, there are no headings for the Introduction or its subparts. However, beginning with the Method section (including any Subjects, Materials or Apparatus, and Procedure subsections) through the Results, Discussion, Conclusion, and References sections, you will generally find clear headings and subheadings. Since the general purpose of the headings and subheadings is "to help readers find specific information" (American Psychological Association, 1983, pp. 25-26), you should use them to help you find and organize the information that you need in order to understand the study. There is not space here to provide details about what each of these sections should contain. Indeed, such details are not necessary here because existing sources give ample information on this topic (e.g., Brown, 1988; pp. 43-62; Hatch & Farhady, 1982, pp. 33-38; Hatch & Lazaraton, 1991, pp. 107-126).

Nevertheless, there are a number of questions that you might want to ask yourself as you read through a statistical study. These questions should help focus the information contained in these key sections and help readers critically evaluate a study. Notice that section and subsection headings are listed in parentheses after each of the questions below. These are meant to suggest where you would typically find the information that would answer each question.

1. What literature is reviewed? Is the review current and complete? Where does the study fit into the field? (Introduction section)
2. What is the purpose of the study? (Introduction section, especially the Statement of Purpose subsection)
3. Who was studied and how were they selected? Were there enough people in the study to make the results meaningful? (Subjects subsection)
4. What tests, questionnaires, rating scales, etc., were used? What do they look like? And, are they reliable and valid for the purposes of the study? (Materials subsection)
5. What actually happened to the subjects during the data gathering process? (Procedures subsection)
6. How were the data organized and analyzed? (Results section)
7. Is there enough information provided to replicate the study? (throughout the Method section including the Subjects, Materials, and Procedure subsections)
8. What were the descriptive results? What other statistical results came out of the study? (Results section)

9. What were the answers to the research questions and what do the answers indicate? (Discussion section)
10. What are the implications of the results, and how do they relate to the field as a whole? (Discussion or Conclusions section)
11. Which conclusions follow directly from the results and which ones are more speculative? (Discussion or Conclusions section)
12. What questions arose in the course of doing the study that might be useful for future research? (Discussion or Conclusions section)

Since answers to these questions are important in understanding any study, you can use the conventional organization of statistical studies as represented by the sections discussed here to find your way around published research articles.

However, be warned that, even though the APA format and organization are well known, some authors do not use exactly the sections and headings listed above. Sometimes there are good reasons for such deviations. For example, in the example article, there is a separate Procedures subsection as advocated in the APA manual. However, in other studies, the same author has chosen to use a combined Materials and Procedures subsection because the two issues were inextricably intertwined to the degree that they made little sense if explained separately. Regardless of the specific sections and headings used in a particular study, you should find sufficient information somewhere in any statistical study to answer the twelve sets of questions listed above.

At this point, you may wish to turn to the example article. Without reading every word, try jumping from section to section while letting the above questions guide what you have to read in order to answer them. In the process, notice that everything is found just about where the APA manual suggests. Notice also that the organization and headings help you to find the information that you need and to generally understand the study.

Examine the Statistical Reasoning Involved in the Study

In order to understand the results of a statistical study, it is necessary to understand the statistical reasoning that underlies all such studies. There are several key concepts that are necessary parts of much statistical reasoning: (a) descriptive statistics, (b) statistical differences, (c) probability levels, (d) statistical tests, and (e) significance versus meaningfulness. These five concepts will be discussed in turn.

Descriptive statistics. Part of the content of any statistical paper describes what happened in the study. As mentioned above, such description is partially accomplished within the various parts of the Method section. However, description also occurs in the Results section, which describes what happened statistically. Descriptive statistics (a phrase used in contrast to inferential statistics) describe or summarize a data set but, by themselves, cannot tell us the extent to which they represent a larger population or other, similar sets. The descriptive statistics that are most often used are indicators of the central tendency and dispersion. The central tendency (which can be viewed as a typical value for a set of numbers) is commonly reported in terms of a statistic called the *mean.* (A statistic is any number that can be computed based on the observed data.) The mean is usually exactly the same as the arithmetic average that we use in daily life. Dispersion (which can be viewed as the variation of the numerical values away from the central tendency) is usually reported in terms of a statistic called the *standard deviation.* The standard deviation summarizes how much the numbers vary away from the mean, or how much they are spread out around the mean. Ordinarily, we would expect most scores (about 68% for a normal distribution) to fall within one standard deviation of the mean. For a normal distribution, 95% of the data fall within 2 standard deviations. For instance, if the data being described are a set of test scores, their standard deviation can be defined as "a sort of average of the differences of all scores from the mean" (Brown, 1988, p. 69).

You now have enough basic information about descriptive statistics to consider the central tendency and dispersion in a real study. Such descriptive statistics most often take the form of a table. Table 2 in the example study is typical. Notice that it describes the results in terms of the various groupings involved in the study, that is, for each group and/or subgroup in the study. The first column labels groups that were created by considering the raters in two faculties (English and ESL) separately as well as combined. Across the top of the table, you will also find labels for the groups that were created by considering the different types of students separately (English and ESL composition students) and combined. The most important thing to note is that the data being grouped and described are the ratings for two types of students (English and ESL) as rated by two types of instructors (from English and ESL faculties).

Notice that for each possible combination of Student Type, and Rater Faculty, the table provides three statistics: the number of students involved in the group *(n),* the mean *(m),* and the standard

deviation *(SD)*. For instance, the descriptive statistics for the English students' compositions as rated by the English faculty indicate that there were 112 compositions involved, that the mean was 2.46 (on a 0-to-5-point scale), and that the standard deviation was 1.11. This information is interesting in itself because it indicates that the test is fairly well centered (i.e., the mean is almost exactly halfway between the lowest possible score of zero and the highest possible score of five), and the scores are spread out to a reasonable degree (i.e., there is room for 2 standard deviations above and below the mean within the range of possible scores from 0 to 5).

However, the information is also useful for comparing groups to each other. Consider the fact that the means for all of the groupings are very similar in this table. That may indicate that there were no major differences among the groups of compositions as produced by the two types of students and rated by the two types of teachers. However, note that the standard deviations for the compositions written by the English students are generally higher than those for the essays written by the ESL students. This indicates that the scores for the English composition students were more spread out than those for ESL students, that is, there was greater dispersion in the writing scores of the English course students.

Statistical differences. As useful and informative as descriptive statistics can be, they are often not enough. There is a type of statistical reasoning that takes over at this stage in most studies: *Inferential statistics* investigate the extent to which descriptive statistics represent a larger population or other, similar data sets. This mode of reasoning hinges on the concept of *significant differences*. The significant differences most often of interest in statistical studies are the differences observed in comparing means, comparing frequencies, or comparing correlation coefficients to zero.

In *comparing means* (i.e., arithmetic averages), it is possible that any observed differences are purely accidental. After all, if you give a test to a group of students on several occasions, you would expect the means to be slightly different because human beings simply do not perform exactly the same on every occasion (e.g., some students' scores might have been affected by the fact that they were sick, tired, depressed, etc., on one of the occasions). Indeed, it would be very surprising if the test results turned out to be exactly the same on successive occasions. The issue that researchers must grapple with is whether the differences that they observe between means are just such chance variations or are due to some other more systematic factor. The question being posed by the researcher and

answered through statistical tests is whether or not there is a significant difference between means.

For example, consider a hypothetical study in which the average number of correctly defined words on a vocabulary test is compared for two groups: one group that received lessons using language teaching Method X for 6 weeks, and the control group that received 6 weeks of instruction based on Method Y. The problem is that the two means will naturally vary to some degree by chance alone. The question that the researchers must resolve is whether there is a significant difference between the means (i.e., whether the observed difference between the means is systematic or occurred by chance alone).

If there is a significant difference, the researcher can say with a certain amount of confidence that the observed difference between the two means was not just accidental. This is an important issue because, if the group learning vocabulary under Method X has a higher mean than the other group, the higher number of vocabulary words learned can probably be attributed to the effects of Method X (provided the experiment was conducted properly). This would constitute an argument in favor of Method X, which might be interesting to other language educators responsible for teaching vocabulary. As described here, it is an argument based entirely on comparing the mean performances of the groups involved.

In *comparing frequencies*, it is also possible that any observed differences are due to chance variations. After all, if we tally the numbers of Taiwanese and Korean students in an ESL class on successive days, we would expect the resulting frequencies (also known as tallies) to be slightly different on different days. It might turn out that there are 7 Taiwanese and 11 Koreans on the first day, 6 Taiwanese and 12 Koreans on the second day, 7 Taiwanese and 9 Koreans on the third day, etc. Indeed, it would be surprising if the frequencies turned out to be exactly the same every day. Similarly, by chance alone, one would expect the numbers of Taiwanese and Koreans to vary from group to group because of chance variations in the proportions of different nationalities. The issue that researchers must come to grips with is whether any difference that is observed between frequencies is a chance variation or is instead due to some other more systematic factor, In short, the question being posed by a researcher who is comparing frequencies is whether or not there is a significant difference between the observed frequencies and the frequencies that would be expected by chance alone.

An example of this type of study might occur in performing a needs analysis for a language course. Perhaps the researcher is

interested in the frequency, or tally, of people who are interested in studying pronunciation as compared to those interested in studying grammar. The question that a researcher might pose is whether there is a significant difference between the existing frequencies and the fifty-fifty split that would be expected by chance. If the researcher can state that there is a significant difference, it will indicate that the observed difference is due to factors other than the chance fluctuations that would occur naturally.

Similar reasoning may be used in *comparing correlation coefficients to zero.* Correlation coefficients are indexes that represent the degree of relationship between two sets of numbers. Correlation coefficients can range from 0.00 (if there is no relationship) to 1.00 (if there is a very strong relationship). Consider the following data set:

Students	EFL Test A	EFL Test B	Last EFL Study (years since)	Age
Maria	100	97	0.0	26
Jaime	98	97	0.5	28
Carla	87	84	2.0	26
Jose	82	85	1.5	28
Juanita	77	74	4.0	27
Jimmy	55	52	3.5	27

EFL Tests A and B appear to be highly related in the sense that as one set of numbers goes up so does the other. The resulting correlation coefficient turned out to be a very high .99. If two sets of numbers are not related at all, the expected correlation coefficient is 0.00. This is the case for Age and Test B in the above example data where the correlation coefficient turned out to be very close to zero at 0.01.

It is important to note that correlation coefficients can also take on negative values anywhere between 0.00 (if there is no relationship) and —1 .00 (if there is a strong but opposite relationship). In such cases, the two sets of numbers are related but in opposite directions. For example, Test B and Last EFL Study in the above example are fairly strongly related but in opposite directions. In other words, as one set of numbers goes up, the other set goes down. The result in this case is a high but negative correlation coefficient of –.86. In short, correlation coefficients can range from –1 .00 (for strong, but opposite relationships) to 0.00 (for no relationship) to +1.00 (for strong relationships in the same direction).

One problem with correlation coefficients is that even for two sets of random numbers some degree of correlation may be found

by chance alone. For instance, on successive sets of randomly selected numbers, correlation coefficients of .12, .07, –.17, –.01, etc., might be found by chance alone. Indeed, it would be surprising if a correlation coefficient of exactly 0.00 were found every time. The issue that researchers must deal with is whether the correlation coefficients that occur in a study are just such random (or chance) variations around zero, or rather are due to some more systematic relationship between the sets of numbers. The question being posed by the researcher is whether there is a significant relationship, i.e., a significant difference between the correlation coefficient that was observed in a study and a zero (or chance) correlation. If the researcher can state that there is a significant difference, the correlation coefficient observed in the study probably varied from zero for other than chance reasons. In simpler terms, it indicates that the relationship between the sets of numbers is probably systematic—not just a chance relationship.

Probabilities. Since there is always at least some possibility that differences are due to chance, researchers use statistical tests to compute a particular significant difference in terms of the probability that observed differences would occur by chance alone. In other words, when a researcher states that there is a significant difference (between the two means, between observed and expected frequencies, or between a correlation coefficient and zero), these results will always be stated in terms of the probability that the observed difference was due to chance fluctuations.

These probabilities are usually expressed as *p* values in statistical studies. They will normally be written as $p < .01$, or $p < .05$, or as exact figures, e.g., $p = .9681$. The *p* stands for probability. In straightforward terms, *p* is the likelihood that the researcher will be wrong in stating that there is a statistical difference (whether between means, between observed and expected frequencies, or between a correlation coefficient and zero) if, in fact, there is no difference. Thus, if $p < .01$, the probability (assuming chance alone) is less than 1% that the observed difference would be so large, giving strong evidence against pure chance. Similarly, if $p < .05$, the probability is less than 5% that an observed difference this large could have occurred by chance alone.

The probability levels of .05 and .01 (also referred to as alpha levels) are used by convention in most social science research to define the threshold of statistical significance. The choice between the .01 and .05 values is governed by how strict the researcher wants to be with regard to the conclusions that are drawn from a study. When a study is about an important medicine, we want to be very

sure that it will not hurt patients. Thus a conservative .01 value might be selected so that there is a 99% probability of nonchance results. If the study is about a new way of teaching reduced forms, the decision is perhaps not quite so crucial, and therefore, we can accept the .05 value, which indicates that we are willing to accept a less restrictive 95% probability of nonchance results.

The determination of significant differences and their associated probabilities takes many forms, but the most commonly reported types are the three that have to do with means, frequencies, and correlation coefficients. It is important to note that a study seldom compares only two means, or contrasts only two frequencies, or examines one correlation coefficient to see if it varies from zero. More commonly, there are a number of means involved, or a number of frequencies, or a number of correlation coefficients to complicate the picture. Nevertheless, the underlying processes of checking significant differences and determining probabilities are the same.

Statistical tests. The process of determining statistical significance as described in the Statistical Differences and Probabilities subsections above is referred to as performing a statistical test. The three most commonly reported types of statistical tests are used in the example study: mean comparisons, comparisons of frequencies, and comparing correlation coefficients to zero.

Example mean comparisons are discussed in the example study:

> In short, the small differences among the means shown in Table 2 can only be interpreted as chance fluctuations, which are not attributable to systematic differences based on the variables used in this study.

In the above quote, the statistical reasoning for comparing pairs of means is explained. In this case, there were three comparisons of interest: (a) the difference between the mean scores for the two types of students, English Composition or ESL composition students; (b) the difference between the mean scores assigned by the English faculty and ESL faculty raters; and (c) the difference between the mean scores that resulted from the two different orders in which compositions were rated (first or second). In all three comparisons, it turned out that there were no significant differences between the two types of students, the two types of raters, and the two types of orders.

The statistical test being used is the *F* test, the results of which are based on the *F* statistic reported in the second column from the right in Table 3 of the example study. The essential information is found in the column on the far right where the probabilities are

given in the column labeled *p.* Notice that each of these *p* values has an asterisk next to it and that the asterisks refer to the statement at the bottom of the table indicating that $p > .05$. This is read as "the probability is greater than .05" and indicates that random chance alone could produce results like these more than 5% of the time; this is not convincing evidence against chance. Thus, the researcher would be wrong in stating that there was a significant difference. (In fact, the *p* values reported in the table are much higher than .05, so they indicate that the observed results are quite consistent with random chance.) The first, second, and fourth *p* values are associated with one *F* ratio each for the main factors in Table 3: Student Type, Rater Faculty, and Order. The quote cited above discusses what these three statistical tests indicate.

Example frequency comparisons are shown in Tables 4 through 7 of the example study. The statistical test being used is known as the chi-square test, or simply x^2. You will notice the asterisks in each table refer to the statement at the bottom of the table that $p < .05$. In this case, the statement would be read as "the *p* value is less than .05" and indicates that the researcher was justified (in those cases marked with an asterisk) in stating that there was a significant difference between the relative frequencies of English faculty and ESL raters who chose a particular feature, that is, the observed differences are unlikely to have occurred due to chance alone. Hence, we conclude that there are systematic differences between English faculty and ESL raters.

Notice that there are two steps involved in interpreting Table 4. First, there is the overall x^2 value to consider. This value, located in the bottom row, is found to be significant at $p < .05$ (as indicated in the line just below the table). This result simply suggests that one or more of the frequencies in the table differed from what would be expected, and more detailed analyses are justified. In order to investigate which of the specific pairs (English faculty and ESL raters) of relative frequencies might be contributing to the overall significant difference, the x^2 values for pairs were also calculated. These x^2 values are reported in the column furthest to the right. They indicate that there was a significant difference (at $p < .05$) for the English faculty raters and ESL raters on Cohesion, Organization, and Syntax (i.e., the frequencies observed for the English and ESL raters on Cohesion, Organization, and Syntax were significantly different). In contrast, the frequencies of response for English and ESL faculty were not significantly different from expectations for the rating categories of Content, Mechanics, and Vocabulary. Similar two-step interpretations can be drawn from each of the other frequency tables (Tables 5-7).

Example comparisons of correlatwn coefficients to zero are shown in Table 1. The statistical test results are based on the Pearson product-moment correlation coefficient, or simply *r*. The asterisks in the table once again refer to the statement at the bottom of the table that $p < .05$. The statement would be read as "the *p* value is less than .05," which, in this case, indicates that the researcher was justified in each case in stating that the correlation coefficient was significantly different from zero. More specifically, random chance produces such strong correlations less than 5% of the time. Based on this evidence against pure chance, the researcher was justified in stating that there was a significant difference between each correlation coefficient and zero. For instance, the results in Table 1 show that, even though it is relatively low in magnitude, the correlation coefficient of .37 between Groups A and B of the English faculty raters differs from zero for reasons other than chance. The same is true for all of the other correlation coefficients in this table. Such is not always the case. Other studies may well find correlation coefficients that are not significant, indicating just chance differences from zero with no systematic association between the sets of numbers involved.

Significance versus meaningfulness. It is important to recognize that a statistically significant difference is just that, and no more. Significant differences, whether working with means, frequencies, or correlation coefficients, simply indicate that we have concluded that the observed differences are due to other than chance factors. In other words, the differences are systematic in some way. It does not indicate that the differences are necessarily interesting or meaningful. In fact, a difference can be statistically significant, yet be so small that it is not at all meaningful or interesting.

For instance, the correlation coefficient of .37 was found to be statistically significant at .05 (i.e., the correlation coefficient is probably different from 0.00 for other than chance reasons), but the meaningfulness of the relationship between the two sets of numbers is a separate issue. In this case, the numbers are scores assigned by two raters and the low correlation coefficient indicates that there was some association between the two sets of scores but there are other important factors that are still not accounted for. The weakness of agreement found here is worrisome because it indicates that the scores may not be very reliable. Thus, this is an example of a correlation coefficient that is statistically significant, but not very meaningful in magnitude.

Similarly, if two means are statistically different at $p < .05$, yet only differ by two points out of 100, then the result might not be at

all meaningful. Likewise, if a set of observed frequencies differs from expected frequencies at $p < .01$, yet differs to an uninteresting degree, the results may not be meaningful. Thus, it is always important to examine the descriptive statistics in any study and think about any statistical tests in terms of descriptive statistics so that you can determine whether any significant differences are also meaningful.

The important thing to remember, then, is that meaningfulness is a separate issue from statistical significance and that meaningfulness will depend on all of the factors involved in the situation in which the study was conducted. When reading a statistical study, you might want to check to make sure that the researcher has kept separate these two issues of significance (i.e., can we rule out chance?) and meaningfulness (i.e., is the difference large enough to be interesting?) and interpreted them clearly.

It is important to remember that statistical studies are no more likely to be infallible than any other form of argumentation. Authors make errors, and computers make errors. However, if a study is properly carried out and the results are adequately described and systematically explained, such studies can help us to view the important issues in our field in new and useful ways.

Evaluate What You Have Read in Relation to Your Professional Experience

So far, you have used the abstract to decide if you wanted to read the study, used the organization of the paper to help you understand how the study was conducted, and used some basic concepts to interpret the statistical reasoning of a study. You are probably now at a point where it makes sense to pull away from the study a bit and think about it more critically. There are six types of questions that may prove useful in thinking about the article after having read it. These questions will enable you to know, comprehend, analyse, apply, synthesize, and evaluate what you have read. (These six categories are taken from Bloom's [1956] taxonomy. Note that they are presented here in a slightly different order from the original.) In short, after reading an article, try to recall basic information about the article by asking yourself Questions 1-3 below; then try to relate the article to your professional life by asking yourself Questions 4–6:

1. *Know:* Who wrote the article? When? In what journal? (Useful for identifying the study when referring to it)
2. *Comprehend:* In a sentence, what was the article about? (Useful for briefly summarizing the study)

3. *Analyze:* What sections was the article divided into? (Useful for recalling the overall structure of the study)
4. *Apply:* How can you apply what you learned in the article to your professional EFL/ESL teaching situation? (Useful for determining whether the article is applicable to your teaching experience)
5. *Synthesize:* How does the article relate to other professional books or papers that you have read? (Useful for seeing how the study fits into the professional literature)
6. *Evaluate:* How good was the quality of the article internally (in terms of style, organization, reasoning, etc.)? How good was it externally (i.e., in terms of everything else you know about the profession)? (Useful for evaluating the overall quality of the article)

Going through these questions (or similar ones) will help you to remember which article you read, comprehend its essential message, analyze the constituent parts of the article, apply what was learned in the article to your professional situation, synthesize what you found in the article with other points of view in the profession, and evaluate the quality of the article (both internally and externally).

Learn More about Statistics and Research Design

Having gone this far in the process of understanding statistical studies, you may now be intrigued by the prospect of learning more. For instance, you may have heard about ANOVAs, regression analyses, factor analyses, and other analyses not directly covered in this article. It is only by learning more that you will be able to understand some of these more complex analyses. In fact, it is only by learning more that you will be able to decide whether the author of a given study chose the correct statistical tests at all, or whether the assumptions that are required for many statistical tests were met.

There are a number of ways to learn more about statistics and research design. In addition to Part 2 of this discussion, there are books specifically designed to help language teachers do statistical research: Butler (1985), Hatch and Farhady (1982), Hatch and Lazaraton (1991), Seliger and Shohamy (1989), Woods and Fletcher (1986), etc. Another book, Brown (1988), is designed to help readers who are only interested in reading (rather than doing) statistical research. If the topics that interest you are more closely related to

the statistics and research in the area of language testing, it may be more appropriate to read references such as Bachman (1990) and Henning (1987). If you have no idea which book to choose, it might be useful to read Hamp-Lyons' (1989, 1990) book reviews which describe a number of the volumes listed above.

I am not advocating that every EFL/ESL teacher read and internalize all of the knowledge in these books. However, I am suggesting that a number of strategies are available to teachers: (a) for some teachers, a thorough reading of one or two of the books listed above may be just what is needed; (b) for other teachers, it may prove useful to use several of the books listed above as references to explore topics in more depth as need arises in reading statistical studies; (c) still other teachers may be more comfortable with the structure provided by taking an organized course in basic research design and statistics at a local college or university.

Regardless of the strategy that is used, learning more about statistical research can help not only in understanding the statistical studies in the professional literature but also in grappling with the research that is reported in the lay media, much of which is done in the same statistical research paradigm that is used in our field. (For an excellent and easy to read treatment of how numbers, figures, and tables are used to fool the general public, you may want to read a book appropriately titled *How to Lie with Statistics,* Huff & Geis, 1954. Yes, it is still in print.) Armed with such knowledge, teachers can then defend themselves against numbers, and understand the reasoning that surrounds their use.

CONCLUSION

This article set out to provide attack strategies for EFL/ESL teachers to use in gaining access to statistical studies. These strategies include using the abstract and conventional organization of statistical papers to guide reading, examining the statistical reasoning, critically evaluating what the results signify to each reader, and learning more about statistical studies. There are a number of reasons why I hope that some readers will find these suggestions useful. First, if the studies that appear in the *TESOL Quarterly* have a larger informed readership, such studies will have greater impact on the field. All of us must use all available information about language learning and teaching to improve the ways that we serve our EFL and ESL students. Second, it is only by having an informed readership that the quality of the statistical studies in the *TESOL Quarterly* can be assured. Though the review process for selection of articles is thorough and fair, there are no guarantees that the

articles that appear in print are 100% correct or uncontroversial. It is therefore our responsibility to read any articles that interest us as carefully and critically as we can so that the interface between teaching and research can be strengthened.
■

ACKNOWLEDGMENTS

I would like to thank Kathleen Bailey, Graham Crookes, Thom Hudson, Andrew F. Seigel, and Ann Wennerstrom for their insightful comments and suggestions on an earlier version of this paper.

THE AUTHOR

J. D. Brown is on the faculty of the Department of ESL at the University of Hawaii at Manoa. He has published numerous articles on language testing and curriculum development, and a book on critically reading statistical studies (Understanding *Research in Second Language Learning,* 1988, Cambridge University Press).

REFERENCES

American Psychological Association. (1983). *Publication manual of the American Psychological Association.* Washington, DC: Author.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bloom, B. (Ed.). (1956). *Taxonomy of educational objectives: Handbook 1. Cognitive domain.* London: Longman.

Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design.* Cambridge: Cambridge University Press.

Butler, C. (1985). *Statistics in linguistics.* Oxford: Blackwell.

Hamp-Lyons, L. (1989). Recent publications on statistics, language testing, and quantitative research methods: I. *TESOL Quarterly, 23* (1), pp. 127-132.

Hamp-Lyons, L. (1990). Recent publications on statistics, language testing, and quantitative research methods: II. *TESOL Quarterly, 24* (2), pp. 293-300.

Hatch, E., & Farhady, H. (1982). *Research design and statistics for applied linguistics.* Rowley, MA: Newbury House.

Hatch, E., & Lazaraton, A. (1991), *The research manual: Design and statistics* for *applied* linguistics. Rowley, MA: Newbury House.

Henning, G. (1987). A *guide to language testing.* Rowley, MA: Newbury House.

Huff, D., & Geis, I. (1954). *How to lie with statistics.* New York: Norton.
Seliger, H. W., & Shohamy, E. (1989). *Second language research methods.* Oxford: Oxford University Press.
Woods, A., Fletcher, P., & Hughes, A. (1986). *Statistics in language studies.* Cambridge: Cambridge University Press.

TESOL QUARTERLY, Vol. 25, No. 4, Winter 1991

Do English and ESL Faculties Rate Writing Samples Differently?

JAMES DEAN BROWN
University of Hawaii at Manoa

This study investigates the degree to which differences exist in the writing scores of native speakers and international students at the end of their respective first-year composition courses (ESL 100 and ENG 100, in this case). Eight members each from the ESL and English faculties at the University of Hawaii at Manoa rated 112 randomly assigned compositions without knowing which type of students had written each. A holistic 6-point (0-5) rating scale initially devised by the English faculty was used by all raters. Raters were also asked to choose the best and worst features (from among cohesion, content, mechanics, organization, syntax, or vocabulary) of each composition as they rated it. The results indicated that there were no statistically significant mean differences between native-speaker and ESL compositions or between the ratings given by the English and ESL faculties. However, the features analysis showed that the ESL and English faculties may have arrived at their scores from somewhat different perspectives.

A fairly large body of literature exists on the teaching of writing in ESL/EFL contexts (see Brown & Bailey, 1984; Connor, 1987; Piper, 1989; Zamel, 1987 for overviews of this work). However, within that literature, there is a surprising lack of information on cooperative efforts between faculty members who teach writing to native English students and faculty who teach ESL students. Perhaps English and ESL faculties simply do not cooperate. If that is true, it is a strange state of affairs given the amount of common ground that writing teachers have—whether they are teaching native-speakers or ESL students.

The administration of the ESL Department at the University of Hawaii at Manoa (UHM) first became interested in this issue of cooperation when summoned by the dean of the college to a meeting with him and the administration of the English Department to discuss "the ESL problem." At that meeting, anecdotal evidence was presented for the particular weakness of foreign students'

writing abilities. The ESL students finish their training in writing with the course ESL 100; in terms of fulfilling the UHM composition requirement, it is treated as an exact equivalent to the ENG 100 composition course, offered by the English Department and required of all native speakers pursuing BA and BS degrees. It was suggested at the meeting that the ESL students should be tested at the end of their training to determine whether they were up to the same "standard" as the students who took ENG 100. There was a good deal of debate over the next few months, but a compromise was reached wherein all ESL students would be tested at the end of ESL 100 if three or four sections of ENG 100 could be tested at the same time for comparison.

Purpose

The central purpose of this study was to investigate the relative writing abilities of native speakers and ESL students at the end of their different first-year composition courses. To that end, the following research questions were posed:

1. Is the holistic scoring method (i.e., the scale used by the English faculty) reasonably reliable when used by English and ESL faculty?
2. Is there a significant difference in the mean performances of students who have completed ENG 100 and ESL 100?
3. Is there a significant difference in the mean scores assigned to writing samples by English and ESL faculty?
4. Is there a significant difference in the means of the first and second scores assigned to writing samples?
5. Are there any significant differences in the best and worst features identified for compositions when assigned by English and ESL faculty?
6. Are there any significant differences in the best and worst features identified for compositions when assigned by English and ESL faculty at different score levels?

The alpha level for all statistical decisions was set at .05.

METHOD

Subjects

The subjects in this study were all students enrolled in first-year composition courses at UHM. There were a total of 112 compositions specially written for this project at the end of instruction during the spring semester. International students enrolled in ESL 100

wrote 56 compositions, and native speakers enrolled in the ENG 100 wrote the other 56 compositions. The ENG 100 compositions were written by students whose instructors had volunteered to participate in this study. The ESL 100 compositions were randomly selected from a larger set written by the entire population of ESL 100 students. This subsample of 56 ESL 100 compositions was taken so that they would be equal in number to the 56 compositions available from the ENG 100 sections.

No effort had previously been made to coordinate the ENG 100 and ESL 100 courses. As a result, they had diverged considerably in curriculum and content. In spring semester, the ESL 100 writing courses included four sections. All of these sections shared a common set of objectives, which were based on a process-oriented approach to writing instruction (see Raimes, 1983; also see Hamp-Lyons, 1986, and Horowitz, 1986, for the debate on this approach). The ENG 100 writing courses included over 50 sections. The objectives for these ENG 100 courses were set by each individual instructor. Consequently, there was a great deal of variation in the content of various sections, ranging from very traditional to very innovative curriculum and teaching methods.

Students involved in this study were undergraduates. The ESL 100 students were 55.7% male and 44.3% female and came from the following regions: East Asia (50%), Southeast Asia (21.5%), South Asia (10.7%), Pacific Basin (7.0%), Africa (5.4%) and Europe (5.4%). These students ranged considerably in overall English proficiency, but they had all scored at least 500 on the TOEFL in order to be admitted to UHM.

The ENG 100 students were all native speakers of English, with 49.1% being male and 50.8% female. It should be noted that, in this particular situation, many of the students could be expected to also be speakers of Hawaiian Creole English (see Sato, 1989, for more on this issue). The degree to which this was true for different students was not treated as a variable in this study.

Materials

Two sets of test questions were used in this research: one that presented a reading passage and analytical writing task about genetic engineering, and one that presented a more open-ended narrative writing task about the problem of watching too much television. These two topics are equally represented within the 56 ESL 100 compositions, as well as within the 56 ENG 100 compositions (see Brown & Durst, 1987, for more on these topics).

A previously established holistic scoring guide was used to rate the students' essays. The descriptors and wording for this 6-point

(0-5) scale were created by the English faculty (see Brown, 1991, for a full description). A strategy for training raters was also developed in the form of a scoring pamphlet with explanation of the scoring process and example compositions for practice in assigning ratings.

While scoring the compositions, the raters were also asked to identify the feature that they thought labeled the best feature of each composition and the one that they felt described the worst quality of each. There were six broad categories from which they could choose: cohesion, content, mechanics, organization, syntax, and vocabulary. These categories were a synthesis of five categories given in Brown and Bailey (1984) and five provided in Jacobs, Zinkgraf, Wormuth, Hartfiel, & Hughey (1981). They were discussed before the rating began until a consensus was developed about what each one meant. The results of this discussion were put on a blackboard to which the raters could refer as they proceeded through their ratings. It is important to note that gathering data on holistic scoring while also asking the raters to judge the worst and best features of each composition is not the way compositions are normally scored. The mixing of these procedures may have affected the way the holistic scores were assigned. Thus the experimental, and perhaps artificial, nature of these arrangements must be considered in interpreting the results.

Procedures

All subjects wrote their compositions during the last week of class in the spring semester. The compositions were then labeled with an identification number so that the raters would not know if a given composition was written by an ENG 100 or ESL 100 student. The students' names and other biodata were recorded separately for the same reason.

After the semester was finished, eight English faculty and eight ESL faculty convened for training in the use of the scoring guide. Each instructor then scored 28 of the compositions written by the subjects described above. The raters were allowed as much time as was necessary to go through this process.

The compositions were given to pairs of teachers from each faculty such that each writing sample would be scored by two members of the English faculty and two from ESL. The compositions were arranged in bundles so that each rater read equal numbers of ESL 100 and ENG 100 compositions. The packets were also counterbalanced so that half of the raters read each composition first then exchanged bundles with their partners. All of this careful

counterbalancing of the compositions was necessary so that simultaneous comparisons could be made between the performances of the two types of students, between the scores assigned by the two faculties, and between the first and second score assignments given by the raters within each faculty.

The distribution of ESL 100 and ENG 100 compositions was otherwise random so that discernible patterns were minimized. The instructors did know that they were dealing with both kinds of students. However, the specific purposes and details of this research were not revealed until after all of the ratings had been completed.

Analyses

The sets of scores on the compositions served as the primary dependent variable throughout much of this study. The principal independent variable of interest here was the students' background, that is, whether they were ESL students as indicated by their enrollment in ESL 100, or native speakers of English as evidenced by their presence in ENG 100. This variable is labeled Student Type in the analyses reported below, and it has two levels. A second independent variable of interest was the raters' background, that is, whether the raters were primarily trained in English literature, as indicated by the fact that they were on the English faculty, or trained in English as a second language. Individual discussions with each of the raters confirmed that the English faculty raters had literature backgrounds with no specific training in teaching writing. This variable is labeled Rater Faculty in the analyses reported below, and it has two levels. The only moderator variable used in this study was the dichotomy between whether the rater was the first or second reader of a given composition. This variable is labeled Order in this study, and it has two levels.

The interval scale scores for each composition were coded along with the nominal data for Student Type, Rater Faculty, and Order. Descriptive statistics were computed; interrater reliability was calculated using the Pearson product-moment correlation coefficient and an adaptation of the Kuder-Richardson formula 20 (K-R 20) taken from Ebel (1979).

Three-way analysis of variance (ANOVA) procedures were calculated with Student Type treated as a grouping factor, while Rater Faculty and Order were treated as repeated measures. Multivariate analyses (including Wilks' lambda and Hotelling-Lawley trace *F* statistics) were also calculated to confirm the more familiar univariate results reported here.

The nominal data gathered on the best and worst features assigned by raters for each composition were analyzed using overall chi-square statistics followed by more detailed analyses based on the same statistic. The purpose here was to zero in on significant differences in raters' views between and within the two faculties.

RESULTS

Reliability

Since the results of this study can be no more reliable than the measures upon which it is based, the issue of reliability (as raised in research question Number 1) will be discussed first. Reliability was initially addressed by exploring the interrater correlation to determine the degree of relationship between the scores assigned by half of the raters from each faculty with those scores given by the other half. Assignment of the raters to these two halves, Group A or B, was random. The correlations between the groups are reported in Table 1. They were all found to be significantly different from zero ($p < .05$, based on two-tailed tests).[1] The combined reliability reported at the bottom of Table 1 is based on the two scores from each faculty taken together as well as on all four scores combined from both faculties. These combined figures

TABLE 1
Correlation Matrix for Rater Groups

	Interrater correlations			
	ENG/A	ENG/B	ESL/A	ESL/B
ENG/A	1.00	.37*	.58*	.36*
ENG/B	37	1.00	.45*	.37*
ESL/A	.58	.45	1.00	.47*
ESL/B	.36	.37	.47	1.00
	Combined reliabilities			
ENG A&B	.54			
ESL A&B	.64			
All raters (ENG & ESL)	.76			

* $p < .05$, $n = 112$.

[1] These significance results along with those reported in Tables 4–7 and the F_{max} test based on the statistics in Table 2 should be interpreted cautiously: The 448 reported essay scores are based on multiple readings of 112 essays.

were calculated using the K-R 20 estimate for rating scales given in Ebel (1979, pp. 282-284).

These reliability estimates are not difficult to interpret. Consider, for example, that according to Table 1, all raters together produced scores which were reliable at about .76. This can be interpreted directly as the proportion of the score variance that is consistent. In other words, approximately 76% of the variance among scores can be considered consistent (or systematic) variance, while the remaining 24% must be viewed as random (or unsystematic) variance. Such information is useful m the sense that it helps understand the degree to which the students' writing abilities are being assessed in a consistent manner.

Mean Differences

The means and standard deviations for the compositions in ENG 100 and ESL 100 as rated by English and ESL faculty are shown in Table 2. Remember that the scores reported here are based on the 6-point scale used at UHM (for more information see the Materials subsection of Brown, 1991). The differences between ENG 100 and ESL 100 students' mean scores appear to be small, as do the differences between the means of scores assigned by the raters from the English and ESL faculties. These differences proved equally unimpressive from a statistical point of view. The source table shown in Table 3 indicates that none of the main effects due to Student Type (ESL 100 vs. ENG 100), Rater Faculty (English faculty vs. ESL faculty), or Order (whether the rater was first or second reader of a composition) was significant. Nor did any of the interaction effects for these factors show any signs of approaching statistical significance. Thus the null hypotheses of no mean

TABLE 2
Descriptive Statistics for Stuent Type and Rater Faculty

Rater faculty	Student type								
	ENG 100 essay scores			ESL 100 essay scores			Essay scores combined		
	n	*m*	*SD*	*n*	*m*	*SD*	*n*	*m*	*SD*
ENG	112	2.46	1.11	112	2.30	.77	224	2.38	.96
ESL	112	2.37	1.16	112	2.31	.96	224	2.34	1.06
Faculties combined	224	2.42	1.14	224	2.30	.87	448	2.36	1.01

TABLE 3
Three-Way ANOVA with Repeated Measures

Source	*SS*	*df*	*MS*	*F*	*P**
Between subjects effects					
Student type	2.315	1	2.315	1.003	0.319
Within groups	253.915	110	2.308		
Within subjects effects					
Rater faculty	0.487	1	0.487	0.819	0.367
Student X Rater	0.362	1	0.362	0.609	0.437
Within groups	65.386	110	0.594		
Order	0.799	1	0.799	1.150	0.286
Student X Order	0.049	1	0.049	0.070	0.792
Within groups	76.396	110	0.695		
Rater X Order	0.276	1	0.276	0.544	0.462
Student X Rater X Order	0.116	1	0.116	0.228	0.634
Within groups	55.864	110	0.508		

* $p > .05$.

differences for Student Type, Rater Faculty, and Order cannot be rejected. Simply put, any differences observed among the means shown in Table 2 can only be attributed to chance fluctuations rather than to systematic differences due to the independent and moderator variables analyzed here.

Table 2 also indicates that the scores for ESL students' compositions are somewhat more homogeneous than those for the ENG students, as indicated by the smaller standard deviations for the ESL students. Similarly (comparing the standard deviations of English and ESL faculty scores), the English faculty raters appear to produce scores that are slightly more homogeneous than those assigned by the ESL raters. In order to determine whether apparent differences among standard deviations are significant, the F_{max} test compares the smallest to the largest standard deviation by squaring their ratio (i.e., it tests homogeneity of variance). In this case of the data reported in Table 2, omitting the Combined row and columns, $F_{max} = 2.2695$, which was significant at $p < .05$ ($k = 4$, $df = 111$). Thus some of the observed differences between standard deviations are interpreted as probably due to other than chance factors.

More importantly, this result indicates that there is a probable violation of a restrictive assumption (i.e., homogeneous variances) that accompany repeated-measures designs like the one reported here. To address this issue, multivariate analyses were also conducted. Since the multivariate *F* statistics for each effect and the

interactions led to the same conclusions as those reached by the repeated-measures design, it is with confidence that the more familiar repeated-measures ANOVA results are presented in Table 3. In short, the small differences among the means shown in Table 2 can be interpreted as chance fluctuations, which are not attributable to systematic differences based on the variables used in this study.

Cross-Tabulation of Features

Given that no statistically significant mean differences were detected in the ratings assigned by the two faculties, one question that remains is whether there are any differences in the ways that ESL and English instructors arrive at their ratings. Recall that the raters were asked to identify the feature that they thought labeled the best quality of each composition and the one that they felt described the worst quality of each, and that there were six broad features from which they could choose as follows: Cohesion, Content, Mechanics, Organization, Syntax, and Vocabulary.

The results for these analyses begin in Table 4, in which the best features identified by the raters in each of the faculties are shown. The comparable figures for worst features are shown in Table 5. Notice that, at the bottom of each of these tables, an overall statistically significant chi-square value is reported, indicating that the English raters differ systematically from those of the ESL raters. Because these overall chi-square values were significant, more

TABLE 4
Overall Best Features Identified by Each Faculty

Best feature	Raters: ENG No.	ENG %	ESL No.	ESL %	Total No.	Total %	χ^2
Cohesion	55	24.6	26	11.6	81	18.1	10.38*
Content	71	31.7	86	38.4	157	35.0	1.43
Mechanics	21	9.4	24	10.7	45	10.0	.20
Organization	32	14.3	65	29.0	97	21.7	11.23*
Syntax	29	12.9	6	2.7	35	7.8	15.11*
Vocabulary	16	7.1	17	7.6	33	7.4	.03
Total	224	100	224	100	448	100	38.39(df=5)*

Note. Due to rounding, column percentages may not total 100.

* $p < .05$ ($df = 1$).

detailed analyses were justified. Hence, chi-square values were calculated for the two faculties on each of the features, Notice that these chi-square (x^2) statistics are reported in the column furthest to the right in each table and that those which were statistically significant have an asterisk. Note also that the percents reported throughout the tables are those for the columns, not for rows.

In Table 4, the feature most often identified as best by the combined raters was Content (chosen for about 35% of the compositions), while Vocabulary and Syntax were the least often associated with the best feature (7.4% and 7.8%, respectively). The two faculties seem to be more or less in agreement on these three features. Markedly divergent differences emerge on the other three features. The English faculty chose Cohesion more than twice as often as the ESL raters (24.6% and 11.6%, respectively) and Syntax nearly five times as often (with 12.9% and 2.7%, respectively). The reverse was true for Organization which was assigned more than twice as often by ESL raters (29.0% for ESL and 14.3% for English faculty).

In Table 5, the features most often identified by both groups as worst were Content (chosen for about 32.8% of the compositions) and Syntax (31.3%), while Cohesion, Mechanics, and Vocabulary were the least often associated with the worst feature (7.8%, 8.5%, and 8.3%, respectively). The two faculties seem to be more or less in agreement on all features except Content and Mechanics for which statistically significant differences were found. The ESL raters

TABLE 5
Overall Worst Features Identified by Each Faculty

Worst feature	Raters						
	ENG		ESL		Total		
	No.	%	No.	%	No.	%	χ^2
Cohesion	14	6.3	21	9.4	35	7.8	1.40
Content	59	26.3	88	39.3	147	32.8	5.72*
Mechanics	28	12.5	10	4.5	38	8.5	8.53*
Organization	21	9.4	30	13.4	51	11.4	1.59
Syntax	80	35.7	60	26.8	140	31.3	2.86
Vocabulary	22	9.8	15	6.7	37	8.3	1.33
Total	224	100	224	100	448	100	21.42(df=5)*

Note. Due to rounding, column percentages may not total 100.

* $p < .05$ ($df = 1$).

chose Content about half again as many times as the English faculty (39.3% and 26.3%, respectively). The reverse was true for Mechanics which was assigned almost three times as often by English faculty (12.5% for English and 4.5% for ESL).

The fact that such marked differences existed in the views of the raters from the two faculties with regard to both the best and worst features, led to further analyses of these features broken down by score levels. Tables 6 and 7 present the same information covered in Tables 4 and 5, but subcategorized into low, middle, and high score ranges. A glance at the totals across the bottom of Tables 6 and 7 will indicate that considerably more scores were assigned in the middle range than in either of the other ranges and that this was true for raters from both faculties. Observe also that the features assigned to compositions as best and worst seem to vary even within faculties (among the three score ranges). The chi-square values are given for the numbers of compositions assigned each feature for the three score ranges.

For the best features reported in Table 6, Cohesion produced significant chi-square values for the low, middle, and high score ranges within each of the faculties. It appears that Cohesion was applied more often as the best feature for lower scoring compositions in both groups of raters. Content shows the reverse pattern for both faculties, that is, it was more often applied to high scoring compositions. Mechanics only evidenced significant differences for the three score ranges of the ESL raters, and even then, it was listed only for low and middle scores. There were no significant differences for score ranges within either faculty on the Organization, Syntax, or Vocabulary features.

For the worst features reported in Table 7, Cohesion exhibited no significant differences. However, Content did show significant differences for the score ranges within each of the faculties. It appears that this feature is more often the worst feature for low scoring compositions in both groups of raters, though more markedly so among ESL raters. Mechanics only showed significant differences within the ESL faculty (in favor of the low and middle ratings). There were no significant differences for the three score ranges within either faculty for Organization, Syntax, or Vocabulary.

In short, with respect to the best features, both groups appear to consider Content an important positive feature. However, their views diverge considerably on Cohesion, Organization, and Syntax assignments, with both groups applying Cohesion to lower scoring compositions and Content to higher scoring ones. Aside from significant differences for Mechanics within the ESL rater group, all

TABLE 6
Best Features for Each Score

Best feature	English faculty scores							ESL faculty scores						
	Low		Middle		High			Low		Middle		High		
	No.	%	No.	%	No.	%	χ^2	No.	%	No.	%	No.	%	χ^2
Cohesion	17	42.5	32	20.0	6	25.0	6.60*	11	21.6	14	9.9	1	3.1	6.69*
Content	4	10.0	56	35.0	11	45.8	8.00*	12	23.5	50	35.5	24	75.0	14.42*
Mechanics	7	17.5	14	8.8	0	0.0	5.13	9	17.6	15	10.6	0	0.0	5.72
Organization	4	10.0	23	14.4	5	20.8	1.24	12	23.5	48	34.0	5	15.6	3.73
Syntax	5	12.5	22	13.8	2	8.3	0.48	2	3.9	4	2.8	0	0.0	1.16
Vocabulary	3	7.5	13	8.1	0	0.0	1.94	5	9.8	10	7.1	2	6.3	0.45
Total	40	100	160	100	24	100		51	100	141	100	32	100	

Note. Due to rounding, column percentages may not total 100.

* $p < .05$ ($df = 2$).

TABLE 7
Worst Features for Each Score

Worst feature	English faculty scores							ESL faculty scores						
	Low		Middle		High			Low		Middle		High		
	No.	%	No.	%	No.	%	χ^2	No.	%	No.	%	No.	%	χ^2
Cohesion	2	5.0	9	5.6	3	12.5	1.70	1	2.0	16	11.3	4	12.5	3.91
Content	19	47.5	35	21.9	5	20.8	8.29*	31	60.8	55	39.0	2	6.3	14.89*
Mechanics	4	10.0	21	13.1	3	12.5	0.25	3	5.9	1	0.7	6	18.8	19.31*
Organization	3	7.5	15	9.4	3	12.5	0.40	6	11.8	21	14.9	3	9.4	0.72
Syntax	8	20.0	64	40.0	8	33.3	3.63	10	19.6	38	27.0	12	37.5	2.35
Vocabulary	4	10.0	16	10.0	2	8.3	0.06	0	0.0	10	7.1	5	15.6	7.26*
Total	40	100	160	100	24	100		51	100	141	100	32	100	

Note. Due to rounding, column percentages may not total 100.

* $p < .05$ ($df = 2$).

other best feature results within and between faculties were not statistically significant.

With respect to the worst features, the English and ESL faculties appear to agree that Syntax is an important negative feature. However, they differ considerably on Content and Mechanics. Content shows an opposite pattern here from that which it produced as a best feature, that is, as a worst feature, it is applied more frequently to the lower scores. Aside from some additional significant differences for Mechanics and Vocabulary within the ESL rater group, all other best feature results within and between faculties can only be attributed to chance.

DISCUSSION

The main thrust of this project was to determine whether the two populations of students enrolled in ENG 100 and ESL 100 differed in their performance in writing at the end of freshman composition training. The results indicate that the observed differences in the overall performance of these two groups of students when they were rated by instructors from each faculty separately or by all of the instructors collectively can be attributed to chance. In addition, it was found that the raters within and between the two faculties did not, on average, vary significantly in the scores that they assigned to compositions, whether written by ENG 100 students or ESL 100 students.

It is important to note that scores did not vary systematically from one kind of student to the other or from one type of faculty rater to the other. In the repeated-measures ANOVA described above, the vast majority of the variance was found within cells where we would like to see it, that is, ideally, all of the variance in scores would be attributable to differences among the students' writing abilities not to differences in their background, the raters' faculty, and/or differences in the order in which the composition was rated. Therefore it was considered a positive result that most of the variance in this study remained within cells.

One problem is that this within-cells variation can be attributed to both consistent variance in the students' abilities and to random variance. The degree to which variance can be assigned to one or the other of these sources is a question of reliability. It is therefore worrisome that the reliability estimates reported in Table 1 were only moderately high. Apparently, even when used under the controlled conditions of this study, the rating scale still needs improvement from the point of view of consistency. This can be accomplished through more intense training of the raters, through

rewriting and improving the descriptors on the scoring scale, and/or through use of more topics and raters on each composition. Such efforts are currently being made at UHM because we owe it to our students, whether native speakers or international students, to use the most reliable assessment scale available.

The best and worst feature analyses indicate that both faculties attend to Content as a primary positive feature. Yet, the English faculty members appear to pay more attention to Cohesion and Syntax than do ESL raters, while the latter group appears to consider Organization more important. In terms of negative features, both groups seem to attend to Syntax as a primary negative feature with Mechanics being of somewhat more interest to English faculty and Content being of more interest to ESL raters.

CONCLUSION

One of the most important results of this study is the apparent lack of difference between the writing produced by native speakers and ESL students at the end of ENG 100 and ESL 100. Perhaps equally interesting is the finding that there was no significant difference between the scores assigned by instructors from the English and ESL faculties. It appears that, on average, English and ESL faculty assign very similar scores—regardless of differences in background or training. They may, however, arrive at those scores from somewhat different perspectives (as indicated by the features analysis).

The primary short-term benefit that was reaped from this cooperative research project (from the local ESL perspective) was the quelling of all discussion of the "ESL problem." However, there are a number of other, more long-term benefits that were gained from the cooperation of the English and ESL faculties at UHM. These benefits would probably apply at other institutions as well. It turns out that there can be a noticeable increase in consultations between faculties on policy and testing issues. Workshops can be conducted to help instructors from both faculties understand the problems and solutions of the other approach to writing instruction. Further cooperative research can be conducted for the benefit of both faculties.

Most important, one might speculate that such cooperation can lead to better instruction so that the students who are affected by these writing programs can make larger gains in their writing abilities. In the long run, both ESL and English faculties have a vested interest in helping all students to fulfill their writing needs, However, it is only by working together that ESL and English

faculties can maximize all instructors' capabilities in the teaching and testing of writing so that all students—native-speakers and international students alike—can benefit equally.

Further Research

Some of the questions raised in the process of doing this research are listed here in the hope that they will stimulate further research:

1. Would similar results be obtained if this study were replicated at other institutions?
2. How would raters from English and ESL faculties differ if they were to use an analytic rating scale designed to produce separate scores for each of the six features examined in this study? Would there be any relationship between the best and worst features identified for compositions and the scores assigned?
3. What alternative and/or additional sources of information (e.g., the American College Test [ACT] Verbal scores, grade point average, Test of Written English scores, portfolios, etc.) should be used in studying the similarities and differences between English and ESL students and instructors?
4. How do international students and native speakers compare in writing performance in regular academic courses?

■

ACKNOWLEDGMENTS

I would like to thank Dr. Russell Durst of the English Department at UHM for his help in collecting the writing samples from that department. I would also like to thank Drs. Despain, Hilgers, Sibley, and Stillians of the English Department for their help in planning and designing this study, as well as Drs. Hudson and Pennington of the ESL Department, Dr. A. Siegel of the University of Washington, and two anonymous *TESOL Quarterly* reviewers for their helpful comments and suggestions on earlier drafts of this paper.

THE AUTHOR

J. D. Brown is on the faculty of the Department of ESL at the University of Hawaii at Manoa. He has published numerous articles on language testing and curriculum development, and a book on critically reading statistical studies *(Understanding Research in Second Language Learning,* 1988, Cambridge University Press).

REFERENCES

Brown, J. D. (1991). 1990 *Manoa Writing Examination* (Manoa Writing Project Technical Report No. 5). Honolulu, HI: University of Hawaii at Manoa.

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning, 34* (4), 21-42.

Brown, J. D., & Durst, R. (1987). Testing writing across the University of Hawaii at Manoa curriculum. *Report on the Educational Improvement* Fund 1986/87 (pp. 34-40). Honolulu, HI: Office of Academic Affairs, University of Hawaii at Manoa.

Connor, U. (1987). Research frontiers in writing analysis. *TESOL Quarterly, 21* (4), 677-684.

Ebel, R. L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Hamp-Lyons, L. (1986). No new lamps for old yet, please. *TESOL Quarterly, 20* (4), 790-796.

Horowitz, D. (1986). Process, not product: Less than meets the eye. *TESOL Quarterly, 20* (1), 141-143.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach.* Rowley, MA: Newbury House.

Piper, A. (1989). Writing instruction and the development of ESL writing skills: Is there a relationship? *System, 17* (2), 211-222.

Raimes, A. (1983). *Techniques in teaching writing.* Rowley, MA: Newbury House.

Sato, C. J. (1989). A nonstandard approach to standard English. *TESOL Quarterly, 23* (2), 259-282.

Zamel, V. (1987). Recent research on writing pedagogy. *TESOL Quarterly, 21* (4), 697-715.

TESOL QUARTERLY, Vol. 25, No. 4, Winter 1991

Communicating About Grammar: A Task-Based Approach

SANDRA FOTOS and ROD ELLIS
Temple University Japan

Providing learners with grammar problems they must solve interactively integrates grammar instruction with opportunities for meaningful communication. This article reports the results of an exploratory study of the use of a communicative, grammar-based task in the college EFL classroom. The two research questions addressed are whether the task successfully promoted L2 linguistic knowledge of a specific grammar point and whether it produced the kind of negotiated interaction which has been assumed to facilitate L2 acquisition. The limited results of this investigation suggest that the grammar task encouraged communication about grammar and enabled EFL learners to increase their knowledge of a difficult L2 rule.

A continuing controversy in second language pedagogy is whether grammar should be taught. On the one hand, there are those who adopt a "zero position." They maintain that the teaching of grammar, has only a minimal effect on the acquisition of linguistic competence in a second language. Krashen (1985), for instance, argues that acquisition only takes place when learners are exposed to roughly tuned input which they are able to comprehend and that learning is limited to a few simple portable rules. On the other hand, there are those who argue for grammar teaching. White (1987) claims that some grammatical forms cannot be acquired solely on the basis of comprehensible input and that formal instruction may be necessary to ensure that learners obtain the data they need to acquire these forms.

In contrast to the disagreement over the role of grammar teaching, there is now broad agreement that learners need opportunities to engage in communication based on an exchange of information. Having learners participate in a variety of tasks which encourage them to negotiate meaning when communication problems arise is considered essential, both to ensure that they obtain sufficient comprehensible input for the acquisition of linguistic competence

(Long, 1983), and to provide the real operating conditions needed to develop the kind of strategic competence which is necessary for the development of fluency (Brumfit, 1984).

The purpose of this article is to demonstrate that it is possible to integrate the teaching of grammar with the provision of opportunities for communication involving an exchange of information. Learners are given grammar tasks which they solve interactively (see Dickens & Woods, 1988). Following a discussion of the roles of formal instruction and communicative language teaching in L2 acquisition, this paper reports on an exploratory study designed to investigate whether this kind of task is successful in developing L2 linguistic knowledge and in promoting the kind of interaction which is believed to facilitate L2 acquisition.

THE ROLE OF FORMAL INSTRUCTION IN L2 ACQUISITION

Bialystock (1981) hypothesizes that learners formulate two distinct kinds of knowledge, "explicit" and "implicit" (p. 34). The former refers to knowledge that is analyzed and abstract. It is available to learners as a conscious representation, so that, if called upon, learners are able to say what it is that they know. Explicit knowledge is not the same as metalinguistics knowledge (i.e., knowledge of grammatical terms), although this may help in its articulation. Implicit knowledge refers to knowledge that is intuitive and procedural. It is not consciously available to learners. Native speakers, for example, are generally unable to describe the rules they use to construct actual sentences. Both explicit and implicit knowledge can be used in communication, but there are limits on learners' ability to access the former. Effective participation in face-to-face conversation, for instance, requires implicit knowledge.

A key issue is the relationship between explicit and implicit knowledge, in particular, whether the two types of knowledge are completely distinct (Krashen, 1981) or whether one type changes into the other (Sharwood Smith, 1981). The position we wish to adopt lies somewhere in between these. Our position is based on studies which have investigated the effects of formal instruction on the acquisition of grammatical knowledge. (For detailed reviews, see Ellis, 1990, and Long, 1988.) These studies suggest the following tentative conclusions:

1. Formal instruction helps to promote more rapid L2 acquisition and also contributes to higher levels of ultimate achievement (Long, 1988).

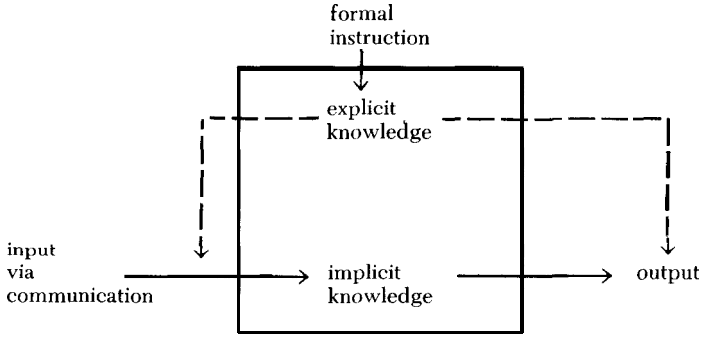
2. There are psycholinguistic constraints which govern whether attempts to teach learners specific grammatical rules result in implicit knowledge. Formal instruction may succeed if the learners have reached a stage in the developmental sequence that enables them to process the target structure (Pienemann, 1984). Conversely, it will not succeed if learners have not reached the requisite developmental stage.
3. Practice is not sufficient to overcome these constraints. There is no clear evidence to suggest that having learners produce sentences that model the target structure results in its acquisition as implicit knowledge. Studies by Schumann (1978), Ellis (1984), and Kadia (1988), among others, suggest that formal instruction directed at developmental or difficult grammatical structures has little effect on performance in spontaneous language use. (The term *developmental* refers here to structures that are acquired in stages and involve the learner passing through a series of transitional phases before mastering the target structure. Examples of developmental structures are negatives and interrogatives.)
4. It is possible, however, that formal instruction directed at relatively simple grammatical rules (such as plural and third-person *–s* or copula *be*) will be successful in developing implicit knowledge, as such forms do not require the mastery of complex processing operations (Pica, 1983; Pienemann, 1984).
5. Formal instruction is effective in developing explicit knowledge of grammatical features. There is substantial evidence to suggest that formal instruction is successful if the learning outcomes are measured by means of an instrument that allows for controlled, planned language use (e.g., an imitation test, a sentence-joining task, or a grammaticality judgment task). It is in this kind of language use that learners are able to draw on their explicit knowledge. Studies by Kadia (1988); Lightbown, Spada, and Wallance (1980); Schumann (1978); and Zobl (1985) all support such a conclusion.
6. Formal instruction may work best in promoting acquisition when it is linked with opportunities for natural communication (Spada, 1987).

Ellis (1990) suggests that the main mechanism by which formal instruction works is by developing explicit knowledge of grammatical features which, subsequently, helps learners to acquire implicit knowledge. Explicit knowledge contributes to L2

acquisition in two major ways. First, knowing *about* a grammatical feature makes the learner more likely to notice that feature in input and, therefore, to acquire it as implicit knowledge. But implicit knowledge will not be achieved until learners are ready to integrate the L2 feature into their interlanguage systems and, in many cases, this will be subject to developmental constraints. Second, explicit knowledge can be used to construct planned utterances, which then serve as input for the language processing mechanisms. The role of explicit knowledge, however, is a limited one because there are restrictions on how much explicit knowledge the typical learner can learn. As indicated above, formal instruction can accelerate knowledge while failing to contribute directly to implicit knowledge of specific linguistic features. Also, explicit knowledge plays only a "monitoring" role in communicative language use. This is a positive role, however, because it accelerates the process of acquiring implicit knowledge and may even be necessary for the acquisition of certain kinds of grammatical rules that evidence suggests (Hammerly, 1987) cannot be acquired solely by means of input derived from communicative language use.

This model, shown schematically in Figure 1, has a number of implications for formal instruction. First, it suggests that the goal of formal instruction should be directed at explicit rather than implicit knowledge. Although formal instruction may succeed in developing implicit knowledge of simple rules (see Conclusion 4 above) and also of developmental rules if the learner is ready for these (see Conclusion 2 above), it is not possible to predict easily and with sufficient precision when these conditions have been met. At the present time, it is more useful, therefore, to limit the formal instruction to explicit knowledge.

FIGURE 1
A Model of Instructed Second Language Acquisition



Second, Ellis' model suggests that formal instruction should be directed at ensuring that learners know about a target structure and can monitor with it (i.e., consciously correct their own erroneous output) but not at enabling them to use the structure in free communication. This in turn suggests that the kind of grammar teaching that is required is one that aims at consciousness-raising rather than practice. Most traditional approaches to grammar teaching are based on providing the learners with opportunities to use the target structure, first in controlled production, and subsequently in free or communicative practice (see Ur, 1988). These opportunities constitute "practice." The approach we have in mind is one that downplays the role of production and, instead, emphasizes the role of cognitive understanding. One way in which this can be achieved is by constructing various problem-solving tasks that require learners to consciously analyze data in order to arrive at an explicit representation of the target feature.

Third, this proposed model suggests that formal instruction needs to be accompanied by instruction that provides learners with opportunities for authentic communication (see Conclusion 6 above).

COMMUNICATIVE LANGUAGE TEACHING AND SECOND LANGUAGE ACQUISITION

One of the main aims of communicative language teaching is to provide opportunities for learners to participate in interaction where the primary goal is to exchange meaning rather than to learn the L2. How does this help acquisition?

The interaction hypothesis advanced by Long (1983) claims that the comprehensible input which results from attempts to negotiate communication difficulties helps to make salient grammatical features which are problematic to learners and thus facilitates acquisition. Long emphasizes the importance of interfactional adjustments in two-way communication. Examples of such adjustments are comprehension checks (e.g., *D'you know what I mean?)* and clarification requests (e.g., *What do you mean by* _____ *?).*Exchanges which involve the resolution of some communication problem by means of such adjustments are believed to promote acquisition. Although Long does not make the distinction between explicit and implicit knowledge, it is clear from his writings that by "acquisition" he is referring to implicit knowledge.

The comprehensible output hypothesis has been proposed by Swain (1985) as a complement to Long's interaction hypothesis. It claims that learners need the opportunity for *pushed output* (i.e., output that is precise, coherent, and situationally appropriate) in

order to develop advanced levels of grammatical competence. Allen, Swain, Harley, & Cummins (1990) show that classrooms with abundant comprehensible input, such as immersion classrooms, may not typically afford many opportunities for such pushed learner output. This may explain why immersion learners often fail to acquire certain marked grammatical features, such as the distinction between French *imparfait* and *passé composé.* Pica, Holliday, Lewin, & Morgenthaler (1989) have observed that learners produce pushed output when they are required to respond to clarification requests.

A task-based approach to language pedagogy can provide opportunities for the kinds of interaction which have been suggested to promote acquisition. Long (1989) proposes four general points regarding the effectiveness of different task types:

1. Two-way tasks produce more negotiation of meaning than one-way tasks, since the former make the exchange of meaning obligatory, whereas the latter do not.
2. Planned tasks, where learners prepare their speech or think about what they will say beforehand, encourage more negotiation than unplanned tasks.
3. Closed tasks, where there is a definite solution or ending, produce more negotiation than open tasks, where there is no clear resolution.
4. Convergent tasks, where the participants must agree on a solution, promote more negotiation than divergent tasks, where different views are permitted.

The adoption of a task-based approach in language pedagogy is closely linked to the use of pair/group work (see Nunan, 1989). A survey of research on pair/group work conducted by Long and Porter (1985), together with the results of studies by other researchers (Doughty & Pica, 1986; Pica & Doughty, 1985; Porter, 1986; Rulon & McCreary, 1986), indicate that learners produce more in pair/group work, use longer sentences, and do not speak any less grammatically than they do in teacher-fronted lessons. Learners also negotiate meaning more, provided that the task requires information exchange. One disadvantage, though, is that the input they receive from other learners may be less grammatical than what they obtain from the teacher.

GRAMMAR TASKS FOR COMMUNICATION

Formal instruction and communicative language teaching can be integrated through the use of grammar tasks designed to promote

communication about grammar. These grammar tasks have two primary aims: to develop explicit knowledge of L2 grammatical features and to provide opportunities for interaction focused on an exchange of information. They can be completed in teacher-directed lessons or they can be used in pair / group work in order to increase opportunities for negotiating meaning.

Grammar tasks will need to aim at raising the learner's consciousness about the grammatical properties of the L2. The tasks will not be designed to provide multiple opportunities for producing sentences containing the target features. Any production that occurs will be incidental and not directed at "acquiring" the target features, only at "learning" them. Also, the grammar tasks will need to incorporate a multiway information gap which requires the exchange of information in order to reach an agreed solution to a problem. The tasks will be designed so that they are closed, i.e., there is a single solution.

An example of such a task is shown in the appendices. The task consists of four task cards (Appendix A) and a task sheet (Appendix B). The task cards list a number of grammatical and ungrammatical sentences illustrating the use of dative verbs. They specify which sentences are correct and which are incorrect. The task sheet provides the learners with some basic grammatical information concerning dative verbs and also supplies them with some useful metalinguistics terminology (e.g., *direct* and *indirect* object). In addition, the task sheet contains a chart to fill in for each of the verbs for which data has been supplied. Finally, it instructs the learners to formulate three rules about the different kinds of dative verbs in terms of the sentence patterns they permit. This task was designed for use in pairs or groups of four learners. It required learners to (a) exchange the information on their task cards in order to complete the chart on the task sheet, (b) talk about the information in order to agree on the results, and (c) report to the class the rules they had formulated.

The study which we now report was based on this task. It was set up to investigate to what extent the task was successful in developing an explicit understanding of how dative verbs work and also whether the task produced the kind of interactions which have been suggested to facilitate L2 acquisition.

THE STUDY

The two research questions addressed were:

1. Is study of a specific linguistic feature (dative alternation) through performance of a grammar task as effective as study of

the same feature through traditional, teacher-fronted grammar instruction, as measured by test scores on a grammaticality judgment test?

2. Is the grammar consciousness-raising task used here interactive in the sense that its performance results in the same kinds and quantity of interfactional adjustments which have been reported in other studies based on two-way information-gap tasks performed in pairs/groups (see Doughty& Pica, 1986)?

Subjects and Design

The subjects for this research were two groups of Japanese EFL college students: first-year English language majors at a women's junior college, determined to be intermediate level on the basis of listening subtest scores of the Michigan English Placement Test and the listening and grammar subtest scores of the Comprehensive English Language Test (CELT), and first-year Business Administration majors, predominantly male, at a private 4-year university. Standardized test scores were not available for the latter group, but these students were considered to be "basic level," having received scores below 60% on a department listening exam. In each case, students from two Oral English classes were combined and assigned to one of three groups on a random basis. In one group, the grammar task was performed by groups of four students and dyads in one classroom, and all participation patterns were audiotaped. (Students were divided into two participation patterns to investigate the findings of Doughty& Pica, 1986, who reported that pairs of students tended to produce more negotiations than did groups.) In another classroom, a traditional, teacher-fronted grammar lesson was presented in English by the native-speaker instructor to the second group; the lesson was audiotaped only at the university. The remaining students served as the control group and worked on a reading assignment in a separate classroom during the treatments.

The design of the pre- and posttests, data sheets, and data cards required several steps. As mentioned, the problematic grammar feature selected was dative alternation. This refers to the position of the indirect object in the sentence. There are three patterns of indirect object placement in English verbs. The first allows placement of the indirect object either after the verb or as a prepositional phrase at the end of the sentence *(I gave her the book; I gave the book to her)*. The second pattern permits placement of the indirect object only as a prepositional phrase and generally is the case with Latinate verbs *(The teacher pronounced the word for the*

students). The third pattern is applicable only to a limited set of verbs, such as the verb *ask* meaning inquire, and necessitates placement of the indirect object immediately after the verb *(She asked the teacher a question).*

Ten verbs were selected on the basis of observed errors in student usage, and a pilot grammaticality judgment test of 20 sentences, 2 per verb, was designed. A grammaticality judgment test was chosen as our measure of language proficiency because, as we have suggested, testing instruments which allow controlled, planned language promote the use of explicit knowledge. In the pilot study, the test was administered to 18 second-year women's junior college English majors. These students were then given a pilot task sheet which listed the verbs used on the test and required the students to generate two rules explaining the possible position of the indirect object and to place the verbs given under the appropriate rule. The students performed the task in groups of four, and two groups were taped for negotiation counts.

Negotiation was obligatory since each student had a task card with five sentences marked correct or incorrect which she had to read to the other group members, who noted the position of the indirect object on their task sheets. Negotiations were observed to consist of comprehension checks, clarification requests, confirmation checks, and repetitions as students listened to and tried to understand the sentences being read, and checked whether they were correct or incorrect. After students performed the task, which took 30 minutes, the same test was given again as a posttest. A significant difference was obtained for pre- and posttest scores *(n* = 18, pretest *m* = 15.94, posttest *m.* = 19.17, paired $t = -10.825$, $df = 17$, $p < .001$) and negotiation counts were greater than 15 for each group.

The test was then redesigned to exclude items which were not problematic and the task sheet and task cards were rewritten. Figures 2 and 3 display the final form of the task sheet and task cards, and the grammaticality judgment test is included in Appendix C. These were administered to the two groups of subjects described above. Two weeks after each grammar treatment, the same test was administered again as a final test to measure longer-term learning.

The content of the traditional grammar lesson was identical to the information given on the task sheet and task cards and took the same amount of time to cover—20 minutes. The teacher wrote the correct and incorrect sentences on the board and pointed out the placement of the indirect object, asking the students whether they thought the sentences were correct or incorrect. The teacher then

provided the answer and, at the end of the lesson, wrote out the three rules governing indirect object placement, and indicated which verbs fit each rule. The board was then erased, and the posttest administered.

Audio recordings were made of all groups and dyads and, for consistency, were transcribed for 10 minutes from the same starting point on all tapes, the reading of the first sentence. Negotiations were considered to consist of the following:

1. Clarification requests, made by listeners when they haven't understood (e.g., *What is question?* or *Which one is correct?)*
2. Confirmation checks, made by listeners when they believe they have understood but want to make sure (e.g., *Is it incorrect sentence?* or *Plan is indirect object?)*
3. Comprehension checks, made by the speaker to be certain that the listener has understood (e.g., *Do you have any questions?* or *Are you satisfied?)*
4. Repetitions, which in the data examined consisted largely of restatements of another's utterance as a type of confirmation check (e.g., *Correct?* or *Is it incorrect?)*
5. Request for repetition, which consisted of the listener's requests for the speaker to repeat a previous utterance (e.g., *Once more please* or *Please repeat)*

RESULTS

Statistical analyses of differences in pretest and control means were performed using a one-way analysis of variance (ANOVA). Paired *t* tests were used to examine the significance of differences between pre- and posttest scores and unpaired *t* tests for determining the significance of differences between posttest scores for the two treatments and between the treatments and the control group. A one-way chi-square test corrected for continuity (Hatch & Farhady, 1982) was used to examine the significance of any differences between negotiation counts for the groups and the dyads.

Statistical Analysis of Test Scores

Table 1 gives the pretest, posttest, and final test score means for all treatment groups and controls. Because of absences, the number of students taking the final test at the university differs from the initial number.

TABLE 1
Mean Scores for the Pretest, Posttest, and Final Test

	Pretest	Posttest	Final test	p^a
Women's jr. college (n=56)				
Task treatment (n=18)	14.39	18.94	16.33	<.05
Grammar treatment (n=18)	15.00	18.69	17.83	<.05
Control (n=20)	14.20	14.25	14.50	*ns*
University (n=34)				
Task treatment (n=12)	13.00	16.17	15.20 (n=10)	<.05
Grammar treatment (n=10)	12.70	19.10	16.83 (n=6)	<.05
Control (n=12)	12.17	12.42	11.63 (n=8)	*ns*

[a] Significance refers to paired t tests between the pre- and posttest means.

For students of the women's junior college, no significant difference existed among the three pretest means as determined by a one-way analysis of variance (ANOVA) (F [2, 53] = .5309, $p > .05$). However, differences were significant at the .05 level between the means of the scores of the pre- and posttest for the task treatment (paired $t = –6.497$, $df = 17$) and for the grammar treatment (paired $t = –4.535$, $df = 17$), as well as for both treatments' posttest scores compared with the control group posttest scores (task vs. control: unpaired $t = 6.926$, $df = 36$; grammar vs. control: unpaired $t = 7.644$, $df = 35$). It is notable that no significant difference was found to exist between the means of the posttest scores of the task treatment and the grammar treatment (unpaired $t = –.0574$, $df = 33$, $p = .570$). Regarding the final test given 2 weeks after the treatments, no significant difference existed between the scores of the posttest and the final test for the grammar treatment, indicating that no significant loss of proficiency occurred (paired $t = 1.399$, $df = 17$, $p = .180$). However, for the task treatment, the difference between the posttest and the final test 2 weeks later was significant (paired $t = 3.803$, $df = 17$, $p < .05$), indicating a loss of proficiency. Nevertheless, comparing the results of the final test to the initial pretest revealed a significant difference (paired $t = –2.475$, $df = 17$, $p < .05$) between the scores before performing the task and the proficiency remaining after 2 weeks. No significant difference was found among the control group scores (Hotelling's $T^2 = .19231$, $p = .590$).

For the students at the university, the one-way ANOVA indicated no significant difference existed among the three pretest means (F [2, 31] = .3345, $p > .05$), although there was a significant difference between the pretest scores for the basic non-English majors and the junior college intermediate-level English majors

(unpaired $t = 5.562$, $df = 4$, $p < .05$), indicating a significant difference in initial proficiency levels between the two groups of students. Differences were significant at the .05 level between the means of the pre- and posttest scores for both the task treatment (paired $t = –3.245$, $df = 11$) and the grammar treatment (paired $t = –8.552$, $df = 9$), as well as for both treatments' posttest scores compared with the control group's posttest scores (task vs. control: unpaired $t = 2.837$, $df = 22$, $p < .05$; grammar vs. control: unpaired $t = 5.094$, $df = 20$, $p < .05$). However, unlike the results obtained with the intermediate English majors, the immediate posttest scores for the grammar treatment were significantly higher than the task treatment's posttest scores (unpaired $t = –2.449$, $df = 20$, $p < .05$).

Comparison of the university students' final test scores with the previous scores was complicated by the fact that fewer students took the final test. For the task group, the difference between the pretest scores and the final test scores is not significant (paired $t = –2.141$, $df = 9$, $p = .061$) at the designated level of .05, although the value of p was quite close, at .061. Nor was the difference between the posttest and the final test significant (paired $t = 1.118$, $df = 9$, $p = .293$), suggesting that there was no substantial loss of proficiency between the two tests in the reduced sample. However, the differences between both pre- and posttest scores and the final test scores were significant for the grammar group (pretest vs. final test paired $t = –2.652$, $df = 5$, $p < .05$; posttest vs. final test paired $t = –2.936$, $df = 5$, $p < .05$), indicating a significant gain in proficiency over the pretest but a subsequent significant loss 2 weeks later. Again, no significant differences among the control group scores were observed (Hotelling's $T^2 = .3343$, $p = .774$).

Combining the task treatments from both schools gives a total of 20 students performing in groups of four, and 10 students in dyads. No significant difference (unpaired $t = 362$, $df = 28$, $p = .362$) was found between the posttest scores of students, regardless of whether they performed the task in groups or in dyads.

Quantitative and Qualitative Analysis of Negotiations

Table 2 gives the total negotiation frequencies in both English (the L2) and Japanese (the Ll), and the negotiation frequencies in English only during the 10-minute sample period for the groups and the dyads for the task treatment and the grammar treatment at each school.

No significant difference in the number of negotiations made by the two types of participation pattern was found for the total

TABLE 2
Total Negotiations and L2-Only Negotiations for the Different Participation Patterns

School	Grammar	Task-group (Total/L2 only)	Task-dyad (Total/L2 only)	Total/L2 only
Junior college (n=30)	0 (n=18)	49/48 (n=8)	33/32 (n=4)	82/80
University (n=22)	0 (n=10)	44/13 (n=8)	68/50 (n=4)	112/63

number of negotiations produced (x^2 = .253, df = 1, p > .05) or for negotiations in the L2 only (x^2 = 2.542, df = 1, p > .05) produced by the four groups and four dyads which were recorded performing the task. At the junior college three pairs and three dyads participated in the task, but because of mechanical failure the data for one group and one dyad could not be obtained.

Although the grammar treatment at the women's junior college was not taped, the teacher stated that no student made any comment, except to answer yes or no when asked if sentences were correct. The grammar treatment at the university was taped, and there was no audible participation by any student during the 20-minute presentation.

At the women's junior college, where the students were strongly requested to use only English during the task, negotiations were almost entirely in English, with only one negotiation each in Japanese for the groups and dyads respectively. At the university, however, where the teacher did not emphasize the need to use only English, the proportion of the L1 used was much higher, with 13 negotiations in English and 31 in Japanese for the groups, and 50 in English, and 18 in Japanese for the pairs.

Finally, the negotiations made by both groups were found to be qualitatively limited in either language and consisted of asking whether a sentence was correct or incorrect, asking for a repetition of a sentence, part of a sentence, or a single lexical item, or making a comprehension check, as in the following portions of protocols from a university group and dyad and a junior college group and dyad:

1. *University Group*
   A: Ready? (Student reads sentence and indicates correctness.)
   B: Yes. (Other three reply in unison.)
   A: Alright? (Student reads next sentence and indicates correctness.)
   B: Huh? (One student only.)
   A: (Student repeats sentence.) One more time?

2. *University Dyad*
   A: (Student reads sentence.)
   B: Correct?
   A: Yes. (Student reads next sentence.)
   B: Please again.
   A: (Student repeats sentence.)
   B: Incorrect?
   A: Yes. (Student reads next sentence.)
   B: Correct?
   A: Yes.
3. *Junior College Group*
   A: (Student reads sentence.) Are you satisfied?
   B: Yes. (Other three reply in unison.)
   A: (Student reads next sentence.) Do you have any questions?
   B: No. (Other three reply in unison.)
   A: Ready?
   B: Yes. (Other three reply in unison.)
   A: (Student reads next sentence.)
4. *Junior College Dyad*
   A: (Student reads sentence and indicates correctness.)
   B: Is it the answer?
   A: Yes. (Student reads next sentence but does not indicate correctness.)
   B: Is it correct?
   A: Yes.
   B: Once more please.
   A: (Student repeats sentence.)
   B: *Plan* is indirect object?
   A: Yes.

DISCUSSION

The Grammar Task Versus the Grammar Lesson

The first research question was to compare the effectiveness of a task approach and a traditional grammar lesson. It should be recalled that no discussion of the grammar point was included in the task treatment except for what was written on the task sheet and task cards, and the students' mastery of the form was gained solely from performance of the task activity. For the English majors at the

junior college, the initial mean percent correct for the task group was 72%, which increased to 95% on the posttest, but then dropped to 82% after 2 weeks—a net gain in proficiency of 10% In comparison, the grammar-lesson students started at an initial mean of 75% correct, increased to 93% correct after the lesson, then decreased to 89% correct after 2 weeks—a net gain of 14% Whereas the posttest scores of the task group and the grammar-lesson group were not statistically different, the task group's mean score on the final test 2 weeks later was significantly lower than the mean score for the grammar-lesson group but was nonetheless significantly higher than the pretest score. Thus, the task appeared to have functioned equally well as the grammar lesson in the short term, and was only slightly less effective in maintaining proficiency than the grammar lesson after 2 weeks.

For the basic level non-English-major university students, the initial mean percent correct for the task group was 65%, which increased significantly to 81% after the task, a gain of 16%. In comparison, the grammar group's initial mean percent correct was 64%, which increased to 96% after the grammar lesson, a gain of 32%. Thus, the gain in proficiency for the task group was significantly lower than for the grammar-lesson group. Longer-term proficiency gains for the grammar-lesson group were more difficult to assess due to absences. However, even with fewer students, the grammar-lesson group showed maintenance of significant proficiency gains after 2 weeks although the task-treatment group did not.

Two possible explanations for the less successful results of task performance by this group compared with the English majors are (a) their lack of familiarity with pair/group work, and (b) as indicated by their comments and questions during the audio recordings, their imperfect understanding of the goals and procedures of the task, information which was presented to them in the second language. It is likely, therefore, that proficiency gains would have been higher with a more detailed explanation of the requirements of the task and previous experience in the pair/group-participation pattern. In contrast, the English majors were familiar with both group work and performance of information-gap activities, and were better able to comprehend the L2 explanation of the task procedures. Furthermore, to prevent the teacher from giving formal instruction about the grammar point, neither group of students received subsequent teacher feedback on the success of their task work. It is reasonable to assume that the learning of explicit knowledge of a grammatical point gained from performance of a grammar task would have been enhanced by feedback concerning how effectively the groups had performed the task.

As a final consideration, it is necessary to establish that there was no practice effect occurring when students took the same test three times, otherwise the favorable results reported above regarding the effectiveness of the grammar task in promoting proficiency gains are questionable. First, the three test scores for the control groups at both the junior college and the university showed no significant variation and, furthermore, the final control group score was actually lower than the previous two control group scores at the university. Second, a similar observed decline in test scores for both task and grammar treatment groups after 2 weeks is additional evidence against the operation of a practice effect.

The Grammar Task and Negotiation of Meaning

One of the problems with communicative task performance is that there have been few ways other than analysis of the nature and frequency of interactions to determine the success of the task in promoting language acquisition. On the other hand, with task types such as listening or reading comprehension tasks, the students' performance in answering questions serves as a criterion-referenced indication of the success of the task. With the type of grammar task presented here, it is possible to test for proficiency gains as well as to analyze the quality and quantity of interactions. The second research question addressed in this report concerned the amount and qualitative nature of the negotiations produced by the grammar task.

The mean number of negotiations reported in a previous study (Doughty & Pica, 1986) for performance of a two-way information-gap task during a 10-minute period by three groups and three dyads combined was 64. With the grammar task investigated here, the total negotiation counts in both the L1 and L2 for groups and dyads combined were 82 for the junior college and 112 for the university. If the L1 negotiations are removed, the L2 negotiation counts come to a combined total of 80 for the junior college groups and dyads (2 of each) and 63 for the university groups and dyads (2 of each). These figures are similar to the average combined count of 64 previously reported for a greater number of groups/dyads. Thus, the grammar task appears to promote similar amounts of interaction in the limited data presented here.

The Doughty and Pica study also hypothesized that in terms of amounts of negotiations, dyads would produce the most, followed by groups, with the least being produced by the teacher-fronted activity. This pattern was found in the data here, with no negotiations produced by the grammar lesson, 61 L2 negotiations

produced by the four groups, and 82 L2 negotiations produced by the four dyads. Although the difference between these figures was not significant, the pattern for the junior college indicates that groups produced more negotiations than dyads.

Why did the university dyads produce so many more negotiations? Examination of the transcripts shows that the extra negotiations made by the university dyads consisted of the words correct or incorrect asked as confirmation checks, as shown in Example 2. In the junior college transcripts, such usages were rare since the sentence readers usually supplied the information to the students who were listening. As mentioned, task performance took 20 minutes, with the bulk of the negotiations occurring during the first 10 minutes, during the reading of the sentences. During the final portion, the students were relatively quiet as they wrote out the rules and did not discuss the nature of the rules in either language. An explanation for this finding is that the students, who had 6 years of previous English study, already possessed explicit linguistic knowledge of the general grammar rules governing English dative alternation and did not have to consult on this point.

We must now consider the broader question of whether the qualitatively limited negotiations observed during performance of this task can be regarded as requests for modifications of input in the sense that the term has been used (see Doughty & Pica, 1986; Long, 1983; Pica & Doughty, 1985; & Pica, Holliday, et al., 1989). Again, it must be recalled that the students heard their interlocutors read a sentence, and requested one of the following (a) a repetition of the sentence or of part of the sentence, (b) an explanation of an unfamiliar word in the sentence, or (c) confirmation of their guess that the sentence was either correct or incorrect. Their interlocutors' response options were to read the sentence again, say yes or no regarding correctness, repeat the questionable word, or make their own comprehension checks, as shown in Examples 1 and 3. Since the students reading the sentences were not originating the language, there was no modification of output taking place and the language was "planned to the extreme. Does such limited discourse represent negotiated interaction?

We suggest that the exchanges observed during performance of this particular grammar task are within the limits of the construct because the focus of the interaction was on meaning, and the negotiations performed were essential to the comprehension of meaning. However, it is clear that a more detailed investigation of the qualitative nature of the negotiated interactions promoted by different types of grammar tasks in different settings is an important future research question. This point is equally true with

regard to the quality of interactions analyzed in other published studies. Further research in the area of negotiated interaction must deal with qualitative aspects of the data, particularly in situations where both interlocutors are nonnative speakers.

CONCLUSIONS

This report has presented the case for use of a particular type of language learning task—one which encourages communication about grammar. It has been argued that grammar tasks may contribute to L2 acquisition in two ways. They may contribute *directly* by providing opportunities for the kind of communication which is believed to promote the acquisition of implicit knowledge, and they may also contribute *indirectly* by enabling learners to develop explicit knowledge of L2 rules which will later facilitate the acquisition of implicit knowledge.

The results of the exploratory study reported in the previous section lend some support to these claims. This study demonstrated that Japanese EFL learners at the college level were able to increase their knowledge of a difficult L2 rule by completing a grammar task. It also showed that the interaction which resulted from grammar task performance was characterized by a similar quantity of conversational modifications to those reported to occur in other two-way information-gap tasks. The learners performing this task had the opportunity to learn about grammar while taking part in communication centered on an exchange of information.

In addition, a number of considerations have been raised. First, the grammar task used here did not result in the same level of longer-term learning as did the traditional, teacher-fronted grammar lesson. As mentioned, possible reasons for this may have been the learners' lack of experience in working in small groups and the absence of teacher feedback on their solution to the task. Clearly, though, it is important to establish that group work discussion can result in development of explicit knowledge, given the importance attached to this in the theoretical framework illustrated in Figure 1. We still need to discover whether and how group work can be made as effective as teacher-directed explanations in developing explicit knowledge.

Second, although the grammar task produced a large number of interfactional modifications, the nature of the exchanges which took place was rather mechanical, as the examples given in the previous section illustrate. This leads us to ask whether it is the quantity of speech modifications that is important for acquiring implicit knowledge, as suggested by the interaction hypothesis, or whether

it is some yet undetermined qualitative aspects of communication which are more important (see Bygate, 1988). The nature of the relationship between interaction and language acquisition is still poorly understood.

Grammar tasks which emphasize consciousness-raising rather than practice appear to be an effective type of classroom activity, and their use is supported by what is currently known about the way a second language is acquired. Furthermore, such tasks provide serious content, in contrast to the trivial content of many information-gap activities, and they accommodate learners who believe that it is important to learn about grammar. They provide opportunities to communicate in the L2 in groups or pairs, and they encourage an active, discovery-oriented approach on the part of the learners, which accords with current views about good educational practice. It should be possible to develop a wide range of grammar tasks, including those which focus on sociolinguistic learning problems as well as on purely formal problems, such as the grammatical feature which was the focus of this study. Hopefully, through experimentation, it will be possible to devise tasks which result in interaction which is qualitatively richer than what we obtained in this exploratory study.

Grammar tasks also have their limitations, however. Some learners may not wish to talk about grammar. They may find it a boring topic, or they may find it difficult to discuss because they lack the basic metalinguistics knowledge needed to do so. Learners may resort extensively to the use of their first language during a grammar task, as is seen in the case of the university group in this study. However, with training in task performance, it should be possible to overcome this limitation, as the English language majors at the junior college demonstrated. It is also possible that grammar tasks are less suitable for beginners, partly because such learners are not able to talk in the second language, and partly because grammar as a discussion topic is less appropriate at this level. In general, we suggest that grammar tasks seem best suited for intermediate/advanced learners who are motivated to study grammar as subject matter.

The use of communicative, problem-solving grammar tasks remains an intriguing proposal in need of further study. Future research will need to address a number of issues. These include (a) developing different formats for grammar tasks, (b) examining the effect of these different formats on the quality and quantity of interaction, (c) examining the effect of the different formats on gains in explicit knowledge, (d) investigating the effect of teacher feedback on the learner's solutions to grammar tasks on learning,

and (e) investigating the role of metalinguistics knowledge in task performance.

ACKNOWLEDGMENTS

Versions of this paper were presented by the authors at the 16th Annual Conference of the Japan Association of Language Teachers, November 1990 in Tokyo. The authors would like to thank two anonymous *TESOL Quarterly* reviewers for their helpful comments on an earlier draft of this article.

THE AUTHORS

Sandra Fotos is Professor of English at Kogakuin University, Tokyo. Her research interests include the role of formal instruction in second language acquisition in the EFL setting.

Rod Ellis, Professor of Applied Linguistics at Temple University Japan, previously taught in Zambia and in London at St. Mary's College and Ealing College of Higher Education. He is author of a number of books and articles in the fields of second language acquisition research and teacher training.

REFERENCES

Allen, P., Swain, M., Harley, B., & Cummins, J. (1990). Aspects of classroom treatment: Toward a more comprehensive view of second language education. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of second language proficiency* (pp. 57-81). Cambridge: Cambridge University Press.

Bialystock, E. (1981). The role of linguistic knowledge in second language use. *Studies in Second Language Acquisition, 4,* 31-45.

Brumfit, C. (1984). *Communicative methodology in language teaching.* Cambridge: Cambridge University Press.

Bygate, M. (1988). Units of oral expression and language learning in small groups. *Applied Linguistics, 9,* 59-82.

Dickens, P., & Woods, E. (1988). Some criteria for the development of communicative grammar tasks. *TESOL Quarterly, 22* (3), 623-646.

Doughty, C., & Pica, T. (1986). "Information gap" tasks: Do they facilitate second language acquisition? *TESOL Quarterly, 20* (2), 305-325.

Ellis, R. (1984). Can syntax be taught? A study of the effects of formal instruction on the acquisition of *Wh*-questions by children. *Applied Linguistics, 5,* 138-155.

Ellis, R. (1990). *Instructed second language learning.* Oxford: Basil Blackwell.

Hammerly, H. (1987). The immersion approach: Litmus test of second language acquisition through classroom communication. *Modern Language Journal, 19,* 331-359.

Hatch, E., & Farhady, H. (1982). *Research design and statistics for applied linguistics.* Rowley, MA: Newbury House.

Kadia, K. (1988). The effect of formal instruction on monitored and spontaneous naturalistic interlanguage performance. *TESOL Quarterly, 22* (3), 509-519.

Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon Press.

Krashen, S. (1985). *The input hypothesis.* London: Longman.

Lightbown, P., Spada, N., & Wallance, R. (1980). Some effects of instruction on child and adolescent ESL learners. In R. Scarcella & S. Krashen (Eds.), *First and second language acquisition processes.* Rowley, MA: Newbury House.

Long, M. (1983). Native speaker / non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4,* 126-141.

Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115-141). Rowley, MA: Newbury House.

Long, M. (1989). Task, group and task-group interactions. *University of Hawai'i Working Papers in ESL, 8,* 1-26.

Long, M., & Porter, P. (1985). Group work, interlanguage talk and second language acquisition. *TESOL Quarterly, 19* (2), 207-228.

Nunan, D. (1989). *Designing tasks for the communicative classroom.* Cambridge: Cambridge University Press.

Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning, 33,* 465-499.

Pica, T., & Doughty, C. (1985). The role of group work in classroom second language acquisition. *Studies in Second Language Acquisition, 7,* 233-248.

Pica, T., Holliday, L., Lewin, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition, 11,* 63-90.

Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6,* 186-214.

Porter, P. (1986). How learners talk to each other: Input and interaction in task-centered discussions. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 200-222). Rowley, MA: Newbury House.

Rulon, K., & McCreary, J. (1986). Negotiation of content: Teacher-fronted and small-group interaction. In R. Day (Ed.), *Talking to learn: Conversation* in *second language acquisition* (pp. 182-199). Rowley, MA: Newbury House.

Schumann, J. (1978). *The pidginiziztion process: A model for second language acquisition.* Rowley, MA: Newbury House.

Sharwood Smith, M. (1981). Consciousness raising and the second language learner. *Applied Linguistics, 2*, 159-168.
Spada, N. (1987). Relationships between instructional differences and learning outcomes A process-product study of communicative language teaching. *Applied Linguistics, 8*, 137-155.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
Ur, P. (1988). *Grammar practice activities.* Cambridge: Cambridge University Press.
White, L. (1987). Against comprehensible input: The input hypothesis and the development of second language competence. *Applied Linguistics, 8*, 95-110.
Zobl, H. (1985). Grammars in search of input and intake. In S. Gass & C. Madden (Eds.), *Input in second language acquisition.* Rowley, MA: Newbury House.

APPENDIX A
Task Cards

Students in groups of 4—one different card to each member
Students in pairs—two different cards to each member

1. Correct: I asked my friend a question.
1. Incorrect: She asked a question to her mother.
2. Correct: Kimiko reviewed the lesson for John.
2. Incorrect: Kimiko reviewed John the lesson.
3. Correct: The teacher calculated the answers for the students.
3. Incorrect: The teacher calculated the students the answers.
4. Correct: The secretary reported the problem to her boss.
4. Incorrect: The student reported the teacher the matter.
5. Correct: I offered her a cup of tea.
5. Correct: I offered a cup of tea to the president.
6. Correct: The teacher pronounced the difficult word for the class.
6. Incorrect: The teacher pronounced the class the difficult word.
7. Correct I bought many presents for my family.
7. Correct: I bought my family several presents.
8. Correct: She cooked a delicious dinner for us.
8. Correct She cooked us a wonderful meal.
9. Correct: She suggested a plan to me.
9. Incorrect: She suggested me a good restaurant.
10. Correct: The teacher repeated the question for the student.
10. Incorrect The teacher repeated the student the question.

APPENDIX B
Task Sheets

There are some verbs in English which can have two objects. One of the objects is called the *direct object*. The other is called the indirect object. An indirect object names the person for whom the action of the verb is performed:

		indirect object	direct object
She	**wrote**	**Susan**	**a letter.**

Different verbs may have the objects in different order, and this is often a problem for students of English. The following exercise will help you understand some confusing verbs.

Directions:
In groups, you are to study correct and incorrect sentences using different verbs. You all have different sentences. You must*read* your sentences to the rest of the group.*Do not show your sentences to the other members! Only read the sentences as many times as necessary!* Work together as a group and decide on the basis of the correct and incorrect sentences where the direct and indirect objects should be located. Fill out the rest of this page. Choose one student to report your results to the rest of the class.**Please speak only in English during this exercise!!**

Verbs:	Possible correct order of direct and indirect object
1.	
2. reviewed:	
3. calculated:	
4. reported:	
5. offered:	
6. pronounced:	
7. bought:	
8. cooked:	
9. suggested:	
10. repeated:	

Conclusion: Write 3 rules concerning the possible order of objects

Rule 1: _____
verbs which follow this rule _____

Rule 2: _____
verbs which follow this mle _____

_____
verbs which follow this rule _____

APPENDIX C
Grammaticality Judgment Test on Dative Alternation

Directions: Read the sentences. Decide if they are corrector incorrect.
Write (0) if correct, or (X) if incorrect.

1. _____ She asked the class a question.
2. _____ She asked a question to the class.
3. _____ She reviewed the sentences for Mary.
4. _____ She reviewed Mary the sentences.
5. _____ She calculated John the math problem.
6. _____ She calculated the math problem for John.
7. _____ She reported the police the problem.
8. _____ She reported the problem to the police.
9. _____ She offered her friend a chocolate.
10. _____ She offered a chocolate to her friend.
11. _____ She pronounced the difficult word for me.
12. _____ She pronounced me the difficult word.
13. _____ She bought her friend a dress.
14. _____ She bought a dress for her friend.
15. _____ They cooked a meal for their friends.
16. _____ They cooked their friends a meal.
17. _____ They suggested the children an idea.
18. _____ They suggested an idea to the children.
19. _____ She repeated the word for me.
20. _____ She repeated me the word.

TESOL QUARTERLY, Vol. 25, No. 4, Winter 1991

Building a Vocabulary Througb Academic Reading

KATE PARRY
Hunter College, The City University of New York

This paper reports a series of longitudinal case studies designed to address the question of how language learners build their vocabularies. Students who were enrolled in an anthropology class were asked to record the words that caused them difficulty as they read their anthropology texts, and to write down, if they could, what they thought the words meant. The resulting lists are analyzed in terms of the kinds of words listed, the accuracy of the glosses, and the probable reasons for misinterpretation; the analysis is considered in relation to data collected in protocols and a translation task. The conclusions are that a range of strategies may be used for learning vocabulary, each involving liabilities as well as assets. Students need to be aware of the range so as to develop flexibility in their responses to unfamiliar words.

The concept of process is by now well established in language research. In the field of writing, Emig (1971) and Perl (1979) initiated a tradition of looking at how people write rather than what they write, and Zamel (1983) and Raimes (1988) have carried this tradition into work with ESL students. The process approach is also increasingly affecting the field of reading, even though reading, as an internal activity, is more difficult to observe. Researchers like Hill and Larsen (1983), using interview data, and Block (1986), using protocols, have demonstrated that it is possible to reconstruct at least something of what goes on in readers' minds, and that how readers arrive at their interpretations is at least as important as what those interpretations are. Such process-oriented work has proved particularly valuable for those of us who teach reading and writing, for the knowledge of what people actually do with text provides a basis from which we can advise our own students as they too struggle to construct and reconstruct meaning.

One process, however, that has been neglected in second language research is that of vocabulary building (Carter, 1987; Levenston, 1979; Meara, 1980, 1983). It is true that a number of

significant studies have recently been done on how second language learners infer word meanings from context (Faerch, Haastrup, & Phillipson, 1984; Haynes & Baker, in press; Huckin & Bloch, in press; Palmberg, 1987), but the question of how those inferences help towards building up a native-like vocabulary has rarely been considered. Consequently, we have little basis, other than our own experience of what works, from which to develop an approach to the teaching and learning of words.

The process is, of course, a difficult one to examine, for it can only occur over a long period of time, and it must be extremely complicated. As Aitchison (1987) shows, the "mental lexicon," contains an immense amount of information, arranged in such a way that it can be checked through with astonishing rapidity. The sheer number of words that most people know is probably enormous: Aitchison estimates, on the basis of various studies that she reviews, that the vocabulary of "an educated adult . . . is unlikely to be less than 55,000 [words] and may be as high as 250,000" (p. 7). Furthermore, Aitchison, citing Fodor (1981), points out that native speakers know a great deal about these words, in terms of both their contexts of use and their exact range of meaning—much more than can be expressed in a dictionary definition. Again, to judge from experiments in word recognition and association and from the mistakes made in aphasia as well as in ordinary slips of the tongue, the mental lexicon must be arranged in a complex network of relationships, quite unlike the simple alphabetical listing that we are accustomed to in dictionaries. How can learners of a second language build up such a large and complicated structure in the few years, or even semesters, in which many of them hope to do it?

We who teach English as a second or foreign language must acknowledge that little vocabulary building gets done through our own direct agency, for we do not cover a wide range of words in class, and many of us spend no time at all on discussing collocations and semantic relationships. We seem, rather, to act on the assumption that if we teach our students grammar, and reading and writing skills, they will build their vocabularies on their own as they engage in other activities; and that this will happen particularly in the context of academic reading. In this paper I report a series of studies through which I tried to find out something about how this may be done.

THE DATA

The research was conducted at Hunter College, which is a four-year liberal arts institution offering bachelor's and master's degrees

within the City University of New York. The college has a large and diverse student population; of its 18,000 students, about 40% speak English as a second language. These ESL students are placed in reading and writing courses on the basis of tests taken on admission and then, like all students, before they can embark on their majors, they must take a range of introductory courses that make up a core curriculum. (Prior to 1988, students would often take these introductory courses concurrently with, or even before, their ESL ones. In that year, however, Hunter College instituted a registration system that required them to pass at least the lower and intermediate levels of ESL before moving into academic subjects at the college level.) Most of these introductory courses have large sections with upwards of a hundred students in each, so the teaching is usually confined to lectures and the assigned work is mainly reading in standard college textbooks. There is thus little opportunity for teachers to check how well the students understand the material, and still less for the students to get linguistic help.

The area chosen for investigation was anthropology, on the grounds that this subject would produce specialized vocabulary that could be expressed in the students' first languages without the use of loan words. Sections of the introductory anthropology course at Hunter College were surveyed to identify students who were still enrolled in, or had only recently completed, the College's ESL sequence and who had received at least a secondary education in their own country and in their own language. These students were given a preliminary vocabulary test (taken from the Michigan Test of English Language Proficiency) and then were asked, as they read the assignments in their anthropology textbooks, to keep lists of any words they encountered that "cause[d them] any difficulty"; they were also asked to write down what they guessed these words to mean and, if they chose to look the words up in the dictionary, to record the definitions that they found there. They were encouraged to express their guessed meanings as precisely as possible, using their first language for the purpose if they found it easier. The glosses were then compared with the original words, with the help of informants who had native or native-like proficiency in both the students' first languages and in English.

Four such studies have now been completed. The first was with a Romanian student, Viviana,[1] in the fall of 1986; the second was with a Japanese, Yuko, a year later; and the third and fourth were with a Greek Cypriot, Dimitri, and a Korean, Ae Young, in the fall of 1988. (Viviana, Dimitri, and Ae Young had all completed their

1 Student names are pseudonyms.

last ESL course the previous term, while Yuko had just begun the three-level sequence; she was also enrolled in the upper level of the College's two developmental reading courses.) By 1988 I had received a research grant, so I was able to expand the research design in Dimitri's and Ae Young's cases to include, in addition to the lists, think-aloud protocols of each student constructing a list, and a translation task that was based on part of the protocol text; these two students were also, as was Yuko, given a posttest at the end of the semester on the words they had identified. The texts used were simply those assigned for the course—which, because of changes within the Anthropology Department, turned out to be different each year. Thus Viviana's lists were based on *Cultural Anthropology* by Ember and Ember (1985), Yuko's on *Cultural Anthropology* by Plog and Bates (1980), and Dimitri's and Ae Young's on *Cultural Anthropology: A Contemporary Perspective* by Keesing (1981). Consequently the four studies cannot be compared on matters of detail; but certain general patterns can be seen, and Dimitri and Ae Young provide more specifically comparable data. This is particularly true of the protocol, for which, in order to ensure that it would be equally unfamiliar to both students, I chose another, slightly more specialized, anthropology text; it was an excerpt from an article by Sudarkasa (1982), entitled "African and Afro-American family structure."

The findings on Viviana and Yuko have been presented elsewhere (Parry, 1987, in press), so this article will discuss them only in general terms; but it will examine in detail the newer and more easily compared information provided by Dimitri and Ae Young. Also, because the four studies have produced an enormous quantity of data, no attempt will be made to cover it all. Instead, just three questions will be addressed: Which words did these students actually identify as difficult? How, and how successfully, did they infer their meanings? What factors seem to have led specifically to misinterpretation?

WORDS IDENTIFIED AS DIFFICULT

First, let us consider how many words the students wrote down in their lists, and what proportion these represented of all the words read. The figures are given in Table 1. It is only possible to give a rough estimate of the total number of words read, for there is no guarantee that the students read all the words on each page, and the number of words per page varies considerably; but even this

TABLE 1
Numbers of Words Listed by Each Student as Proportion of Total Read

Student	No. pages read	No. words read	No. words listed	% listed
Viviana	99	55,000?	66	0.12%?
Yuko	93	48,000?	168	0.35%?
Dimitri	182	72,000?	91	0.13%?
Ae Young	11	7,500?	119	1.60%?

rough estimate shows clearly that the proportion of words listed was tiny. This does not, of course, mean that the students knew all the words they did not record—indeed, there are several cases where a word recorded on a particular page appears, unrecorded, on a page that was read previously. It is more likely that the students read holistically much of the time and only stopped to work on particular words when they felt they were losing their grip on the text.

The tendency to skate over unfamiliar words may, however, be more marked in some students than in others. Another striking point about Table 1 is that Ae Young both read much less than did all the others and recorded proportionately many more words. The book she was reading (Keesing, 1981) certainly seems to be more difficult than the other two texts, in that it puts a greater emphasis on theory and lays out the factual information in a less systematic way; but this does not account for the difference between her and Dimitri. The explanation may, of course, be that Dimitri simply knew more words; yet he scored slightly less than she did on the Michigan Test, getting 58% against her 63%. (These figures should be interpreted cautiously since the validity of testing vocabulary in this way has been questioned. See Bright & McGregor, 1971). Another explanation—perhaps an additional rather than an alternative one—may lie in their styles of approaching the task, which the protocols showed to be very different.

When making the protocol, Dimitri read through a whole passage (which is six pages long and contains about 1800 words) quite rapidly, stopping only to mark each difficult word as he came across it. Then he read through the whole passage again, this time writing each word on his list, guessing its meaning, and writing the guess down, usually in English (he identified 11 words, and wrote down a guessed meaning for 9). The proportion of words he identified (0.6%) is higher than that recorded for his lists as a whole. There are at least two factors that could account for this: As noted above, there is no assurance that he read all of every page that he

listed, especially since there are many boxed case studies in Keesing's book that he may or may not have skipped over; second, in doing the work in the presence of an observer, he was undoubtedly self-conscious, and this may have made him less confident of his knowledge. This is an obvious methodological problem in studies that use protocols.

After writing a word down, guessing its meaning, and entering his gloss, Dimitri looked it up in the dictionary before going on. He made his guesses rapidly, without much explicit looking at the preceding text, and without articulating cues in the subsequent text at all; and in a couple of instances he simply gave up without a struggle, leaving the space for guessed meaning blank. Then he took rather a long time to find the word in the dictionary (he had difficulty with alphabetical order) and wrote down the definition he found there without, in most cases, referring back to the context.

Ae Young, on the other hand, only read the text once, and stopped at each unfamiliar word. She then spent a long time going back over the previous two or three sentences and sometimes going forward as well; on one occasion she translated into Korean, although on her list she expressed only three of her 23 guesses in that language. Also, in the two cases where she wrote down no guess, she worked on the word and its context for a long time before giving up. Finally, like Dimitri, she looked up each of her words in the dictionary. She was quicker at finding the place (suggesting, perhaps, that she uses the dictionary more often) but then, when she found it, spent time not only writing the definition down but also checking how it fitted into the context. Altogether, she worked more meticulously and analytically, with the result that she took 2 hours to do the whole protocol, whereas Dimitri took only 1 hour. It will also be noted that in the protocol, as in her other lists, Ae Young recorded a relatively high proportion (1.4%) of all the words read.

The fact that Ae Young took so long on the protocol passage also suggests why she read so little altogether. She began the semester by reading very slowly and recording many words; but soon the lists stopped coming in, and she said it was because she found she could get all the information from the lectures, and it was not worth spending the time it took to read the text. If this is a general pattern for her, it further explains why she seemed to know fewer words than the others: Since reading takes her so long, she is discouraged from doing much of it, and consequently she is exposed to relatively little vocabulary.

This supposition is further supported by another feature that distinguishes Ae Young's lists from the others': the proportion of

words listed that are of relatively high frequency. This was determined by checking each of the words listed by each student against the frequency counts published by Kucera & Francis (1967) and Hofland & Johansson (1982). Each of these counts shows how many times a particular string of letters occurs in a corpus of a million running words, the corpus being drawn from 500 texts in 15 subject areas. In order to compare the students' lists, I defined as *relatively frequent* any word that appears 10 or more times in both of these two corpora, adding together in this tally the singular and plural forms of nouns and all the finite and nonfinite forms of verbs. Then I identified in each student's lists the words that were of relatively high frequency, and obtained the results given in Table 2.

TABLE 2
Relatively Frequent Words Listed by Students

Student	No.	% of all words listed
Viviana	1	2
Yuko	10	6
Dimitri	8	9
Ae Young	36	31

These figures show that for all the students most of the words that caused trouble were ones that occur very infrequently; but Ae Young recorded a strikingly higher proportion of relatively frequent ones than did any of the others. This makes sense if she is reading less, even if that reading is done more carefully, for she will presumably have come across fewer words altogether, and will have had relatively few encounters with even the more frequent ones.

We should not, however, overemphasize this point, for even the words I have classified as relatively frequent cannot really be considered common: Once in 100,000 is not a high rate of occurrence. The point brings home the major difficulty that advanced ESL students face, namely, that the further they progress the more they will have to learn large numbers of words that they will come across only rarely.

The lists provided by the four students also give some interesting indications of what kinds of words these infrequent items might be. One might expect them to include a high proportion of specialized vocabulary, specific to the field of anthropology; and this would make the problem of learning them less, since within that field they would appear relatively often. A glance at the texts shows that such

words do indeed occur, but to judge from the students' lists, they were not the primary cause of difficulty. Indeed, the number of anthropology-specific words is so small that it is possible to list them all individually. They are presented in Table 3 under the names of the students who recorded them:

TABLE 3
Words Listed by Each Student That Are Specific to Anthropology

Viviana	Yuko		Dimitri	Ae Young
betrothal	avunculocal	pastoralism	hominid	kinship
endogamous	emic	patrilineal	potlatch	lineage
kinship	etic	polyandry	sedentarism	patrilineal
	exogamous	polygyny	shaman	primate
	lineage	sibling	sorcery	stratification
	matrilineage	stratification		

As can be seen, my interpretation of the category anthropology-specific has been quite liberal: Words like *primate* and *stratification* are likely also to be found in other fields.

In addition to the particularly anthropological terms listed above, the students recorded a number of terms that are specific not so much to anthropology as to the ways of life that anthropologists describe—ways of life that differ markedly from any followed in New York City. Table 4 gives a list of such words; again, the number is small:

TABLE 4
Words Relating to Ways of Life Described in Anthropology

Viviana	Yuko	Dimitri	Ae Young
corralling	antelope	arboreal	arboreal
fallow	calipee	carnivore	betel
graze	fallow	fowl	foraged
machetes	forage	granary	peasant
millet	glean	hoofed	
stubble	harpoon	maize	
	herbivorous	ochre	
	herding		
	sleds		

Altogether, then, only a small proportion of the words identified by the students can be described as specialized: The two categories listed above account for about 20% of Viviana's, Yuko's, and Dimitri's lists, and only 12% of Ae Young's.

By contrast, a much larger proportion of each student's words is what Martin (1989) has described as "bridging" vocabulary—general items that characterize formal prose and often express relationships or other abstractions.[2] Words of this type account for more than 40% of Viviana's, Yuko's, and Dimitri's lists, and over 50% of Ae Young's. The lists of these words are too long to give in full, but Table 5 shows the first 10 recorded by each student:

TABLE 5
The First 10 General Words Recorded by Each Student

Viviana	Yuko	Dimitri	Ae Young
ensuing	vitally	rudimentary	circuitous
supplant	lurid	repertoire	viable
autonomous	expertise	depiction	cumulation
tenuous	emissions	vantage	folly
stipulating	refinement	dexterity	accumulation
prevalent	arbitrary	swirl	diverse
augmented	interlock	meager	crucial
fluctuate	askance	glimpse	collective
encroachment	condone	perishable	impoverished
reverts	unsubstantiated	tinker	connotation

The predominance of words such as these suggests that where students most need to build up their vocabularies is not in any particular subject matter but in the register of formal expository prose.

While the students' lists show some common pattern in the type of words they include, it is remarkable, when we look at particular items, how little overlap there is between them. It has not been possible to check the overlap between Viviana's and Yuko's lists because they are not on a computer database, but all the other lists have been compared, with the results given in Table 6.

There was no word identified by more than two students, despite the fact that they were all working in the same subject area. Even Dimitri and Ae Young, who were actually reading the same book, were not apparently aware of the same problems; indeed, the highest rate of overlap is between AE Young and Yuko. The

2 It has been suggested tome that this kind of vocabulary is what Trimble (1965), following Cowan (1974), characterizes as "sub-technical." I think not, however, for Cowan defines sub-technical vocabulary as "context-independent words which occur with high frequency across disciplines" (p. 391) and Trimble adds a further set, defined as "those 'common' words that occur with special meanings in specific Scientific and technical fields" (p. 129). The words I have in mind correspond to neither of these descriptions, for, while they do occur across disciplines, it is not with high frequeney, nor are they used in anthropology (or any other scientific field) with meanings that are "special."

TABLE 6
Amount of Overlap Between Students' Lists

Student	Same words	Related words
Dimitri/Yuko	3	4
Dimitri/Ae Young	5	1
Dimitri/Viviana	2	0
Ae Young/Yuko	8	1
Ae Young/Viviana	2	3

pedagogical problem is obvious: If the items that cause difficulty are different for each student, how is vocabulary to be approached in class? The implication is that we should be teaching not specific words but strategies and principles for inferring word meanings; so now let us consider the students' inferences.

GUESSING MEANINGS FOR DIFFICULT WORDS

The first point to come out of the students' lists of inferences is that all four were very willing to attribute meaning to the words that they did not know. Table 7 shows the number of instances in which each student left a blank in the guessed meaning column of the list:

TABLE 7
Words for Which Students Wrote No Gloss

Student	No.	% of all words listed
Viviana	13	20
Yuko	18	11
Dimitri	8	9
Ae Young	3	3

Such figures suggest that an unfamiliar word is rarely seen as a hole in the fabric of the text—it is assumed to have meaning, and that meaning is derived from the context. Moreover, an analysis of the guessed meanings in relation to the original words indicates that much of what the students infer is in fact correct, or at least partially so. The method of classifying the students' guesses is discussed in detail elsewhere (Parry, in press), so here I shall simply summarize, A *correct guess* is one which my informant characterized as a good translation of the word in question; or, when the guess was given in English (as most of Dimitri's and a large proportion of Ae Young's

were), it is one which I judged to be either a synonym or a good definition of the word. An *incorrect* guess is one which, when decontextualized and compared with the original word, seems to bear no relation to it—beyond one of a most general kind such as their both being abstract nouns; antonyms were also included in this category. A *partly correct* guess always does bear some relationship to the original, such as one of superordination, or belonging to the same superordinate category, or sharing one or more semantic features; the gloss and the original word therefore overlap but do not coincide in meaning. Table 8 shows how many and what proportion of each student's guesses fall into each category.

TABLE 8
Classification of Glosses by Relationship to Original Words

Student	Correct	Incorrect	Partly correct
Viviana	11 (17%)	8 (12%)	34 (52%)
Yuko	55 (33%)	33 (20%)	60 (36%)
Dimitci	20 (22%)	27 (30%)	35 (39%)
Ae Young	14 (12%)	58 (49%)	44 (37%)

Note. Percentages are of total numbers of words listed, not of words glossed.

There is, of course, a problem in presenting the data in this way, for the figures suggest a greater degree of accuracy than it was possible to obtain. Word meanings are by their nature indeterminate, and they depend, at least to some extent, on the perceptions of individual language users. Thus the translations of the first language glosses might be questioned by other translators, although my informants tried to correct for this by consulting dictionaries or other speakers of their languages in cases where they were uncertain. My own classification of the glosses is open to question too, for obviously the categories shade into each other, and there is unlikely to be agreement between any two assessors on every decision. This difficulty was obviated, in the case of Yuko's data, by publishing the classification of each gloss so that readers might judge for themselves (Parry, in press). It is not possible to do so for the more extensive data collected from Dimitri and Ae Young; but I did compare my classification with that of an independent assessor, and we reached 77% agreement on Dimitri's glosses and 84% on Ae Young's. Moreover, the 38 differences that there were between us followed a consistent pattern: 26 of them concerned glosses that I classified as incorrect and my colleague classified as partly correct, and a further 8 concerned glosses I classified as

partly correct and my colleague as correct (the number of each kind of disagreement was exactly the same for each student). It would seem, then, that the figures given above are based on a judgment that is relatively stringent.

This is an encouraging observation, for even by my more stringent classification, the students did remarkably well. Each recorded a substantial proportion of correct glosses (Yuko's 33% is particularly remarkable) and got more than half of the total at least partly correct. It would seem, then, that students are well advised to try and make inferences, even though most of the time they will not arrive at an accurate representation of meaning; and it can be hypothesized that a trace of each inference will remain to be modified in subsequent encounters with the word.

How do readers arrive at such inferences? Huckin and Bloch (in press) present an interesting model of what goes on, based on protocols collected from Chinese learners of English as they carried out a difficult translation task. The model postulates a cognitive structure for dealing with new vocabulary called a "generator-evaluator." This generator-evaluator consists of a number of modules that are interconnected and operate in parallel; they include vocabulary knowledge, text schemata, syntax and morphology, world knowledge, and the text representation (i.e., the representation of the particular text being read). When an unknown word is encountered, an attempt is made to hypothesize its meaning, using the modules in the generator-evaluator, and the modules are again used to test the hypothesis. If it is found to be unsatisfactory, or if no hypothesis is produced at all, the learner may read on to generate more context and try again; or he or she may "satisfice," putting the word into a working-memory "buffer," and coming up with a hypothesis perhaps much later. If, on the other hand, a hypothesis is generated and accepted, it is used to update the modules within the generator-evaluator; hence it may help to deal with other items already in the buffer and also modify the text representation as the learner continues reading. The hypothesized meaning for the word may also, of course, become part of the learner's vocabulary knowledge, but that certainly does not happen in every case.

Ae Young's protocol can be seen in terms of this model, with the exception that since she had a dictionary (Huckin and Blochs informants did not) she had an additional means of testing her final hypothesis before continuing. Here, for example, is what she said as she came to the first unfamiliar word in the text. The words of the text, which she read aloud, are printed here in italics, while her own comments are in ordinary type; repetitions are shown, hesitations are indicated by slashes, and miscues are placed in square brackets:

1. *Stated another way, this means that in the / matri-matrilineal society, a woman—a woman's chi-children, male and female, belong to the lineage to which / she herself belongs. [The] lineage / would also include her brothers and sisters born of the same mother; her mother and her mother's brothers and sisters born of the same mother; her mother's mother[s], and so on. By the same principle, it follows that children / of u/terine—* Uterine. What is it? [writes] U/te/rine. OK.

She was apparently not able to generate a hypothesis at this point, so she reviewed the context and read on a little:

2. *Uterine sisters—children—it follows that children of uterine sisters also belong to the same / matrilineage. By the same principle / it foll–it follows that children of uterine sisters / also belong to the same mat-matrilineage.* Uterine sister, her mother's mothers. OK.

She still did not succeed in generating a hypothesis, but she did not give up; instead, she went over the previous sentence again, this time with more success:

3. *That lineage would also include her brothers and sisters born of the same mother, her mother an* mother—Uterian—I don't know-OK. Uterine. It is step sis—step sisters?

Then she tested the hypothesis:

4. *It follows that children of uterine sisters / also belong to the same matrilineage.* Oh, it means like—I think it's like—cousins, I mean—I mean—um-um cousins whose—um mothers are sister. I don't know, I never heard of it, so I think I have to look up this dictionary.

And, accordingly, she did so.

In this case the cues that Ae Young used were all contextual; but there were also cases where she used what Faerch et al. (1984) describe as intralingual cues—that is, her understanding of English morphology—together with contextual ones. For example, in the following passage she did not know the word *bilaterally:*

5. A second difference between the lineage and the extended family concerns the manner in which descent is reckoned in the two groupings. In the lineage descent is traced through one line only, whereas in the extended family descent (or more technically, filiation) is traced bilaterally, through both parents. (Sudarkasa, 1982, p. 136)

Her first hypothesis for the word was apparently derived from the context:

6. *Bil-bilat-bil-bilaterally* It means *whereas in the extended family descent is traced bi-bila-bdaterally* It means like both parents, the mother and the father—no, it's not, it's not the same thing. Filiation is traced. Bilaterally. Umm. Bi-bilaterally *through both parents* Right, it means through both parents.

Then she read on a bit:

7. *Thus in a patrilineal or matrilineal society, from the point of view of any given person, the lineage compr-comprises relatives linked through the father-line or the mother-line, as the case may be. In both types of society, however, the extended family comprises relatives on both the mother and the father side.* Umm.

At this point she went back to the beginning of the paragraph, stopping at "through both parents." Then she went back again:

8. Hmhm. *Between the lineage and the extended family concerns the manner in which descent is reckoned in the two groupings—* reckoned in the two groupings. *In the lineage descent is traced through one line only. In the lineage descent is traced / through* [at this point she wrote *bilaterally* down] In *the lineage descent is traced through one line only whereas in the extended family descent or more technically filiation is traced bil-l-laterally through both parents.* Bilaterally. It means like something both. Because it has *bi-* so it means—like both. Bilaterally. From, I mean, do with both parents, is that mean? Yeah, I think so, it's j-just like through both parents, I think it's the meaning.

So she wrote her inference down "through both parents."

Dimitri's strategies can also be seen in terms of Huckin and Bloch's model, although, since he was less articulate than Ae Young, it is more difficult to infer what he was doing. Here, for example, is what he did with the word *uterine* when he read the text for the first time:

9. *By the same principle, it follows that children of uterine* [he marked the word in his text] *sisters also belong to the same matrilineage[s].* This word uterine—[here he lapsed into silence until asked to speak up] *By the same principle that children uterine sisters—* I believe it's the opposite uh that the lineage—matri—the—what's this—matri-lineage [he had already marked *matrilineage* as unfamiliar] uh—it's the opposite of that.

Then he read on, without apparently testing his hypothesis, and did not come back to the word until he was reading the text for the second time—in Huckin and Bloch's terms, he seems to have put it into the buffer. On the second reading he came up with a different hypothesis, which he expressed and wrote down in Greek: It means "half-brother or sister." At that point he resorted to the dictionary.

In the case of *bilaterally,* which Dimitri also did not know, he simply marked it on the first reading and went straight on—his strategy was, apparently, to use the buffer extensively. Then, on the second reading, he read up to the word twice and generated a hypothesis:

10. *In the lineage descent / is traced through [the] line—through one line only, whereas in the extended family descent (or more technically, filiation) is traced / bilater-rarily- bilaterally* [he wrote the word down]. *In the lineage descent is traced through one line only whereas in the extended family descent is traced bilaterally*. It means—by—by—both descent and technical.

He then apparently tested the hypothesis by rereading the sentence up to "through both parents." This led him to a new hypothesis:

11. That means by both—both parents, I can say, through—

At this point he turned to his dictionary.

These examples give some idea of how the students' lists—at least Dimitri's and Ae Young's—may have been arrived at. They suggest one important common feature—namely, that the students depended primarily on context for hypothesizing meaning, even though they sometimes used intralingual cues (and Dimitri also occasionally interlingual ones, as when he recognized *labyrinthine* as a Greek cognate). But the protocols also suggest significant differences of strategy: Dimitri tended to set particular difficulties aside in favor of constructing a more general representation of the text; Ae Young apparently preferred to build up a more exact representation at the local level—we have seen how frequently she went over the sentences surrounding a word she was working on. It would seem that Dimitri's method is the more efficient. Not only did he complete this task more quickly, but in his lists as a whole, a large majority of his inferences are at least partly right, whereas Ae Young produced fully 49% that were wrong.

There is, however, another side to this: Dimitri may have got through the task more quickly, and his inferences may have been more generally correct, but the posttest showed that he did not seem to be adding these words to his vocabulary. The posttest was different for the two students, being constructed by taking every fifth word that each had recorded (which meant that, since Ae Young had recorded more words, she got a few more on her test), and it was presented in two stages. First, the students were given the words in isolation and asked to write down the meanings of any they could remember, using whichever language they chose. Second, the words were presented again, this time within the context of the single sentences in which they had appeared when the students first recorded them. Not surprisingly, the students did much worse at the first stage than at the second; but it is striking that whereas Ae Young attempted to define 7 out of 24 words that she was given at this point, Dimitri attempted only 1 out of 19. Further, when it came to the second list, the single sentences that were

provided obviously triggered memories for Ae Young, for her definitions corresponded closely to either her previous guesses or to the dictionary definitions that she had found: The result was that she gave 19 definitions in all, of which 6 were correct, 9 were partly correct, and only 4 were incorrect. Dimitri, on the other hand, seems to have remembered less and depended on the reduced context more: Of the 18 definitions that he produced, only 2 were correct, 3 were partly correct, and 13 were incorrect—and some of the incorrect definitions were for words that he had guessed correctly in the first place. It seems, therefore, that Ae Young's meticulous approach leads to better long-term memory for the items. What she remembered, though, was often her own inferences rather than the dictionary definitions; where these inferences were wrong, there is obviously a problem, so it is to her problems of interpretation that we now turn.

PROBLEMS WITH GUESSING

The first and most significant point about Ae Young's 58 incorrect guesses is that the largest group (23) make perfectly good sense in the context. Consider, for example, her guess for *circuitous* in the opening passage of the book (Keesing, 1981):

> 12. This journey through anthropology will take us through remote corners of the world—African deserts and coral lagoons in the South Pacific—and then will take us back to the crises and complexities of the 1980s and the challenges of the century that dawns ahead. We may well pause before embarking to ask why such a *circuitous* route, which will take us through ways of life now vanished or transformed, is worth taking. (p. 1)

Ae Young's gloss for this, a Korean expression for "special, not common," is perfectly reasonable. Another example comes from this sentence:

> 13. In any case, what people tell the ethnographer about their way of life must be cross-checked, *substantiated,* and filled out by detailed records of actual events and transactions. (p. 7)

The gloss for *substantiated* is "examine''—one which again indicates that, while Ae Young may not have correctly inferred the meaning of the word (and while she may not have fully grasped its morphological properties), she was reading the text surrounding it with understanding.

There are some cases, however, where she seems to be going quite badly off track. How, for example, could she interpret *missionaries* as "big houses" in the following context?

14. By the time an anthropologist arrived on the scene there were usually colonial administrators collecting taxes and imposing peace, as well as trade stores and missionaries. (p. 5)

If we examine the immediate context more closely, however, we see that *missionaries* is paired with *stores;* the assumption that the unknown word must refer to another kind of building is not altogether unreasonable. Further, there may be a problem with the syntax here, which is actually ambiguous: Did Ae Young think that the colonial administrators were imposing trade stores and big houses as well as peace? And a final point about this guess is that it may also have been influenced by some vague memory of the slightly similar word, *mansions.*

If the last point is correct, "big houses" for *missionaries* is partly a case of what Huckin and Bloch (in press) characterize as "mistaken ID. " Another, much clearer, example of this is her gloss for the word *peasant,* which appears in this context:

15. But anthropologists no longer work only, or even predominantly, in such societies: they study *peasant* villagers, including those in Europe . . . (p. 1)

The gloss for this is written in Korean and translates back as "peacock." This apparently makes no sense at all; but Ae Young herself explained what had happened, for she wrote in the comment, "I confused it with pheasant." It seems that in this case she thought she knew the word, but since she could not understand how it could fit in the context, she noted it as problematic.

Mistaken IDs, however, only account for a few of Ae Young's incorrect glosses, and it is more difficult to suggest why the others occurred. One reason undoubtedly was the sheer accumulation of unknowns, as in the following example:

16. Ironically, these approaches, in their concern with experimental rigor, systematic inference, laboratories, prediction, measurement, and so on, have very often *emulated*—even *parodied—latter* nineteenth-century natural science. (p. 5)

Given the concentration of unfamiliar words here it is remarkable how well Ae Young did: She guessed *rigor* as "result"; *inference,* in Korean, as "influence" (this is probably another case of mistaken ID); *emulated* as "changed; and *parodied,* again in Korean, as "absorb." The whole makes some sense, especially if she thought, as students often do, that the *nineteenth* century refers to the present one. Still more remarkable is the fact that she could put an interpretation—though definitely not the intended one—on a sentence like this:

17. Some [anthropologists] spend their time chasing through African *thickets* in *pursuit* of *fuzzy primates.* (p. 2)

Her glosses were, respectively, "green grass" in Korean, "clothes" in English (was she thinking of suit?), "faint, dizzy" in Korean (is this another mistaken ID—a compound of *faint* and *dizzy?),* and "pattern, design" in English; perhaps she had a vision of anthropologists going about their business in camouflage dress.

Another factor causing Ae Young difficulty may have been occasional confusion over morphology. Like Yuko (Parry, in press), Ae Young seems sometimes to be unaware of the significance of suffixes: she glossed the word *accumulation,* for example, as "complex one" in Korean, thus apparently missing the abstraction implied by *-ation.* Similarly, in the following sentence, the *-al* ending of *colonial* and *neocolonial* failed to alert her to the fact that the two words are adjectives, although it seems that she did know the related noun, *colony:*

18. We will see how the anthropologist has, usually unwittingly, been aligned with *colonial* and *neocolonial* forces. (p. 6)

She glossed *colonial* as "a part of country" and *neocolonial* as "group of colonies."

The above example also shows problems with syntax. For a native speaker, *colonial* and *neocolonial* could not be considered nouns, even if their form allowed it, because neither is preceded by a determiner. Further, they are followed by the word *forces,* which Ae Young's interpretation does not seem to account for at all. Another example of such apparent interaction of morphological with syntactic problems is in the case of *perspectives:*

19. Yet anthropological approaches and *perspectives* remain distinctive. (p. 4)

Her guess for *perspectives* was "acquire"; perhaps she thought that *approaches* was a verb (which the form of the word allows), and confused *anthropological* with anthropology, but it is hard to imagine what she thought *remain distinctive* meant.

From Ae Young's data, then, we can hypothesize five factors leading to misinterpretation of unknown words:

1. The general context, which allows plausible but nevertheless incorrect inferences as to meaning. This seems to be the most significant factor, accounting for 23 of Ae Young's incorrect glosses.
2. The particularly strong influence of words in the immediate context, as opposed to the wider context. This may account for an additional 4 incorrect glosses.

3. Mistaken ID. There are 4 cases where this seems to have happened.
4. A concentration of unknown words in a small amount of text, so that each word has correspondingly little contextual support. There are at least 10 cases where this factor must have made it extremely difficult to guess the meanings of the words.
5. Confusion about, or at least failure to use, morphological and syntactic cues. Eight of Ae Young's incorrect inferences can be explained by this factor.

Some glosses appear in more than one of these categories, and there are 8 that are not accounted for at all, for they appear to be inexplicable. Indeed, such explanations as are offered here must be taken with reservation: There is no way of knowing what Ae Young was actually thinking. Too much weight, therefore, should not be put on the figures suggested here, and this analysis of error should be considered a starting point for enquiry rather than a definitive statement.

But how much does such misinterpretation matter, as far as vocabulary building is concerned? In Ae Young's case it does seem to have mattered somewhat, for there is some evidence that her initial misinterpretation of a particular word did, sometimes at least, affect her subsequent understanding of it. There is one word, *crucial,* which comes up twice in her lists. In the first case she glossed it, in Korean, as "exact, definite," which is incorrect, but plausible in the context. In the second case she provided a similar gloss, "accurate" (in English); but in this context the guess is much less plausible. Again, in her posttest there were eight items for which she had originally supplied incorrect glosses; of these she reproduced four in the test. It seems, then, that it is possible for learners to place idiosyncratic repesentations on particular words and to remember and build on such representations; maybe it is such idiosyncratic representations on words that she did not record that account for the misinterpretations that I have been unable to explain.

Dimitri's case is in marked contrast to Ae Young's. It will be remembered that he had many fewer incorrect glosses than she had, and when they are classified in terms of the five factors outlined above, the proportions are quite different: Of the 27 glosses in question, 20 could be explained as plausible guesses in the context, 3 as arising out of syntactic confusion, one possibly as mistaken ID (he glossed *arrogance,* in English, as "propoganda," suggesting that he may have thought they were related words), and there are 3 that defy explanation. His posttest performance was indeed poor, but

this was not, it seems, because he remembered previous incorrect inferences. It was rather that he did not remember his previous inferences at all, at least, not enough to use them for defining the words. Consequently, he apparently tried to work out the meanings again from context, but the context was too reduced on this occasion for the strategy to be successful.

The figures suggest that Dimitri's style of reading—relatively rapidly without spending too much time on each new word—is more efficient as a means of vocabulary building than is Ae Young's, even if each item is not remembered to the point that it can be defined. But is it better for the purposes of studying a particular academic subject? This question must be raised because of a difficulty that became apparent in the translation task. The purpose of including this task as part of the research design was to see whether students were in fact misinterpreting or simply not noticing words that they did not record. Dimitri's translation suggests that the latter was indeed happening. What the students were asked to do was to translate two paragraphs from the protocol passage into their first language, the subject of these two paragraphs being differences between the lineage and the extended family in African kinship. In doing the translation, Dimitri worked rapidly, as he had in doing the protocol, and with apparent confidence; it came therefore as a surprise to see that he did not seem to know the word *lineage* and, which is worse, he did not realize that he did not know it. The result is that although his translation starts well, it soon becomes confused, and parts of it degenerate into nonsense. Ae Young's translation, on the other hand, shows that she understood the basic distinction, even though her Korean text, or rather its back-translation, reads very peculiarly. Can we conclude, then, that Ae Young, by reading more slowly and carefully, was in fact comprehending more—and could this account for the fact that she ended with a B for her anthropology course, whereas Dimitri got only a C?

CONCLUSION

The last section has appropriately ended with a question, for these studies raise more questions than they answer. What does it mean when a student records a word as difficult? How much can we assume that students understand of words that they do not record? HOW accurately do the students' inferences reflect their thoughts? To what extent do students depend on the context for these inferences, and to what extent are they affected by some memory, however vague, of previous encounters with the words, or

with similar ones? With so many unknown factors, the results of these studies must obviously be treated with caution. Nevertheless, they do provide some evidence on vocabulary acquisition, and here, by way of conclusion, I shall highlight the most important points and use them as a basis for some pedagogical suggestions.

First, the figures quoted at the beginning of this article suggest that there is, as we have always suspected, a strong correlation between how much people read and how many words they know. Ae Young read very little—at least of her anthropology text—and she recorded a higher proportion of words than did any of the other students. Amongst those words there were also proportionately many more of relatively high frequency. This seems to indicate a circular process: she reads little and so comes across relatively few words; consequently she is slowed down in her reading by the many words that are unfamiliar; this, in turn, means that she does not read much, and so on. Dimitri, by contrast, read a great deal and seemed to know a large proportion of the words, especially of the more frequent ones. Here the circle seems a more virtuous one: He has to stop for few words, it is relatively easy for him to infer their meanings, so he can read more quickly, encounter more new words, and reencounter sooner those that he is already getting to know. Thus, one aspect of our traditional practice in teaching ESL is confirmed: To establish a firm foundation for the vocabulary building to be done in academic courses, we should encourage our students to read as much as they can before they leave our classes. As Krashen (1989) has argued, plenty of comprehensible input may be the single most important factor in second language acquisition, especially when it comes to vocabulary building.

Second, the protocols showed that individuals may have significantly different strategies, and that these strategies may radically affect the way in which they learn new words. Ae Young worked long and carefully over the words she did not know, with the result that she both remembered them relatively well (as shown in the posttest), and developed a relatively precise understanding of the text (as shown in the translation task). Dimitri, on the other hand, worked much more quickly, remembered less well, and overlooked a crucial word in the translation passage, so that he missed the whole point. Here there seems to be a contradiction: How can Dimitri build his vocabulary so successfully if he does not remember the new words he comes across—and fails to notice some altogether? The answer may be that the process of vocabulary learning varies at different stages and for different kinds of words. Given that such a large proportion of inferred meanings is at least

partly correct, those words that are relatively frequent may be learnt by simple encounter and reencounter: A trace of each inference may remain in memory, even though the learner cannot consciously recall it, and this trace may be reinforced and revised in subsequent encounters. Thus Dimitri's strategy may have helped him to get such relatively frequent words in place before embarking on this anthropology course; and we should encourage the students in our classes to adopt strategies similar to Dimitri's—rapid reading without spending too much time on each word. Less frequent and more specialized words, however, may need more careful treatment to be remembered, since they are unlikely to be reencountered soon. Ae Young's strategy may have been more appropriate for this purpose, and so she may have acquired a better understanding of crucial terms, such as *lineage*. What we need to think about, as language teachers, therefore, is how students can learn to recognize those words whose meaning they must know accurately in order to understand a particular text. This implies some work on cohesion: Words that are introduced in a leading sentence and then thematized by repetition or by the use of reference items are obvious candidates for dictionary work; so also are those that are printed in bold or italic type. Most other words, however, need not be looked up, and they should not be because it takes so much time and interrupts the flow of reading. In short, our students must learn both Dimitri's and Ae Young's strategies, and they must develop criteria for deciding which is appropriate in any particular case.

Third, there is a real danger of misinterpretation, and if incorrect inferences are remembered, it may have a cumulative effect. Some of the factors that lead to misinterpretation may be guarded against: We should spend some class time on guessing from context so that we can demonstrate the necessity for using the text beyond the immediate sentence, checking for mistaken IDs, and using morphological and syntactic cues. But even when students have become very good at this, there remain difficulties that are in the nature of text and in the way words are joined together to construct it. In a text each word derives meaning from its context, but it also contributes meaning that is derived from all the other contexts in which it may be found—or, rather, may have been found in the experience of individual readers. When readers are dealing with new words, however, they have only the present context to go on, and so they have only half of the equation; indeed, ESL and EFL students may have less than half, because their understanding of the other, familiar words is unlikely to match that of a native speaker. We must also, therefore, alert our students to the problem,

impressing on them that building a vocabulary requires a great deal of work, and that as more pieces are added and more connections made, there must be a continuous process of reviewing, and sometimes undoing, construction that has already been done,

The same may be said of building a theory of vocabulary acquisition. Viviana, Yuko, Dimitri, and Ae Young have suggested a few of the pieces that we might use in its construction, but much more research must be done before we can expect to have a reliable model. It is important that researchers turn their minds to the problem, for it is only with such a theory that we can do justice in our teaching to vocabulary building as a process. ■

ACKNOWLEDGMENTS

This article is a revised and expanded version of a paper presented at the 24th Annual TESOL Convention in San Francisco, March 1990. This research was supported in part by grant number 666415 from the PSC-CUNY Research Award Program of the City University of New York. I should like to thank Lelia Ruckenstein, John Ingulsrud, Yasuko Watt, Joo-Eun Ju, and Agathi Raptu for their help as translators; Sonja Hubbert for her work in processing the data; Franklyn Horowitz for providing an independent classification of the glosses; and Ann Raimes and Harriet Johnson for helpful comments on early drafts of the article. Above all, I would like to thank Viviana, Yuko, Dimitri, and Ae Young, without whose cooperation this project would have been impossible.

THE AUTHOR

Kate Parry is Associate Professor in the Department of English at Hunter College of The City University of New York. She has previously taught in Uganda and Nigeria. Besides her work on vocabulary acquisition, she is interested in literacy assessment, the development of a pedagogical grammar, and the social history of English.

REFERENCES

Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon.* Oxford: Basil Blackwell.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly, 20* (3), 463-494.

Bright, J. A., & McGregor, G. P. (1971). *Teaching English as a second language: Theory and techniques for the secondary stage.* London: Longman.

Carter, R. (1987). Is there a core vocabulary: Some implications for language teaching. *Applied Linguistics, 8,* 178-190.
Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly, 8* (4), 389-399.
Ember, C. R., & Ember, M. (1985). *Cultural anthropology* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
Emig, J. (1971). *The composing processes of twelfth graders.* Urbana, IL: NCTE.
Faerch, C., Haastrup, K., & Phillipson, R. (1984). *Learner language and language learning.* Clevedon, England: Multilingual Matters.
Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science.* Cambridge: MIT Press.
Haynes, M., & Baker, I. (in press). American and Chinese readers learning from lexical familiarizations in English text. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning.* Norwood, NJ: Ablex.
Hill, C. A., & Larsen, E. (1983). *What reading tests call for and what children do.* Washington, DC: National Institute of Education.
Hofland, K., & Johansson, S. (1982). *Word frequencies in British and American English.* Harlow, England: Longman.
Huckin, T., & Bloch, J. (in press). Strategies for inferring word-meanings in context: A cognitive model. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning.* Norwood, NJ: Ablex.
Keesing, R. M. (1981). *Cultural anthropology: A contemporary perspective* (2nd ed.). New York: Holt, Rhinehart & Winston.
Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73* (4), 440-464.
Kuçera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.
Levenston, E. A. (1979). Second language acquisition: Issues and problems. *The Interlanguage Studies Bulletin, 4,* 147-160.
Martin, A. V. (1989). Bridging vocabulary: An essential component of ESL proficiency. Paper presented at the 23rd Annual TESOL Convention, San Antonio, TX.
Meara, P. M. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching & Linguistics: Abstracts, 13,* 221-246.
Meara, P. M. (1983). *Vocabulary in a second language* (Specialized Bibliography No. 3). London: Centre for Information on Language Teaching and Research.
Palmberg, R. (1987). On lexical inferencing and the young foreign-language learner. *System, 15,* 69-76.
Parry, K. J. (1987). Learning vocabulary through an academic subject: A case study. In S. Dicker & L. Fox (Eds.), *Improving the odds: Helping ESL students succeed. Selected papers from the 1987 CUNY ESL Council Conference.* New York: The City University of New York, Instructional Resource Center.

Parry, K. J. (in press). Too many words Learning the vocabulary of an academic subject. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning.* Norwood, NJ: Ablex.
Perl, S. (1979). The composing processes of unskilled college writers. *Research in the Teaching of English, 13,* 317-336.
Plog, F., & Bates, D. G. (1980). *Cultural anthropology* (2nd ed.). New York: Alfred A. Knopf.
Raimes, A. (1988). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning, 37,* 439-468.
Sudarkasa, N. (1982). African and Afro-American family structure. In J. B. Cole (Ed.), *Anthropology for the eighties.* New York: The Free Press.
Trimble, L. (1985). *English for science and technology: A discourse approach.* Cambridge: Cambridge University Press.
Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly, 17* (2), 165-187.

TESOL QUARTERLY, Vol. 25, No. 4, Winter 1991

A Comparison of Three Learning Strategies for ESL Vocabulary Acquisition

THOMAS S. BROWN
Boise, Idaho

FRED L. PERRY, JR.
American University in Cairo

The purpose of this study was to compare three learning strategies–differentiated according to Craik & Lockhart's (1972) "depths of processing" theory—for ESL vocabulary. Six intact ESL classes at two levels of proficiency were divided into three treatment groups (keyword, semantic, and keyword-semantic). These Arabic-speaking students then received 4 days of instruction. Both recognition and cued-recall instruments were used to measure effects both 1 day and 9 days after treatment. Cued-recall results immediately after treatment revealed that the keyword method facilitated vocabulary acquisition for lower-proficiency students. The delayed results for both the recognition and cued-recall tests suggested that the combined keyword-semantic strategy increased retention above the other strategies. Possible applications of these findings are discussed.

Students learning English for higher education face a formidable task. Nagy and Herman (1987) summarized a number of studies investigating the acquisition of vocabulary in native English speakers, and estimated that by the last year of high school the typical student has learned 40,000 words, an average of around 3,000 words per year. A logical extrapolation is that an ESL student who is learning academic English would have to learn on average more words per year than this. The question is, How can students increase their learning power for new vocabulary?

Oxford (1986) has argued that a greater emphasis needs to be placed on identifying effective second language learning strategies and teaching students how to use them. Previously, she had prepared a taxonomy of second language learning strategies (Oxford-Carpenter, 1985) which provides a useful overview for this

purpose. Under the heading of Vocabulary Acquisition she has listed a number of strategies which include keyword and elaboration (henceforth referred to as semantic processing). The purpose of the present paper is to compare these two strategies, in addition to a combination of the two, with regard to their facilitative effect in learning new vocabulary,

A considerable amount of research has taken place since the late 1970s concerning vocabulary learning strategies utilized by native speakers of English. Emphases have mainly been on three strategies: keyword (Pressley, Levin, & Delaney, 1982), contextual (Sternberg, 1987), and semantic processing (Beck, McKeown, & Omanson, 1987). Of these three, the keyword method has received most attention and has been shown to be superior to contextual and no-strategy conditions (Pressley, Levin, McDaniel, 1987). Sternberg (1987) claims that "most vocabulary is learned from context," but adds that "what the claim does not imply is that teaching specific vocabulary using context is the most effective, or even a relatively effective, way of teaching that vocabulary" (p. 89). The third strategy, semantic processing, is just beginning to receive attention in the research literature. Only the keyword and semantic processing strategies will be investigated in the following study.

The keyword method was first described by Atkinson (1975). To remember a new word's definition, a keyword is chosen which is acoustically similar to the new word, yet has a meaning of its own independent of the new word's meaning. A visual association through an image is then made between the keyword and the new word's meaning. "To remember that *carlin* means old woman, a subject might use the keyword car, and imagine an old woman driving a car" (McDaniel & Pressley, 1984, p. 598). In this example, the keyword, *car,* which is acoustically similar to the new word, serves as a cue for an image which in turn aids the recall of the word's definition.

Unlike the keyword method, no single semantic technique is universally recognized as the semantic processing method. However, two defining characteristics of semantic processing are used in this study: Focus must be on the meaning of the new word, and the learner must act upon the meaning of the new word in a way that is considered integrative in relation to already existing semantic systems. This goes beyond merely associating a word with its definition. Beck et al. (1987) state, "students should be required to manipulate words in varied and rich ways, for example, by describing how they relate to other words and to their own familiar experiences" (p. 149).

Although several people (Channell, 1980; Martin, 1976) have proposed various learning strategies for acquiring English

vocabulary in a second language environment, to date little research has been carried out regarding the effectiveness of various learning strategies for vocabulary in English as a second language. Some research exists on Russian (Atkinson, 1975; Atkinson & Raugh, 1975), Spanish (Levin, Pressley, McCormick, Miller, & Shriberg, 1979; Pressley, 1977; Pressley, Levin, Hall, Miller, & Berry, 1980; Raugh & Atkinson, 1975), Latin (Pressley, Levin, Nakamura, et al., 1980), and German (Desrochers, Gelinas, & Wieland, 1989; Hall, 1988). All of these studies looked at the keyword method in comparison with a no-strategy condition. To our knowledge, Crow and Quigley (1985) have published the only study looking at the effectiveness of ESL vocabulary learning strategies. Their study examined several semantic processing strategies and found them to be superior to no-strategy conditions. It remains to be determined how the keyword and semantic processing strategies compare in facilitating acquisition of new vocabulary.

Craik and Lockhart's (1972) depths-of-processing theory provides a theoretical basis for comparing the keyword with semantic methods. They suggested that retention is dependent on the level at which information is processed: As one moves from the shallow sensory level of processing to the deeper semantic level, memory traces become more permanent. At the sensory level, the stimulus is processed in terms of its visual or acoustic properties. In contrast, stimuli processed at the semantic level are analyzed for meaning and related to existing cognitive structures.

Craik and Tulving (1975) expanded the theory to suggest that retrieval is also enhanced by elaboration; that is, further processing at a certain depth of memory. For a memory trace to undergo elaboration at the sensory level, additional acoustical or visual processing must occur. For example, elaboration at the sensory level would occur when several phonetic features of a word, such as both its vowels and consonants, are given special attention. Semantic elaboration would occur if a meaningful response to a question about the meaning of the word was given (Perry, 1982).

Criticism of the depths-of-processing model has centered on the circular nature of the definition of the essential construct (Benton, Glover, & Bruning, 1983). However, others have been able to operationalize the construct as did Johnson-Laird, Gibbs, & deMowbray (1978) who defined it as the number of decisions made requiring meaningful information about the subject. Benton et al. (1983) found that as the number of decisions toward textual materials produced by yes/no type questions increased, so did the amount of information recalled.

A reasonable suggestion is that the type of processing used with keywords involves mainly elaboration at the sensory level combined with some attention at the semantic level (Pressley, Levin, & Delaney, 1982). Through the merging of acoustical and imagery elaboration produced by a keyword, a link is established between the L2 new word and the L1 meaning. However, the focus is on acoustical similarity and the interlinking associative image, which does not necessarily encourage active processing at the semantic level. Pressley, Levin, Digdon, Bryant, & Ray (1982) argued that the keyword method is semantic to the degree that a definition must be meaningfully related to the keyword; however, the relationship between the definition and the keyword is arbitrary in that the meaning of the keyword and the new word are independent of one another. The primary association between a new word and its definition is acoustical and visual. In line with Craik and Tulving's (1975) thinking, some retention benefits for vocabulary processed through keywords should be expected. However, semantic processing is not the main focus of the keyword method and should not produce better retention for the meaning of a new L2 word than a method that relies more heavily on semantic processing.

The primary emphasis in any semantic processing method for learning L2 vocabulary is on the semantic association between the new word and its definition. Any procedure that causes the learner to act on the meaning of a new word by tying it into existing knowledge structures would fit into this category. For example, Crow and Quigley's "semantic field" approach (1985) had subjects manipulate synonyms along with target words in meaningful sentences. Eeds and Cockrum (1985) had subjects use their target words in seven different meaningful tasks. In the current study, the semantic processing method used is defined as follows: Two different examples of usage were provided in order to activate appropriate semantic structures; also a question was asked whose answer necessitated the use of the new word. Perry (1982) reviewed theoretical positions and research that would support such a definition. According to depths-of-processing theory, this semantic processing method should better aid retention than the keyword device, which appears to primarily elicit sensory processing. Words learned through examples and a question related to a learner's experience should be remembered longer than words learned through arbitrary acoustic and visual associations.

In addition, one could argue from the depths-of-processing theory that elaboration at all levels would enhance memory above that at any one level alone. In other words, using a keyword would

provide an initial link between an L2 word and its meaning in Ll, whereas semantic elaboration would further fix the semantic association within existing knowledge structures. If this is the case, then the combination of the two methods should produce better results than either on its own.

Several moderating variables have shown to be of importance in the study of L1 vocabulary learning strategies and should be taken into consideration. First, the keyword method has mainly been found to be effective with individual presentation rather than with group presentation (Hall, Wilson, & Patterson, 1981; Pressley et al., 1987). On the other hand, various semantic processing methods have been found to be effective with group presentation (Crow & Quigley, 1985; Eeds & Cockrum, 1985; McKeown, Beck, Omanson, & Pople, 1985). This is an important factor since most vocabulary learning in the classroom occurs through group presentation. Second, there is some evidence that the superiority of the keyword method over no-strategy conditions decreases in long-term memory (McDaniel, Pressley, & Dunay, 1987). This may weaken the usefulness of the keyword method since the objective is for students to remember the meanings of new words over time. Results for semantic processing methods have demonstrated beneficial effects over long time periods (Crow & Quigley, 1985). Finally, in studies of native-speaking students, the effectiveness of the keyword method has differed according to the verbal ability of the student: those of lower ability finding it more useful than those of higher ability (McDaniel & Pressley, 1984). Evidence for the semantic processing method suggests that both ability groups can benefit (Eeds & Cockrum, 1985).

The purpose of this study is to compare the three strategies of learning ESL vocabulary discussed above: the keyword method, a semantic processing method and a combined keyword-semantic condition. All methods were administered to groups of subjects in actual classroom situations. Upper and lower English proficiency levels were included in order to examine the differential effects which may occur among methods. Measures of recognition as well as cued-recall were used to look at whether any differences existed between retention of information and the ability to retrieve information. In order to study the differential effects over time, treatment effects were studied both immediately after, shortly after, and 9 days after administration of the treatments.

Answers to the following questions were sought:

1. Will the semantic processing method produce superior results in comparison to the keyword method over a longer time period?

According to the depth and elaboration of processing theory, there should be a superior effect.

2. Will the combined keyword-semantic method facilitate longer term retention than either the keyword or the semantic processing methods used alone? Again, the stated theory would suggest that there might be.
3. Are there any differential effects for the instructional methods being examined immediately after treatment as compared with longer time spans?
4. Are there differential effects for methods over varying time spans for two different levels of English proficiency?

METHOD

Design

A nonequivalent control-group design (Borg & Gall, 1989) was used in this study. There was a control group and a treatment group; however, subjects were not randomly assigned. This design was chosen for two reasons: one practical, the other theoretical. First, the instructional program used in this study could not tolerate the breaking up of classes to facilitate random assignment of subjects. Second, in order to make more ecologically valid generalizations to real classroom environments, authentic classroom situations were used with authentic students who were presumably studying with genuine motivation to succeed in the course: Students' performance in the classroom determined entrance into university. One could argue that experimental design with random assignments of students to treatment and control groups would make our findings more generalizable from the point of randomization of individual differences. However, Snow (cited in Borg & Gall, 1989) has argued that fully randomized experimental designs often lack ecological validity due to the inauthentic environments in which studies are carried out. Because the intention of this study was to provide the practitioner with findings that are closer to their own classroom settings, the authors selected the design described here.

Subjects

Six intact classes from the English Language Institute (ELI) at the American University in Cairo (AUG) participated in this study. Three classes were upper level and three were lower level. Upper level students were defined as those scoring between 82 and 75 on the AUC English Entrance Test (parallel in form and content to the

Michigan Test of English Language Proficiency) while lower-level students were those scoring between 74 and 65. The classes had a balanced ratio of males and females, Egyptians and students from other Arab nations. A total of 94 students received at least some instruction and testing. We considered only students who received all instruction and testing, leaving a total of 60 subjects.

Materials

Vocabulary. The words to be learned were chosen according to two criteria. First, they had to be appropriate for ESL students at the university level. Of the 40 nouns and verbs chosen, 31 were taken from a computerized word list containing words found in first-year college textbooks (Praninskas, 1972). The least frequently occurring words on the list were chosen. The remaining words were taken from readings that ELI teachers planned to use after completion of the final posttest.

Second, the words were to be unfamiliar to the students. Two steps were taken to insure novelty. Words chosen from the frequency list were eliminated if they appeared in sections of the textbooks or handouts that teachers had taught before experimental instruction began or would be taught before the final posttest. Also, a pretest checklist containing 51 words presented in short sentences with minimal contextual clues was given to subjects several days before instruction to eliminate words that students already knew. Subjects were asked to rate how well they knew the words on a 5-point, Likert-type scale. The 11 words with means greater than 3.5 were eliminated, leaving 40 words for instruction.

Definitions of the target words were paraphrased from English dictionary entries. Unlike many previous keyword studies (e.g., Pressley, Levin, Kuiper, Bryant, & Michener, 1982) in which definitions usually consisted of one or two synonyms, these definitions consisted of short phrases. This was due to the fact that the more abstract nature of the vocabulary chosen for instruction in this study required longer definitions than the more concrete vocabulary chosen for instruction in most keyword studies.

Keywords. Native Arabic speakers chose keywords for the selected vocabulary. The criteria for keyword selection were (a) the keyword should sound like at least one syllable of the chosen word, preferably either the stressed or first syllable; (b) the keyword should be a concrete noun or verb; and (c) the keyword should be Egyptian Colloquial Arabic. Seven of the 40 keywords varied slightly from these criteria. For these seven words, the acoustic

similarity to their associated target words was greater than any other concrete nouns or verbs that could be found. Two abstract nouns were used *(forsa,* "chance," for fortress; *shart,* "condition," for charter); one adjective *(baayiz,* "spoiled," for despise) and one adjective phrase *(intu wih-shiin,* "you are ugly," for intuition) were used; one question word *(fain,* "where," for reign) and one exclamation of greeting *(halo,* "hello," for allusion) were used.

Instruments

A 40-item, four-choice multiple-choice test was constructed to test retention. Each item consisted of a sentence requiring the use of one of the target words. The distracters were chosen from among the 40 words that were studied and were the same part of speech as the correct answer.

To test ability to retrieve the target words, a cued-recall measure was used. Subjects were asked to write the English definition of each of the 40 target words which were listed on a test paper.

Procedure

The six ELI classes were divided into three experimental groups: keyword, semantic, and keyword-semantic. In each treatment group there was one upper- and one lower-level class. Keyword classes were presented with the new word, its definition, and a keyword, as in Example 1.

	Keyword	Translation
1. CORD	/kora/	ball

a covered electrical wire

Please plug in the cord; I want to watch the news on television.
When bands set up their electrical instruments, the ground is covered with cords.
What connects the refrigerator to electricity?_____

Students were also given practice in making interactive images.

Semantic classes were given the new word, its definition, two examples of the word's use in sentences, and a question which they were required to answer using the new word, as in Example 2.

	Keyword	Translation
2. CORD	/kora/	ball

a covered electrical wire

The keyword-semantic classes received the new word, its definition, the keyword, and the example sentences and question. They also practiced making interactive images.

The students in each class received a day of instruction on how to use their method as well as a second day of instruction and practice testing. Instruction and testing took about 15 minutes each day. For the following 4 days, each class was given 5 minutes to learn 10 new words followed by an additional 5 minutes to do a cued-recall test.

The day after instruction ended, a comprehensive cued-recall test was given. On the following class day, all students took a comprehensive multiple-choice test. Nine days later, the cued-recall test was repeated, and the following day, the multiple-choice test was repeated. All tests were unannounced.

In order to maintain high student interest in instruction and testing, students were told that this was the first time. that their method had been used among native-Arabic-speaking students learning English and that their results would be carefully analyzed to see how helpful their method of instruction was. They were also told that they would get their training materials back in order to study the words for their AUC English exam.

Grading. Answers elicited on the cued-recall tests were evaluated by three raters. They used three criteria for determining acceptability of answers. First, synonyms were acceptable as long as they corresponded to the required definition; second, answers had to be the same part of speech as that of the required definition; and third, if raters thought the given definition was unclear but tended toward the correct answer, it was to be scored correct. Answers were counted correct if they had been accepted by two of the three raters.

RESULTS

Preexperiment Measures

Because intact groups were used, classes were compared on the basis of their total scores for the AUC entrance test as well as the vocabulary component of that test in order to check for preexperimental differences. While a difference between levels of proficiencies was to be expected, a main effect for experimental groups or an interaction between groups and level would warrant an analysis of covariance (ANCOVA). A 3 x 2 (Experimental Group X Proficiency Group) analysis of variance (ANOVA) was performed on each of the dependent variables using the Multiple General Linear Hypothesis Module in the SYSTAT package for PCs. Results

showed no significant main effect for experimental grouping for total score or for the vocabulary component, $F(2, 54) < 1$. In addition, no significant interaction between experimental groups and proficiency levels was found either for total score or the vocabulary component. Because no preexperimental differences were found, the use of an ANCOVA was not required.

Experimental Measure

In order to control for Type I errors, a MANOVA was performed on the data with the three cued-recall tests and the two administrations of the multiple-choice test used as dependent variables and Group and Level as the independent variables.

The MANOVA for the Group main effect was found to be significant, $F(10, 100) = 2.018$ (Wilks' lambda = 0.692), $p < .03$. As a result, the univariate ANOVAs were explored for each test. The effect for Group on the multiple-choice test given one day after the treatment was significant, $F(2, 54) = 5.583$, $p < .01$. A Newman-Keuls pair-wise comparison revealed only one significant result: The keyword-semantic group's performance was superior to that of the keyword group. The semantic group's performance was between these two, but the differences were not significant, $p = .10$. (See Table 1 for descriptive statistics.)

No other result for Group was significant at the $p < .05$ level. However, power[1] coefficients (Kirk, 1982, pp. 143-145) for the various univariate tests were computed and they were found to be low in several cases. For example, for the effect of Group on the Day 1 cued-recall test, the power coefficient equaled 0.50. According to Kirk (1982, p. 144) the recommended power is 0.80. One factor that affects power is the number of subjects used in the experiment. For Groups, the cell size is 20. To achieve a power of 0.80 for the same mean square values for the Day 1 cued-recall test, the required sample size was computed and the finding showed that cell sizes of 50 would be needed.

The MANOVA for the main effect, Levels, was statistically significant, $F(5, 50) = 12.798$ (Wilks' lambda = 0.439), $p < .0001$. The univariate ANOVAs for all of the tests were significant below the .01 level. However, these findings were to be expected and simply confirmed that the two levels were different in their English proficiency. The MANOVA for the interaction of Group by Levels was not statistically significant, $p < .18$.

1 For a discussion of statistical power, see Research Issues in this issue of the *TESOL Quarterly*.

TABLE 1
Descriptive Statistics for Test Type by Experimental Group and Proficiency Level

		Test type				
		Cued-recall			Multiple choice	
Group/level		Daily	Day 1	Day 9	Day 1	Day 9
Keyword						
Upper	*M*	32.20	15.30	14.90	31.00	31.30
	SD	5.43	7.89	8.43	5.60	5.74
Lower	*M*	29.20	6.30	6.70	17.30	19.90
	SD	6.61	3.70	3.95	4.17	5.65
Total	*M*	30.70	10.80	10.80	24.15	25.60
	SD	6.09	7.57	7.66	8.51	8.06
Semantic						
Upper	*M*	34.00	17.20	16.80	33.60	33.60
	SD	4.16	4.89	7.69	3.27	5.50
Lower	*M*	22.30	10.10	8.10	20.80	23.00
	SD	10.87	5.02	4.75	7.38	7.26
Total	*M*	28.15	13.65	12.45	27.20	28.30
	SD	10.01	6.05	7.65	8.60	8.30
Keyword-Semantic						
Upper	*M*	32.50	19.10	18.20	34.80	33.30
	SD	3.92	4.01	5.73	4.73	5.40
Lower	*M*	31.60	13.30	12.70	25.80	27.40
	SD	6.52	12.30	10.31	8.19	9.35
Total	*M*	32.05	16.20	15.45	30.30	30.35
	SD	5.26	9.39	8.59	7.98	8.02

Note. No. of subjects per cell = 10 (upper and lower), 20 (total).

DISCUSSION

Several findings provide possible insight for comparing the three strategies for learning ESL vocabulary addressed in this study. First, the strongest effect was found for the keyword-semantic method which produced significantly better results than the keyword method alone and was slightly better (although not significantly so) than the semantic method. These findings are especially interesting because they are consistent with the predictions made from the depths-of-processing theory (Craik & Lockhart, 1972; Craik & Tulving, 1975). That is, first, information processed at the semantic level produces better memory traces than that processed at acoustical and visual levels; and second, when elaboration occurs at a number of levels, memory traces are even stronger. If assumptions that recognition tasks measure information stored in memory and

cued-recall tasks reflect facility in retrieving information from memory are warranted (cf. Richardson-Klavehn & Bjork, 1988), then the conclusion can be made that the combination of these methods produced both stronger memory traces and better retrieval paths than if used alone.

Second, although upper-level students consistently outperformed those at the lower level, the retention of both was aided by the combined keyword-semantic strategy; the combined strategy significantly outperformed the keyword strategy alone. This pattern in the results suggests that the combined strategy worked better for differing proficiency levels and is therefore a possible choice for teachers in the selection of learning strategies with which to train their students.

Finally, the effects on retention by the combined methods was found in authentic classroom situations. This is important, because for a strategy to be useful, research must demonstrate that certain learning strategies are not only effective in the laboratory but in the classroom as well.

One possible factor that could have been instrumental in affecting these results was the level of concreteness of the target words. Many of the new vocabulary words themselves, such as *rectify* and *rival*, were fairly abstract and difficult to image. As a consequence, the use of abstract academic vocabulary may have encouraged the use of semantic strategies by keyword subjects (Hall, 1988; Hall et al., 1981; Raugh & Atkinson, 1975). In addition to many abstract target words, only 83% of the keywords were judged to be concrete (for a description of the remaining 17%, see the discussion of keywords in the Materials subsection). A follow-up questionnaire showed that some students seemed to switch to a keyword-semantic-like method rather than use images. One student reported remembering the meaning of the word *fortress*, whose keyword was *forsa*, meaning chance, by making a sentence, "Attacking the crusader's fortress was a chance for victory for Saladin." This student supplied a meaningful rather than sensory association between the meaning of the keyword and the meaning of the new word. If students systematically have difficulty generating images, this circumstance would restrict the keyword method to concrete words, limiting its usefulness in ESL classrooms, especially those for academic purposes. The findings of this study suggest combining the two strategies in order to provide the student with more versatility in mastering heterogeneous vocabulary.

An additional factor that could have played a role in the results of this study is the format in which the treatments were given. In many non-English L2 keyword studies, the subjects were native English

speakers. Consequently, the foreign word was presented, an English keyword supplied, and an English definition given. In this study, an English L2 word was given to nonnative English speakers, an Arabic keyword supplied, and an English definition given. The mismatch comes in the last step. To make this study more comparable, an Arabic definition should have been given. Future research should take this into consideration. However, the format used in this study is often the one used in the ESL classroom and therefore might be considered to have more ecological validity.

This study has provided some initial evidence to suggest that the keyword method in combination with the semantic processing method promotes more vocabulary acquisition than the keyword method used alone. The findings also indicate that using both simultaneously enables students of varying proficiencies to become more versatile in handling words with differing levels of concreteness. These are not the only strategies to be considered, however. Other strategies need to be compared as well, in order to gain an overall picture of the optimal use of learning strategies for vocabulary learning.

■

ACKNOWLEDGMENTS

This research is based in part on a master's thesis by the first author completed at the American University in Cairo. The authors wish to thank Mr. Abbas Tonsi, Ms. Sanaa Makhlouf, Mr. Yasser El Kasef, Mr. Maged Botros, Ms. Maysa Moslay, and Ms. Susan Esnawi for providing and editing keywords. In addition, we wish to thank Ms. Patricia Barnes, Ms. Hoda Garas, Ms. Sanaa Makhlouf, Ms. Marian Sarofim, Ms. Cynthia Sheikholeslami, Ms. Barbara Thornton, and their students for participating. Finally, we would like to thank Ms. Russanne Hozayin and Dr. Paul Stevens for their helpful editorial comments.

THE AUTHORS

Thomas Brown received an MA in TEFL at the American University in Cairo (AUC) in 1989 before teaching EFL in AUC'S Division of Consulting Instruction and Training. Currently he is developing a vocational ESL program at Micron Technology, Inc.

Fred L. Perry, Jr. is Associate Professor in the EFL Unit of the English Language Institute at the American University in Cairo, where he teaches research methodology and language testing. He is also head of the ELI Testing and Evaluation Unit. His research interests include language testing, cognitive factors in language learning and teaching, and internationally mobile children.

REFERENCES

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30,* 821-828.

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 126-133.

Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147-163). Hillsdale, NJ: Lawrence Erlbaum.

Benton, S. L., Glover, J. A., & Bruning, R. H. (1983). Levels of processing: Effect of number of decisions on prose recall. *Journal of Educational Psychology, 75,* 382-390.

Borg, W. R., & Gall, M. D. (1989). Educational research: An introduction (5th ed.). New York: Longman.

Channell, J. (1980). Applying semantic theory to vocabulary teaching. *English Language Teaching Journal, 35,* 115-122.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology, 104,* 268-294.

Crow, J. T., & Quigley, J. R. (1985). A semantic field approach to passive vocabulary acquisition for reading comprehension. *TESOL Quarterly, 19* (3), 497-513.

Desrochers, A., Gelinas, C., & Wieland, L. D. (1989). An application of the mnemonic keyword method to the acquisition of German nouns and their grammatical gender. *Journal of Educational Psychology, 81,* 25-32.

Eeds, M., & Cockrum, W. A. (1985). Teaching word meanings by expanding schema versus dictionary or reading context work. *Journal of Reading, 28,* 492-497.

Hall, J. W. (1988). On the utility of the keyword mnemonic for vocabulary learning. *Journal of Educational Psychology, 80,* 554-562.

Hall, J. W., Wilson, K. P., & Patterson, R. J. (1981). Mnemotechnics: Some limitations of the mnemonic keyword method for the study of foreign language vocabulary. *Journal of Educational Psychology, 73,* 345-357.

Johnson-Laird, P. N., Gibbs, G., & deMowbray, J. (1978). Meaning, amount of processing, and memory for words. *Memory & Cognition, 6,* 372-375.

Kirk, R. E. (1982). *Experimental design* (2nd ed.). Monterey, CA: Brooks/Cole.

Levin, J. R., Pressley, M., McCormick, C. B., Miller, G. E., & Schriberg, L. K. (1979). Assessing the classroom potential of the keyword method. *Journal of Educational Psychology, 71,* 583-594.

Martin, A. V. (1976). Teaching academic vocabulary to foreign graduate students. *TESOL Quarterly, 10,* 91-97.

McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method in context. *Journal of Educational Psychology, 76,* 598-609.

McDaniel, M. A., Pressley, M., & Dunay, P. K. (1987). Long-term retention of vocabulary after keyword and context learning. *Journal of Educational Psychology, 79,* 87-89.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20,* 522-535.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hills dale, NJ: Lawrence Erlbaum.

Oxford, R. (1986). *Second language learning strategies: Current research and implications for practice (TR3).* Los Angeles: University of California, Center for Language Education and Research.

Oxford-Carpenter, R. (1985). *A new taxonomy of second language learning strategies.* Washington, DC: CAL/ERIC Clearinghouse on Languages and Linguistics.

Perry, F. L., Jr. (1982). Test-like events: An aid to learning. *Singapore Journal of Education, 4,* 44-47.

Praninskas, J. (1972). *American University Word List.* London: Longman.

Pressley, M. (1977). Children's use of the keyword method to learn simple Spanish vocabulary words. *Journal of Educational Psychology, 69,* 465-472.

Pressley, M., Levin, J. R., & Delaney, H. D. (1982). The mnemonic keyword method. *Review of Educational Research, 52,* 61-91.

Pressley, M., Levin, J. R., Digdon, N., Bryant, S. L., & Ray, K. (1983). Does method of item presentation affect keyword method effectiveness? *Journal of Educational Psychology, 75,* 686-691.

Pressley, M., Levin, J. R., Hall, J. W., Miller, G. E., & Berry, J. K. (1980). The keyword method and foreign word acquisition. *Journal of Experimental* Psychology, 6, *163-173.*

Pressley, M., Levin, J. R., & McDaniel, M. A. (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 107-127). Hillsdale, NJ: Lawrence Erlbaum.

Pressley, M., Levin, J. R., Nakamura, G. V., Hope, D. J., Bispo, J. G., & Toye, A. R. (1980). The keyword method of foreign vocabulary learning: An investigation of its generalizability. *Journal of Applied Psychology, 65,* 635-644.

Pressley, M., Levin, J. R., Kuiper, N. A., Bryant, S. L., & Michener, S. (1982). Mnemonic versus non-mnemonic vocabulary learning strategies: Additional comparisons. *Journal of Educational Psychology, 74,* 693-707.

Raugh, M. R., & Atkinson, R. C. (1975). A mnemonic method for learning a second language vocabulary. *Journal of Educational Psychology, 67,* 1-16.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475-543.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Lawrence Erlbaum.

TESOL QUARTERLY, Vol. 25, No. 4, Winter 1991

What Does Language Testing Have to Offer?

LYLE F. BACHMAN
University of California, Los Angeles

Advances in language testing in the past decade have occurred in three areas: (a) the development of a theoretical view that considers language ability to be multicomponential and recognizes the influence of the test method and test taker characteristics on test performance, (b) applications of more sophisticated measurement and statistical tools, and (c) the development of "communicative" language tests that incorporate principles of "communicative" language teaching. After reviewing these advances, this paper describes an interfactional model of language test performance that includes two components, language ability and test method. Language ability consists of language knowledge and metacognitive strategies, whereas test method includes characteristics of the environment, rubric, input, expected response, and relationship between input and expected response. Two aspects of authenticity are derived from this model. The situational authenticity of a given test task depends on the relationship between its test method characteristics and the features of a specific language use situation, while its interfactional authenticity pertains to the degree to which it invokes the test taker's language ability. The application of this definition of authenticity to test development is discussed.

Since 1989, four papers reviewing the state of the art in the field of language testing have appeared (Alderson, 1991; Bachman, 1990a; Skehan, 1988, 1989, 1991). All four have argued that language testing has come of age as a discipline in its own right within applied linguistics and have presented substantial evidence, I believe, in support of this assertion. A common theme in all these articles is that the field of language testing has much to offer in terms of theoretical, methodological, and practical accomplishments to its sister disciplines in applied linguistics. Since these papers provide excellent critical surveys and discussions of the field of language testing, I will simply summarize some of the common themes in these reviews in Part 1 of this paper in order to whet the appetite

of readers who may be interested in knowing what are the issues and problems of current interest to language testers. These articles are nontechnical and accessible to those who are not themselves language testing specialists. Furthermore, Skehan (1991) and Alderson (1991) appear in collections of papers from recent conferences that focus on current issues in language testing. These collections include a wide variety of topics of current interest within language testing, discussed from many perspectives, and thus constitute major contributions to the literature on language testing.

The purpose of this paper is to address a question that is, I believe, implicit in all of the review articles mentioned above, What does language testing have to offer to researchers and practitioners in other areas of applied linguistics, particularly in language learning and language teaching? These reviews discuss several specific areas in which valuable contributions can be expected (e.g., program evaluation, second language acquisition, classroom learning, research methodology). Part 2 of this paper focuses on two recent developments in language testing, discussing their potential contributions to language learning and language teaching. I argue first that a theoretical model of second language ability that has emerged on the basis of research in language testing can be useful for both researchers and practitioners in language learning and language teaching. Specifically, I believe it provides a basis for both conceptualizing second language abilities whose acquisition is the object of considerable research and instructional effort, and for designing language tests for use both in instructional settings and for research in language learning and language teaching. Second, I will describe an approach to characterize the authenticity of a language task which I believe can help us to better understand the nature of the tasks we set, either for students in instructional programs or for subjects in language learning research and which can thus aid in the design and development of tasks that are more useful for these purposes.

PART 1: LANGUAGE TESTING IN THE 1990s

In echoing Alderson's (1991) title, I acknowledge the commonalities among the review articles mentioned above in the themes they discuss and the issues they raise. While each review emphasizes specific areas, all approach the task with essentially the same rhetorical organization: a review of the achievements in language testing, or lack thereof, over the past decade; a discussion of areas of likely continued development; and suggestions of areas in need

of increased emphasis to assure developments in the future. Both Alderson and Skehan argue that while language testing has made progress in some areas, on the whole "there has been relatively little progress in language testing until recently" (Skehan, 1991, p. 3). Skehan discusses the contextual factors—theory, practical considerations, and human considerations—that have influenced language testing in terms of whether these factors act as "forces for conservatism" or "forces for change" (p. 3). The former, he argues, "all have the consequence of retarding change, reducing openness, and generally justifying inaction in testing" (p. 3), while the latter are "pressures which are likely to bring about more beneficial outcomes" (p. 7). All of the reviews present essentially optimistic views of where language testing is going and what it has to offer other areas of applied linguistics. I will group the common themes of these reviews into the general areas of (a) theoretical issues and their implications for practical application, (b) methodological advances, and (c) language test development.

THEORETICAL ISSUES

One of the major preoccupations of language testers in the past decade has been investigating the nature of language proficiency. In 1980 the "unitary competence hypothesis" (Oller, 1979), which claimed that language proficiency consists of a single, global ability was widely accepted. By 1983 this view of language proficiency had been challenged by several empirical studies and abandoned by its chief proponent (Oller, 1983). The unitary trait view has been replaced, through both empirical research and theorizing, by the view that language proficiency is multicomponential, consisting of a number of interrelated specific abilities as well as a general ability or set of general strategies or procedures. Skehan and Alderson both suggest that the model of language test performance proposed by Bachman (1990b) represents progress in this area, since it includes both components of language ability and characteristics of test methods, thereby making it possible "to make statements about actual performance as well as underlying abilities" (Skehan, 1991, p. 9). At the same time, Skehan correctly points out that as research progresses, this model will be modified and eventually superseded. Both Alderson and Skehan indicate that an area where further progress is needed is in the application of theoretical models of language proficiency to the design and development of language tests. Alderson, for example, states that "we need to be concerned not only with . . . the nature of language proficiency, but also with language learning and the design and researching of achievement

tests; not only with testers, and the problems of our professionalism, but also with testees, with students, and their interests, perspectives and insights" (Alderson, 1991, p. 5).

A second area of research and progress is in our understanding of the effects of the method of testing on test performance, A number of empirical studies conducted in the 1980s clearly demonstrated that the kind of test tasks used can affect test performance as much as the abilities we want to measure (e.g., Bachman & Palmer, 1981, 1982, 1988; Clifford, 1981; Shohamy, 1983, 1984). Other studies demonstrated that the topical content of test tasks can affect performance (e.g., Alderson & Urquhart, 1985; Erickson & Molloy, 1983). Results of these studies have stimulated a renewed interest in the investigation of test content. And here the results have been mixed. Alderson and colleagues (Alderson, 1986, 1990; Alderson & Lukmani, 1986; Alderson, Henning, & Lukmani, 1987) have been investigating (a) the extent to which "experts" agree in their judgments about what specific skills EFL reading test items measure, and at what levels, and (b) whether these expert judgments about ability levels are related to the difficulty of items. Their results indicate first, that these experts, who included test designers assessing the content of their own tests, do not agree and, second, that there is virtually no relationship between judgments of the levels of ability tested and empirical item difficulty. Bachman and colleagues, on the other hand (Bachman, Davidson, Lynch, & Ryan, 1989; Bachman, Davidson, & Milanovic, 1991; Bachman, Davidson, Ryan, & Choi, in press) have found that by using a content-rating instrument based on a taxonomy of test method characteristics (Bachman, 1990b) and by training raters, a high degree of agreement among raters can be obtained, and such content ratings are related to item difficulty and item discrimination. In my view, these results are not inconsistent. The research of Alderson and colleagues presents, I believe, a sobering picture of actual practice in the design and development of language tests: Test designers and experts in the field disagree about what language tests measure, and neither the designers nor the experts have a clear sense of the levels of ability measured by their tests. This research uncovers a potentially serious problem in the way language testers practice their trade. Bachman's research, on the other hand, presents what can be accomplished in a highly controlled situation, and provides one approach to solving this problem. Thus, an important area for future research in the years to come will be in the refinement of approaches to the analysis of test method characteristics, of which content is a substantial component, and the investigation of how specific characteristics of test method affect test

performance. Progress will be realized in the area of language testing practice when insights from this area of research inform the design and development of language tests. The research on test content analysis that has been conducted by the University of Cambridge Local Examinations Syndicate, and the incorporation of that research into the design and development of EFL tests is illustrative of this kind of integrated approach (Bachman et al., 1991),

The 1980s saw a wealth of research into the characteristics of test takers and how these are related to test performance, generally under the rubric of investigations into potential sources of test bias; I can do little more than list these here. A number of studies have shown differences in test performance across different cultural, linguistic or ethnic groups (e.g., Alderman & Holland, 1981; Chen & Henning, 1985; Politzer & McGroarty, 1985; Swinton & Powers, 1980; Zeidner, 1986), while others have found differential performance between sexes (e.g., Farhady, 1982; Zeidner, 1987). Other studies have found relationships between field dependence and test performance (e.g., Chapelle, 1988; Chapelle & Roberts, 1986; Hansen, 1984; Hansen & Stansfield, 1981; Stansfield & Hansen, 1983). Such studies demonstrate the effects of various test taker characteristics on test performance, and suggest that such characteristics need to be considered in both the design of language tests and in the interpretation of test scores. To date, however, no clear direction has emerged to suggest how such considerations translate into testing practice. Two issues that need to be resolved in this regard are .(a) whether and how we assess the specific characteristics of a given group of test takers, and (b) whether and how we can incorporate such information into the way we design language tests. Do we treat these characteristics as sources of test bias and seek ways to somehow "correct" for this in the way we write and select test items, for example? Or, if many of these characteristics are known to also influence language learning, do we reconsider our definition of language ability? The investigation of test taker characteristics and their effects on language test performance also has implications for research in second language acquisition (SLA), and represents what Bachman (1989) has called an "interface" between SLA and language testing research.

METHODOLOGICAL ADVANCES

Many of the developments mentioned above—changes in the way we view language ability, the effects of test method and test taker characteristics—have been facilitated by advances in the tools that are available for test analysis. These advances have been in

three areas: psychometrics, statistical analysis, and qualitative approaches to the description of test performance. The 1980s saw the application of several modern psychometric tools to language testing: item response theory (IRT), generalizability theory (G theory), criterion-referenced (CR) measurement, and the Mantel-Haenszel procedure. As these tools are fairly technical, I will simply refer readers to discussions of them: IRT (Henning, 1987), G theory (Bachman, 1990b; Bolus, Hinofotis, & Bailey, 1982), CR measurement (Bachman, 1990b; Hudson & Lynch, 1984), Mantel-Haenszel (Ryan & Bachman, in press). The application of IRT to language tests has brought with it advances in computer-adaptive language testing, which promises to make language tests more efficient and adaptable to individual test takers, and thus potentially more useful in the types of information they provide (e.g., Tung, 1986), but which also presents a challenge not to complacently continue using familiar testing techniques simply because they can be administered easily via computer (Canale, 1986). Alderson (1988a) and the papers in Stansfield (1986) provide extensive discussions of the applications of computers to language testing.

The major advance in the area of statistical analysis has been the application of structural equation modeling to language testing research. (Relatively nontechnical discussions of structural equation modeling can be found in Long, 1983a, 1983b.) The use of confirmatory factor analysis was instrumental in demonstrating the untenability of the unitary trait hypothesis, and this type of analysis, in conjunction with the multitrait/multimethod research design, continues to be a productive approach to the process of construct validation. Structural equation modeling has also facilitated the investigation of relationships between language test performance and test taker characteristics (e.g., Fouly, 1985; Purcell, 1983) and different types of language instruction (e.g., Sang, Schmitz, Vollmer, Baumert, & Roeder, 1986).

A third methodological advance has been in the use of introspection to investigate the processes or strategies that test takers employ in attempting to complete test tasks. Studies using this approach have demonstrated that test takers use a variety of strategies in solving language test tasks (e.g., Alderson, 1988c; Cohen, 1984) and that these strategies are related to test performance (e.g., Anderson, Cohen, Perkins, & Bachman, 1991; Nevo, 1989).

Perhaps the single most important theoretical development in language testing in the 1980s was the realization that a language test score represents a complexity of multiple influences. As both Alderson and Skehan point out, this advance has been spurred on, to a considerable extent, by the application of the methodological

tools discussed above. But, as Alderson (1991) notes, "the use of more sophisticated techniques reveals how complex responses to test items can be and therefore how complex a test score can be" (p. 12). Thus, one legacy of the 1980s is that we now know that a language test score cannot be interpreted simplistically as an indicator of the particular language ability we want to measure; it is also affected to some extent by the characteristics and content of the test tasks, the characteristics of the test taker, and the strategies the test taker employs in attempting to complete the test task. What makes the interpretation of test scores particularly difficult is that these factors undoubtedly interact with each other. The particular strategy adopted by a given test taker, for example, is likely to be a function of both the characteristics of the test task and the test taker's personal characteristics. This realization clearly indicates that we need to consider very carefully the interpretations and uses we make of language test scores and thus should sound a note of caution to language testing practitioners. At the same time, our expanded knowledge of the complexity of language test performance, along with the methodological tools now at our disposal, provide a basis for designing and developing language tests that are potentially more suitable for specific groups of test takers and more useful for their intended purposes.

ADVANCES IN LANGUAGE TEST DEVELOPMENT

For language testing, the 1980s could be characterized as the decade of "communicative" testing. Although two strains of communicative approaches to language testing can be traced, as with many innovations in language testing over the years, the major impetus has come from language teaching. One strain of communicative tests, illustrated by the *Ontario Assessment Pool* (Canale & Swain, 1980a) and the A *Vous la Parole* testing unit described by Swain (1985), traces its roots to the Canale/Swain framework of communicative competence (Canale, 1983; Canale & Swain, 1980b). The other, exemplified by the *Test of English for Educational Purposes* (Associated Examining Board, 1987; Weir, 1983), the *Ontario Test of English as a Second Language* (Wesche et al., 1987), and the international English Language *Testing Service* (e.g., Alderson, 1988b; Alderson, Foulkes, Clapham, & Ingram, 1990; Criper & Davies, 1988; Seaton, 1983) has grown out of the English for specific purposes tradition. While a number of lists of characteristics of communicative language tests has been proposed (e.g., Alderson, 1981a; Canale, 1984; Carroll, 1980; Harrison, 1983; Morrow, 1977, 1979), I will mention four characteristics that would

appear to distinguish communicative language tests. First, such tests create an "information gap," requiring test takers to process complementary information through the use of multiple sources of input. Test takers, for example, might be required to perform a writing task that is based on input from both a short recorded lecture and a reading passage on the same topic. A second characteristic is that of task dependency, with tasks in one section of the test building upon the content of earlier sections, including the test taker's answers to those sections. Third, communicative tests can be characterized by their integration of test tasks and content within a given domain of discourse. Finally, communicative tests attempt to measure a much broader range of language abilities—including knowledge of cohesion, functions, and sociolinguistic appropriateness—than did earlier tests, which tended to focus on the formal aspects of language—grammar, vocabulary, and pronunciation.

A different approach to language testing that evolved during the 1980s is the adaptation of the FSI oral interview guidelines (Wilds, 1975) to the assessment of the oral language proficiency in contexts outside agencies of the U.S. government. This "AEI" (For *A*merican Council for the Teaching of Foreign Languages/ *E*ducational Testing Service/ *I*nteragency Language Roundtable) approach to language assessment is based on a view of language proficiency as a unitary ability (Lowe, 1988), and thus diverges from the view that has emerged in language testing research and other areas of applied linguistics. This approach to oral language assessment has been criticized by both linguists and applied linguists, including language testers and language teachers, on a number of grounds (e. g., Alderson, 1981b; Bachman, 1988; Bachman & Savignon, 1986; Candlin, 1986; Kramsch, 1986; Lantolf & Frawley, 1985, 1988; Savignon, 1985). Nevertheless, the approach and ability levels defined have been widely accepted as a standard for assessing oral proficiency in a foreign language in the U.S. and have provided the basis for the development of "simulated oral proficiency interviews" in various languages (e.g., Stansfield & Kenyon, 1988, 1989). In addition, the approach has been adapted to the assessment of EFL proficiency in other countries (e.g., Ingram, 1984).

These two approaches to language assessment—communicative and AEI—are based on differing views of the nature of language proficiency, and are thus likely to continue as separate, unrelated approaches in the years to come. Lowe (1988) has explicitly articulated such a separatist view, in stating that the "concept of Communicative Language Proficiency (CLP), renamed Communicative Language Ability (CLA), and AEI proficiency may prove

incompatible" (p. 14). Communicative language testing and AEI assessment represent two different approaches to language test design, and each has developed a number of specific manifestations in language tests. As a result, language testing will be enriched in the years to come by the variety of tests and testing techniques that emerge from these approaches.

This summary has focused on common areas among four recent reviews of language testing. In addition to these common areas, each of the reviews mentions specific areas of progress or concern. Skehan (1991) and Alderson (1991) both note that until very recently other areas of applied linguistics have provided very little input into language testing. Skehan, however, is encouraged by the relevance to language testing of recent work in sociolinguistics, second language acquisition, and language teaching, and points out the need for language testing to be aware of and receptive to input from developments in other areas of applied linguistics such as the SLA-based approach to assessing language development of Pienemann, Johnston, & Brindley (1988). Skehan and Alderson both argue that language testing must continue to investigate new avenues to assessment, such as formats that measure communicative abilities more successfully (e. g., Milanovic, 1988); "series tasks," in which specified language interactions are scored in terms of how particular aspects of information are communicated; group testing; self-assessment; and computer-based language testing. Alderson discusses two additional areas to which language testing needs to turn its attention in the years to come: "washback" effects and learner-centered testing. He points out that while we generally assume that tests have an impact on instruction (washback), there is virtually no empirical research into how, if at all, instructional impact functions, under what conditions, and whether deliberate attempts to design tests with positive instructional impact are effective. Alderson also argues persuasively for the greater involvement of learners in the activity of testing, in the design and writing of tests, and in the setting of standards for success. In this regard, I would mention the work of Brindley (1989) in assessing language achievement in learner-centered instructional settings and the papers in de Jong & Stevenson (1990), which address issues in individualizing language testing. A final area of development, mentioned by Bachman (1990b), is the renewed interest in language aptitude and developments in both the definition of the theoretical construct and in approaches to its measurement (Perry & Stansfield, 1990).

As a result of the developments of the 1980s, language testing has emerged as a discipline in its own right within applied linguistics.

Alderson (1991) notes that since 1980 language testing has seen the creation of an internationally respected journal, *Language Testing*, as well as several regular newsletters; five new texts on language testing as well as over a dozen volumes of collected papers have been published; and there are now at least two regular major international conferences each year devoted to language testing. The field of language testing has seen the development of both a model of language test performance that can guide empirical research, and the application of a variety of research approaches and tools to facilitate such research. In sum, language testing can now claim its own research questions and research methodology. As Bachman (1990a) states, "perhaps for the first time in the history of language testing it is possible to see a genuine symbiotic relationship between applied linguistic theory and the tools of empirical research as they are applied to both the development and the examination of a theory of performance on language tests [and to] the development and use of better language tests" (p. 220).

Also as a result of developments in the past decade, language testing is in a better position, I believe, both to make contributions to its sister disciplines in applied linguistics and to be enriched by developments in those disciplines. The next part of this paper briefly describes what I consider two contributions that language testing has to offer to the areas of language learning and language teaching.

PART 2: AN INTERACTIONAL APPROACH TO LANGUAGE TEST DEVELOPMENT

Language tests are used for a variety of purposes; these can be grouped into two broad categories. First, the results of language tests may be used to make inferences about test takers' language abilities or to make predictions about their capacity for using language to perform future tasks in contexts outside the test itself. Second, decisions (e.g., selection, diagnosis, placement, progress, grading, certification, employment) may be made about test takers on the basis of what we infer from test scores about their levels of ability or their capacity for nontest language use. A major consideration in both the design and use of language tests, therefore, is the extent to which the specific test tasks we include elicit instances of language use from which we can make such inferences or predictions. What this implies is that in order to investigate and demonstrate the validity of the uses we make of test scores, we need a theoretical framework within which we can describe language test performance as a specific instance of

language use. Specifically, in order to make inferences about levels or profiles of ability, or predictions about capacity for using language to perform future tasks in nontest language use contexts, we need to demonstrate two kinds of correspondences: (a) that the language abilities measured by our language tests correspond in specifiable ways to the language abilities involved in nontest language use, and (b) that the characteristics of the test tasks correspond to the features of a target language use context.

In an instructional setting, for example, in which we may want to use a test to measure learners' degrees of mastery of different components of language ability that have been covered in the curriculum, we need to demonstrate that the content of the test is representative of the content of the course. Specifically, we will want to demonstrate that the components of language ability included in the test correspond to those covered in the course and that the characteristics of the test tasks correspond to the types of classroom learning activities included in the program. Demonstrating correspondences such as these provides some justification for interpreting test scores as evidence of levels of ability in the different components tested.

Another example would be a situation in which we need to select individuals for possible employment in a job which requires a specified level of proficiency in a foreign language. In this case, we need to demonstrate that the tasks included in the test are representative of the language use tasks required by the future job. Demonstrating this correspondence provides some justification for using the test scores to predict future capacity for using the foreign language effectively in the target employment situation.

Demonstrating correspondences between test performance and language use is equally important for justifying the use of language tests in applied linguistics research. For example, if we were interested in investigating the interlanguage development of a specific component of ability in a target language, for example, sensitivity to appropriate register, and wanted to use a test as one of our research instruments, we would need to be sure that the test we used measured this aspect of language ability. Similarly, we would want to specify the characteristics of the tasks included in the test, so as to minimize any variations that may arise between performance on this test and other elicitation procedures we may want to use.

In this part of the paper I will present a framework that I believe provides a basis for relating test performance to nontest language use. This framework includes a model of language ability for describing the abilities involved in language use and test performance and a framework of test method characteristics for relating

the characteristics of tests and test tasks to features of the language use context. I will then suggest how this framework can be used to clarify our thinking about the notion of authenticity and for designing test tasks that are authentic.

LANGUAGE ABILITY

The language ability of the language user is one feature of language use. When we design a language test, we hypothesize that the test taker's language ability will be engaged by the test tasks. Thus, in order to relate the abilities we believe are involved in test performance to the abilities involved in language use, we need a model of language ability. The model I will describe here is a refinement of my 1990 model that Adrian Palmer and I are developing (Bachman & Palmer, in press). We define language ability essentially in Widdowson's (1983) terms as the capacity for using the knowledge of language in conjunction with the features of the language use context to create and interpret meaning. Our model of language ability includes two types of components: (a) areas of language knowledge, which we would hypothesize to be unique to language use (as opposed to, for example, mathematical knowledge or musical knowledge), and (b) metacognitive strategies that are probably general to all mental activity.

This view of language ability is consistent with research in applied linguistics that has increasingly come to view language ability as consisting of two components: (a) language knowledge, sometimes referred to as competence, and (b) cognitive processes, or procedures, that implement that knowledge in language use (e.g., Bachman, 1990a; Bialystok, 1990; Spolsky, 1989; Widdowson, 1983). It is also consistent with information-processing, or cognitive, models of mental abilities, which also distinguish processes or heuristics from domains of knowledge (e.g., Sternberg, 1985, 1988).

Language use involves the integration of multiple components and processes, not the least of which are those that constitute language ability. It is unlikely that every language test we develop or use will be intended to measure all the components in our model. Nevertheless, even though we may be interested in focusing on only one or a few of these in a given testing context, we need to be aware of the full range of language abilities as we design and develop language tests and interpret language test scores, For example, even though we may only be interested in measuring an individual's knowledge of vocabulary, the kinds of test items, tasks or texts we use need to be selected with an awareness of what other

components of language ability they may evoke. We believe, therefore, that even though a given language test may focus on a narrow range of language abilities, its design must be informed by a broad view of language ability.

Language Knowledge[1]

What we refer to as language knowledge can be regarded as a domain of information that is specific to language ability and that is stored in long-term memory. For our purposes, we do not attempt to characterize how this knowledge is stored. That is, we use the term knowledge to refer to both conscious and tacit, analyzed and unanalyzed knowledge. While the importance of such distinctions has been recognized in other areas of applied linguistics, it remains to be seen how relevant they are to the design, development, and use of language tests.

Language knowledge includes two broad areas: organizational knowledge and pragmatic knowledge. These are constantly changing, as new elements are learned or acquired, and existing elements restructured. The learning or acquisition of areas of language knowledge is beyond the scope of my discussion here, and for purposes of describing how they pertain to language use, I will treat them as more or less stable traits or constructs. The areas of language knowledge are given in Figure 1 below.

Discussion of these elements of language knowledge is beyond the scope of this paper. I would simply indicate that this model of language ability has evolved from earlier models, particularly that of Canale & Swain (Canale, 1983; Canale & Swain, 1980b) as a result of both empirical research and review of relevant literature in applied linguistics. The model presented here thus includes a much wider range of elements and provides a more comprehensive view of language ability than have earlier models.

Strategic Competence

The second component of language ability is what I have called strategic competence, and have described as consisting of three sets

1 This description of *language knowledge* is essentially the same as Bachman's (1990b) discussion of *language competence*. The change in terminology from *competence* to *knowledge* reflects the view that the former term now carries with it a great deal of unnecessary semantic baggage that makes it less useful conceptually than it once was. I would note two changes from Bachman's 1990 model: (a) "Vocabulary" has been removed from "organizational competence" and placed within a new area, "propositional knowledge," under "pragmatic knowledge," and (b) "illocutionary competence" has been renamed "functional knowledge."

FIGURE 1

Areas of Language Knowledge

From Language Testing in Practice by L. F. Bachman and A. S. Palmer, in press, Oxford, Oxford University Press. Copyright by Oxford University Press. Reprinted by permission.

- **Organizational knowledge**
  (Determines how texts—oral or written—are organized)
  - **Grammatical knowledge**
    (Determines how individual utterances or sentences are organized)
  - **Textual knowledge**
    (Determines how utterances or sentences are organized to form texts)
- **Pragmatic knowledge**
  (Determines how utterances/sentences, intentions and contexts are related to form meaning)
  - **Propositional knowledge**
    (Determines how utterances/sentences are related to propositional content)
  - **Functional knowledge**
    (Determines how utterances/sentences are related to intentions of language users)
  - **Sociolinguistic knowledge**
    (Determines how utterances/sentences are related to features of the language use context)

of processes: assessment, planning, and execution (Bachman, 1990b). In applying this model to the practical design and development of language tests, Adrian Palmer and I have refined this view of strategic competence as consisting of three sets of metacognitive strategies: assessment, goal setting, and planning. These are given in Figure 2 below.

Since I will be referring to these in the discussion of interfactional authenticity below, I briefly discuss these metacognitive strategies here. First, however, I would point out a dilemma in describing a set of strategies that we hypothesize operate simultaneously and are thus essentially unordered. In situated language use, the metacognitive strategies and areas of language knowledge interact with each other simultaneously. Thus there is no particular ordering or sequencing in the way they operate. Furthermore, the strategies and areas of language knowledge are integrated and interactive, by which I mean that all the components of language ability, although distinct from each other, interact with each other and are fully integrated in any instance of language use. However, since the language used to describe the model is linear, only one strategy can be described at a time; I must begin with one and end with another, This dilemma applies to examples, as well, which must necessarily be described in terms of a sequence of events. However, if the reader will keep in mind that my purpose here is to provide a

FIGURE 2

Metacognitive Strategies

From *Languzge Testing in Practice* by L. F. Bachman and A. S. Palmer, in press, Oxford, Oxford University Press. Copyright by Oxford University Press. Reprinted by permission.

Assessment (Taking stock of what you need, what you have to work with, and how well you've done)

1. Determining the desirability of achieving a particular goal and what is needed to achieve it in a particular context
2. Determining what knowledge components—language knowledge, schemata, and affective schemata—are available for accomplishing that goal
3. Determining the extent to which the communicative goal has been achieved by a given utterance

Goal-setting (Deciding what you're going to do)

1. Identifying and selecting one or more communicative goals that you want to achieve
2. Deciding to attempt or not attempt achieving the communicative goal selected

Planning (Deciding how to use what you have)

1. Selecting relevant areas of language knowledge for accomplishing the given communicative goal
2. Formulating a plan for implementing these areas in the production or interpretation of an utterance

general conceptualization for guiding test development and use, rather than to give a detailed description of how language is processed in the mind, I believe that this dilemma in presentation will not be a problem.

Assessment. The strategies of assessment provide a direct link between the context in which language use takes place, the discourse that is used, and the areas of language knowledge that the language user employs in producing or interpreting utterances. Assessment strategies perform three types of functions:

1. Assessing features of the context to determine whether it is feasible to achieve a given goal and if feasible, what is needed to achieve it in a particular context
2. Assessing what areas of language knowledge are available for accomplishing that goal
3. Assessing the extent to which the communicative goal has been achieved

In performing these functions, assessment strategies draw upon and interact with the different areas of language knowledge as well as with real-world knowledge schemata and affective schemata.

Goal setting. From the perspective of the language user, goal setting involves, essentially, deciding what you are going to do, and includes the following functions:

1. Identifying a set of possible communicative goals
2. Choosing one or more goals from this set of possible goals
3. Deciding whether or not to attempt to achieve the goal(s)

Since one of the primary advantages of a language test, as opposed to other ways of obtaining information about an individual's language ability (such as naturalistic observation), is that it is designed to elicit a specific sample of language use, the test taker's flexibility in setting goals for performance on test tasks is necessarily limited. Thus, even though test takers may have some flexibility in setting goals for test performance, this is generally not as much as language users enjoy in nontest language use. As I will argue below, however, one way to increase the degree of interfactional authenticity of a language test task is to increase the test taker's involvement in goal setting.

Planning. Strategies of planning involves the following:

1. Selecting the relevant areas of language knowledge for accomplishing the given communicative goal
2. Formulating a plan for implementing these areas in the production or interpretation of an utterance

Strategies in Language Use

Using language involves interpreting or producing utterances with propositional content, functional purpose, and contextual appropriateness. This involves all the strategies and areas of language knowledge simultaneously and interactively. As discussed above, each of the three strategies interacts with all of the areas of language knowledge. In addition, the strategies themselves function interactively. Consider goal setting and assessment: We may modify or abandon a particular goal on the basis of assessment strategies. If, for example, we wanted to invite someone over for dinner, and saw that person at a party, we might decide to speak to them. If, however, we determined that it was inappropriate to extend the invitation in the presence of another person whom we did not want to invite, we would most likely modify our communicative goal and engage in polite conversation until the opportunity to extend the dinner invitation arose.

One implication of this model is that variations in language ability can be attributed to two sources. First, the areas of language

knowledge may vary over time and across different language users, so that language knowledge may contain varying combinations of elements from the native language, interlanguage, and the target language. These areas of knowledge may vary both in terms of the presence *or* absence of different elements and in the nature of the elements that are present. Second, the metacognitive strategies may vary over time and across different language users. Thus these strategies may be used more or less effectively by different individuals in completing a given language test task or by the same individual in completing different test tasks. Furthermore, the strategies may be used in differing proportions and in differing degrees for different language test tasks. This means that test takers' performance on a given test task reflects both their knowledge of the language elements being measured and their capacity for effectively relating this language knowledge to the characteristics of the test task so as to arrive at a successful solution to the problem posed.

TEST METHOD CHARACTERISTICS

So far I have described a model of language ability that consists of areas of language knowledge and metacognitive processes. I have argued that this model can provide a basis for demonstrating the correspondences between the abilities measured by the test and the abilities required for nontest language use.

The second correspondence we need to demonstrate in order to make inferences about abilities or to make predictions about future language use on the basis of language test scores is the correspondence between the characteristics of the test task and the features of a target language use context. It is widely recognized that the features of the language use context, such as the relationship between the language users, the topic, and the purpose, influence the way we use language. It would thus not be surprising to find that characteristics of the method of testing affect the way individuals perform on tests. Indeed, one of the major findings of language testing research over the past decade is that performance on language tests is affected not only by the ability we are trying to measure but also by the method we use to measure it (cf. references in the Theoretical Issues section of Part 1). In a language test, the instances of language use that we elicit are shaped, as it were, by the test tasks we present. Thus, it is the discourse that is included in the test question and the nature of the test task that will determine, to a large extent, how the test taker processes the information presented and responds to the particular task.

Language teachers also realize that the selection of a testing method is important. Frequently one of the first questions asked in testing classes or at conferences is the speaker's opinion of the "best" way to test a particular component of language ability. Teachers are clearly aware that the way they test language ability affects how their students perform on language tests and hence the quality of the information obtained from their tests. These teachers also want to be sure that this test method effect works in ways that will be fair to their students and that will enable them, the teachers, to make inferences about the language abilities they want to measure.

When we consider the different test methods that are commonly used for language tests, we realize that they are not single wholes, but rather collections of characteristics. "Multiple-choice" test items, for example, vary in a number of ways, such as in their length, syntactic complexity, level of vocabulary, topical content, and type of response required, to name but a few. Similarly, the "composition" test method encompasses a wide variety of prompts that can differ in terms of characteristics such as the intended audience, purpose, and specific organizational pattern requested. It is thus clear that we cannot be very precise in our thinking about test methods if we think of them only as holistic types. In order to organize our thinking about test methods, therefore, we first need a way to characterize them. A framework for doing this is presented in Figure 3 below.

Again, given space limitations, I will not discuss these here, but will simply point out that these test method characteristics have been used in several empirical studies to analyze the content of test items and to investigate the relationships between test content and the difficulty and discrimination of test items (Bachman, Kunnan, Vanniarajan, & Lynch, 1988; Bachman et al., 1989; Bachman et al., 1991; Bachman et al., in press).

CHARACTERIZING AUTHENTICITY

I have argued that in order to justify the use of language tests, we need to be able to demonstrate that performance on language tests corresponds in specified ways to nontest language use. I have described a model of language ability that I believe provides a basis for specifying the language ability of the language user as test taker. I have also presented a framework for specifying the characteristics of test tasks in a way that I believe enables us to investigate how these are related to the features of a given target language use context. In the remainder of this paper I will attempt to describe how these models can be used to assess the "authenticity" of a given

FIGURE 3

Test Method Characteristics

From *Fundamental Considerations in Language Testing* (p. 119) by L. F. Bachman, 1990, Oxford, Oxford University Press. Copyright 1990 by Oxford University Press. Reprinted by permission.

- Characteristics of the testing environment
- Characteristics of the test rubric
  - Test organization
  - Time allocation
  - Scoring
  - Instructions
- Characteristics of the test input
  - Format
  - Language of input
    - Length
    - Organizational characteristics
    - Pragmatic characteristics
- Characteristics of the expected response
  - Format
  - Language of expected response
    - Length
    - Organizational characteristics
    - Pragmatic characteristics
- Relationship between input and response
  - Reciprocal
  - Nonreciprocal
  - Adaptive

test task, and suggest some ways this notion of authenticity can be used in the practical design of test tasks.

A number of language testing specialists have discussed the features that characterize an authentic test task (e.g., Alderson, 1981b; Canale, 1984; Carroll, 1980; Harrison, 1983; Morrow, 1977, 1979). When we try to define *authenticity,* however, we notice that it is one of those words like *real* (as in "He's really real") that sounds good but leaves us wondering exactly what it means. In fact, in language testing circles, authenticity has been defined in a number of different ways. One approach, for example has been to define it as direct, in the sense of getting at the ability without going through an intermediate representation of the ability. However, it is impossible to directly observe the neurological programming in the brain that may account for language ability. Thus, all language tests are indirect; they simply permit us to observe behavior from which we can make inferences about language ability.

A second approach has been to define authenticity in terms of similarity to real life. The problem with this definition is that "real life" language use consists of an infinite set of unique and widely varied speech events. Thus, the way in which language is used in real life can vary enormously and includes language use situations as different as announcing a sporting event and keying information from a questionnaire into a computer. Because of the great variation in the language used in real life, we have no basis for knowing what kind of real life language tasks to use as our primary criteria for authenticity.

A third approach to defining authenticity is to appeal to what was once called face validity, which is really nothing more than face appeal. This is just as problematic as the previous two approaches, but in a different way. This definition refers to a purely subjective response on the part of the evaluator and offers us no criteria for use in creating appealing tests. Moreover, there is also the issue, noted by Davies (1977) some years ago, that what is appealing to experts in language testing might be different from what is appealing to teachers, students, or to parents of students.

All three of these approaches to authenticity capture some intuitively useful aspects of authenticity, but they are problematic in many ways. Furthermore, they are not clearly enough defined, I believe, to provide a basis for test development. Adrian Palmer and I (Bachman & Palmer, in press) have attempted to define authenticity in a way that we think still captures the spirit of these approaches but does so in a way that avoids the problems of these approaches and is also specific enough to guide test development. For this purpose, we define two different types of authenticity: situational authenticity and interfactional authenticity.

Situational Authenticity

We define situational authenticity as the perceived relevance of the test method characteristics to the features of a specific target language use situation. Thus, for a test task to be perceived as situationally authentic, the characteristics of the test task need to be perceived as corresponding to the features of a target language use situation. For example, one set of test method characteristics relates to certain characteristics of vocabulary (e.g., infrequent, specialized) and topics (e.g., academic, technical) included in the test input. If test takers were specialists in engineering, it is likely that inclusion of technical terms and topics from engineering would tend to increase the situational authenticity of the test. While situational authenticity may appear to be essentially the same as the real-life

approach described above, it is fundamentally different in a critical way. The real-life approach assumes that there is a well-defined domain of target language use tasks outside the test itself and that these tasks themselves can be sampled in order to achieve authenticity. In contrast, we define the situational authenticity of a given test task in terms of the distinctive features that characterize a set of target language use tasks. Thus, in designing a situationally authentic test, we do not attempt to sample actual tasks from a domain of nontest language use, but rather try to design tasks that have the same critical features as tasks in that domain.

This definition allows for possibly different perceptions of situational authenticity. Thus, different test takers may have different ideas about their target language use situations. And the perceptions of test takers about the relevance of the characteristics of the test task to their target language use situations may be different from those of the test developers. What this implies is that situational authenticity must be assessed from a number of perspectives and that these all must be taken into consideration in the development and use of language tests.

Language testers and teachers alike are concerned with this kind of authenticity, for we all want to do our best to make our teaching and testing relevant to our students' language use needs. For a reading test, for example, we are likely to choose a passage whose topic and genre (characteristics of the test input) match the topic and genre of material the test user is likely to read outside of the testing situation. Or, if the target language use situation requires reciprocal language use, then we will design a test task in which reciprocity is a characteristic of the relationship between test input and expected response.

Interfactional Authenticity

What we call interfactional authenticity is essentially Widdowson's (1978) definition of authenticity and is a function of the extent and type of involvement of task takers' language ability in accomplishing a test task. The different areas of language knowledge and the different strategies can be involved to varying degrees in the problem presented by the test task. In contrast to situational authenticity, where the focus is on the relationship between the test task and nontest language use, interfactional authenticity resides in interaction between the test taker and the test task.

In order to make these definitions of authenticity useful for test development, we need to be able to specify both the characteristics of the test task and the nature of the involvement of the test taker's

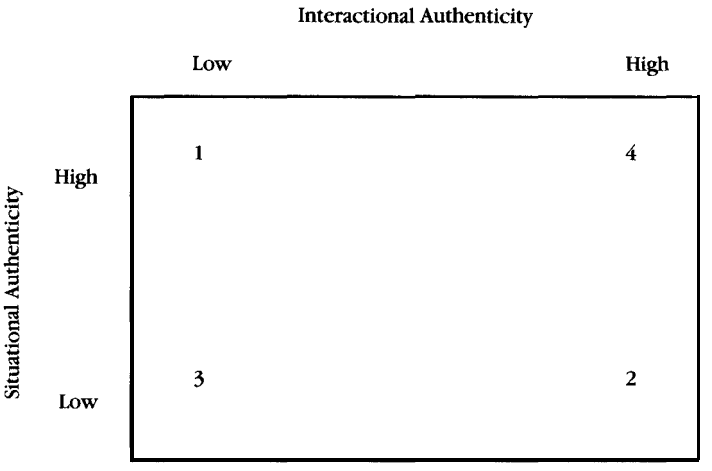
language ability. We propose that the situational authenticity of a given test task can be assessed largely in terms of the framework of test method characteristics and that these characteristics can also provide a basis for designing test tasks that are situationally authentic. Assessing interfactional authenticity and designing tasks that are interactionally authentic, however, is more complex, since this requires us to consider both the characteristics of the test task and the components of the test taker's language ability. For this reason, I will focus the remainder of the paper on a discussion of the relevance of interfactional authenticity to the design of test tasks.

But before I discuss ways of increasing the relative interfactional authenticity of test tasks, I present some examples of test tasks that vary in terms of their relative situational and interfactional authenticity. This will accomplish a number of objectives, I hope. It will help the reader understand how actual test tasks differ in terms of their authenticity. It will also illustrate what Bachman and Palmer (in press) believe are some fundamental facts about authenticity. First, both situational and interfactional authenticity are relative, so that we speak of "low" or "high" authenticity, rather than "authentic" and "inauthentic." Second, we cannot tell how authentic a test task is just by looking at it; we must also consider the characteristics of the test takers and the specific target language use context. Third, certain test tasks may be useful for their intended purposes, even though they are low in either situational or interfactional authenticity. Finally, in either designing new tests or analyzing existing tests, estimates of authenticity are only best guesses. We can do our best to design test tasks that will be authentic for a given group of test takers, but we need to realize that different test takers may process the same test task in different ways, often in ways we may not anticipate.

The first example is from an institution abroad where Adrian Palmer and I once worked, in which some of the typists in our department did not understand English very well. Nevertheless they were excellent typists and produced high quality typescripts, even from handwritten documents. These typists had developed a high level of mechanical control of English, and this was all that was required for their job. A screening test for new typists in this situation might involve simply asking job applicants to type from a handwritten document. If the applicants knew that their on-the-job use of English would be limited to exactly this kind of typing, they would probably perceive the typing test as highly relevant to the job. Clearly, however, the test meets very few of the criteria for interfactional authenticity. This example illustrates a test task which would be evaluated as highly situationally authentic but low in

terms of interfactional authenticity. The number *1* in the upper left corner of the diagram in Figure 4 indicates where this example test falls, in terms of authenticity. How useful is this test likely to be for its intended purpose? Probably quite useful.

FIGURE 4
Aspects of Authenticity



We can use the same testing situation to invent a second example. Suppose for a moment that these applicants were capable of carrying on a conversation in English, and that we tested them by interviewing them in English. If the topics we talked about in the interview were of interest to them, the interview might actually involve the same types of interactions involved in nontest conversation. If we used the scores from this interview to select individuals whose sole use of English was to type from handwritten documents, how would this example rate with respect to authenticity? I would judge this task to be relatively low in situational authenticity and relatively high in interfactional authenticity. The number 2 in the diagram indicates where this example falls. How useful is this test likely to be for predicting applicants' ability to type from handwritten documents? Probably not very useful.

For our third example, suppose we gave international students applying to U.S. universities a test of English vocabulary in which they were required to match words in one column with meanings in a second column. How does this stack up in terms of authenticity?

This task, I would say, is relatively low in both situational and interfactional authenticity. The number 3 in the diagram shows where this example falls. How useful is this test likely to be in predicting readiness for academic study in English? Given the generally low predictive utility of language tests for forecasting academic achievement, it is difficult to say how useful this test might be.

One final example: Suppose we used a role play in which a prospective sales person was required to attempt to sell a product. The role play might involve a face-to-face oral conversation with an interlocutor who plays the role of a potential customer. The test taker might be required to engage the hypothetical customer in a conversation, decide what kind of approach to use in the selling task, and carry out the task. How authentic is this test task? I would rate this task as relatively high in both situational and interfactional authenticity. The number 4 in the diagram indicates where this test task falls. How useful is this test likely to be in selecting successful sales persons? It could be quite useful, particularly if we included some criteria about successful completion of the task in the scoring.

INCREASING THE INTERACTIONAL AUTHENTICITY OF TEST TASKS

The four examples just given illustrate how test tasks can differ in their relative authenticity, and suggest how authenticity may be useful in evaluating existing tests. I now turn to considerations in designing and developing test tasks that are relatively interactionally authentic.

The two main steps in developing interactionally authentic test tasks are as follows:

1. For a given test task, assess the degree to which the strategies of language ability are involved in successfully completing the test task.
2. Explore ways of increasing the interaction of each strategy.

Assessing Levels of Interfactional Authenticity

The involvement of the components of language ability in a given test task can be assessed in a number of ways. We could, for example, imagine ourselves to be typical test takers and then speculate about the levels of involvement of the strategies and areas of language knowledge. What specific areas of language knowledge will be engaged by the test task? Grammatical knowledge?

Sociolinguistic knowledge? What strategies will be evoked—assessment, planning—and to what extent? Another way of assessing interfactional authenticity is through observing test takers and asking them to self-report on the strategies they used in attempting a given test task. This method of introspection is being used increasingly by language testing researchers to investigate the strategies used by test takers and how these strategies are related to test performance (e.g., Anderson et al., 1991; Cohen, 1984; Nevo, 1989).

Increasing Interfactional Authenticity

Interfactional authenticity can be increased, I believe, by increasing the level of involvement of the test taker's language ability. Bachman and Palmer (in press) suggest that this can be done by designing the test method characteristics of the test tasks to provide for the following: (a) requirement, (b) opportunity, (c) feasibility, and (d) interest. In order to assure that the strategies or given areas of language knowledge must be involved for successful completion of the test, one might set up tasks that cannot be done without their involvement. Opportunity for involvement may be provided by allocating adequate time or providing necessary information, tools, references, and so forth. Feasibility of involvement might be increased by adjusting the difficulty of the test task so that the involvement of the strategies and areas of language knowledge will be within the test takers' ability range. Finally, involvement of language ability may be increased by making test tasks interesting to test takers. In this regard, interest may be increased through increased situational authenticity.

Let me offer an example of how one might evaluate interfactional authenticity and try to increase it for a specific writing test task. Suppose we want to develop a writing test as a final examination for a composition course. And suppose one procedure that comes to mind involves simply giving the candidates one hour to write a 250-word composition on a prescribed topic. Our initial task specification, or prompt, might simply be "Write a 250 word composition describing your most frightening experience. You will not be graded on what you say. You will be graded on how well you express your ideas."

Assessment strategies. In this task candidates might assess their experiences with frightening situations and their ability to write about them. This could affect their choice of topics (goal-setting). As they write their essays, they might take some time to assess the

correctness of their grammar and spelling. After finishing a draft of their essays, they might spend some time in revision and editing, which involve reassessment of the language produced. Since the instructions indicate that the essay will not be graded on content, however, the candidates might not reassess the information in the essay.

In this task the involvement of assessment is likely to be rather limited. Since the main focus of the task seems to be on how well the candidates express ideas, assessment is likely to be limited primarily to this domain. Moreover, since no detailed grading criteria are provided, the candidates are not prompted to assess what they have written in any particular areas. They might construe "how well you express your ideas" in the prompt to mean correct grammar and spelling. Of course, they may have a much richer set of internal criteria for assessment, both of content and of language, but nothing in the prompt specifically evokes this.

We could increase the involvement of assessment strategies in this task by providing more specific information in the prompt about the criteria for grading. We could also encourage test takers to go back and revise their writing, and provide adequate time for this.

Goal setting. In this example, test takers are assigned a topic, and the involvement of goal setting might be limited to picking a particular frightening experience to describe. They are given no purpose for writing the essay other than to produce something to be graded. And they are given no indication that what they say is of any importance. Their thinking is likely to be limited to "What can I write the best essay about?" Their emotions might be involved in this decision since the topic of the essay is about an emotional experience. On the other hand, the instructions state that the content of what is written will not influence the grade, and this might tend to cause test takers to minimize the importance of what they write, and hence their emotional involvement with it. Also, since the task does not require them to come up with anything new, it might not evoke feelings of interest, concern or excitement.

To increase the involvement of goal setting we might consider allowing the test taker a choice of topics to talk about, the choice to be made on the basis of what the test taker finds interesting. So we might provide the test taker with a list of topics including political, social, educational, and religious issues. We might also give the test taker a choice of purpose or intended audience.

Planning. In our example writing task, no instructions are given as to how to organize the essay or the role that organization plays in the

grade the essay will receive. In addition, the task is such that a simple chronological organization would suffice, so the test taker might feel little need to do much planning. Because the task is one that requires little planning, the test taker needs to do very little formal organization of ideas. Thus, the way the task is specified (particularly the lack of indication that the essay needs to be well planned) is likely to result in little involvement of planning.

One way to increase the involvement of planning would be to explicitly state in the prompt that planning is part of the task, and then allowing test takers some time to organize and prepare their thoughts on the topic selected. We might even go so far as to require an outline as part of the test taker's response.

These examples suggest the following principles for increasing interfactional authenticity of test tasks:

1. Tasks should consist of multiple, sequential subtasks.
2. Goal setting and planning should be required at the beginning of the task.
3. Request for and provision of feedback should be given at the end of each subtask.
4. Feedback should be used in subsequent subtasks.
5. There should be provision for changing goals and for additional planning.

Clearly, these principles are related to some of the features that have been included in the design of communicative language tests, as discussed in Part 1 above. An information gap can be created by including complementary types of information in different tasks. Information presented aurally via a videotape recording, for example, could be complemented by different information on the same topic presented visually in a reading passage. Another feature of communicative tests is that of task dependency. This can be built into language tests by sequencing tasks so that information generated in completing one task is used to complete subsequent tasks. Answers to listening and reading comprehension questions, for example, could provide the content for a writing task. These principles are consistent with much practice in the development of communicative language tests, and thus cannot be claimed to be original. Nevertheless, it is comforting to find that principles for the design and development of language tests that are derived from theoretical models of language ability and test method characteristics are in keeping with practice that is derived from language teaching.

CONCLUSION

I believe the framework for language testing presented here makes two contributions to the field. First, it grounds practical considerations in test design, development, and use firmly on a theoretical framework of the nature of language ability and test tasks, and thus provides a principled basis for making practical test development decisions. Second, and equally important, this framework enables us to specify and assess the relationship between language test performance and nontest language use, and thus provides a principled basis for addressing issues of validity and authenticity.

■

ACKNOWLEDGMENTS

I would like to thank Adrian Palmer for his useful comments and suggestions in preparing this paper. In particular, the material in Part 2 is largely the result of our collaborative effort to apply a theory of language test performance to the practical design and development of language tests. I also thank Charles Alderson and Peter Skehan for making available their recent reviews of language testing. I am grateful to Kari Sajavaara and Eduardo Cascallar for their valuable comments and suggestions with regard to metacognitive strategies. An earlier version of the second part of this paper was presented at the PennTESOL-East 1991 Spring Conference in Philadelphia, March 1991.

THE AUTHOR

Lyle F. Bachman is Professor of Applied Linguistics at the University of California, Los Angeles. His publications include journal articles in the areas of language testing and program evaluation. His book *Fundamental Considerations in Language Testing* (Oxford University Press) was awarded the 1991 Mildenberger Prize for outstanding research publication in foreign language teaching. His research interests include the development and validation of tests of communicative language ability, the analysis of test content, and the effects of test-taker characteristics on test performance.

REFERENCES

Alderman, D. L., & Holland, P. W. (1981). *Item performance across native language groups on the Test of English as a Foreign Language* (TOEFL Research Report No. 9). Princeton, NJ: Educational Testing Service.

Alderson, J. C. (1981a). Reaction to Morrow paper (3). In J. C. Alderson & A. Hughes (Eds.), *Issues in language testing* (ELT Documents No. 111, pp. 45-54). London: The British Council.

Alderson, J. C. (1981 b). Report of the discussion on communicative language testing. In J. C. Alderson & A. Hughes (Eds.), *Issues in language testing* (ELT Documents No. 111, pp. 55-65). London: The British Council.

Alderson, J. C. (1986, January). Levels of reading comprehension: Do they exist? Paper delivered at Salford University, England.

Alders on, J. C. (1988a). *Innovation in language testing: Can the microcomputer help?* (Language Testing Update Special Report No. 1). Lancaster, England: Lancaster University, Institute for English Language Education.

Alderson, J. C. (1988b). New procedures for validating proficiency tests of ESP? Theory and practice. *Language Testing, 5* (2), 220-232.

Alderson, J. C. (1988c, March). Testing reading comprehension skills. In P. L. Carrell, J. Devine, & W. Grabe (Organizers), *Research in reading in a second language.* Colloquium conducted at the 22nd Annual TESOL Convention, Chicago, IL.

Alderson, J. C. (1990, March). *Judgments in language testing.* Paper presented at the 12th Annual Language Testing Research Colloquium, San Francisco, CA.

Alderson, J. C. (1991): Language testing in the 1990s: How far have we come? How much further have we to go? In S. Anivan (Ed.), *Current developments in language testing* (pp. 1-26). Singapore: SEAMEO Regional Language Centre.

Alderson, J. C., Foulkes, J., Clapham, C., & lngram, D. (1990, March). *International English language testing service (lELTS).* Panel presented at the 12th Annual Language Testing Research Colloquium, San Francisco, CA.

Alderson, J. C., Henning, G., & Lukmani, Y. (1987, April). *Levels of understanding* in *reading comprehension tests.* Paper presented at the 9th Annual Language Testing Research Colloquium, Miami, FL.

Alderson, J. C., & Lukmani, Y. (1986, March). Reading in a second language. In D. E. Eskey, J. Devine, & P. L. Carrell (Organizers), *Research on reading in a second language.* Colloquium conducted at the 20th Annual TESOL Convention, Anaheim, CA.

Alderson, J. C., & Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing, 2* (2), 192-204.

Anderson, N., Cohen, A., Perkins, K., & Bachman, L. (1991). An investigation of the relationships among test-taking strategies, item content, and item difficulty in an EFL reading test. *Language Testing, 8* (1), 41-66.

The Associated Examining Board. (1987). *Test in English for Educational Purposes (TEEP).* Aldershot, England: Author.

Bachman, L. F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition, 10* (2), 149-164.

Bachman, L. F. (1989). Language testing-SLA interfaces. In R. B. Kaplan (Ed.), *Annual Review of Applied Linguistics 1988* (pp. 193-209). New York: Cambridge University Press.

Bachman, L. F. (1990a). Assessment and evaluation. In R. B. Kaplan (Ed.), *Annual Review of Applied Linguistics 1989* (pp. 210-226). New York: Cambridge University Press.

Bachman, L. F. (1990b). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F., Davidson, F., Lynch, B., & Ryan, K. (1989, March). *Content analysis and statistical modeling of EFL proficiency tests.* Paper presented at the llth Annual Language Testing Research Colloquium, San Antonio, TX.

Bachman, L. F., Davidson, F., & Milanovic, M. (1991, March). *The use of test method characteristics in the content analysis and design of EFL proficiency tests.* Paper presented at the 13th Annual Language Testing Research Colloquium, Princeton, NJ.

Bachman, L. F., Davidson, F., Ryan, K., & Choi, I-C. (in press). An *investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL comparability study*. Cambridge: University of Cambridge Local Examinations Syndicate.

Bachman, L. F., Kunnan, A., Vanniarajan, S., & Lynch, B. (1988). Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries. *Language Testing, 5* (2), 128-159.

Bachman, L. F., & Palmer, A. S. (1981). The construct validation of the FSI oral interview. *Language Learning, 31* (1), 67-86.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly, 16* (4), 449-465.

Bachman, L. F., & Palmer, A. S. (1988). The construct validation of self-ratings of communicative language ability. *Language Testing, 5* (2), 260-275.

Bachman, L. F., & Palmer, A. S. (in press). *Language testing in practice.* Oxford: Oxford University Press.

Bachman, L. F., & Savignon, S. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal, 70* (4), 380-390.

Bialystok, E. (1990). *Communication strategies: A psychological analysis of second-language use.* Oxforal: Basil Blackwell.

Bolus, R. E., Hinofotis, F. B., & Bailey, K. M. (1982). An introduction to generalizability theory in second language research. *Language Learning, 32* (2), 245-258.

Brindley, G. (1989). *Assessing achievement in the learner-centred curriculum.* Sydney, Australia: National Centre for English Language Teaching and Research.

Canale, M. (1983). On some dimensions of language proficiency. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 333-342). Rowley, MA: Newbury House.

Canale, M. (1984). Testing in a communicative approach. In G. A. Jarvis (Ed.), *The challenge for excellence in foreign language education* (pp. 79-92). Middlebury, VT: The Northeast Conference Organization.

Canale, M. (1986). The promise and threat of computerized adaptive assessment of reading comprehension. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 29-45). Washington, DC: TESOL.

Canale, M., & Swain, M. (1980a). A domain description for core FSL: Communication skills. In *The Ontario assessment instrument pool: French as a second language, junior and intermediate divisions* (pp. 27-39). Toronto, Canada: Ontario Ministry of Education.

Canale, M., & Swain, M. (1980b). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1* (l), 1-47.

Candlin, C. (1986). Explaining communicative competence limits of testability? In C. W. Stansfield (Ed.), *Toward communicative competence testing: Proceedings of the second TOEFL invitational conference* (pp. 38-57). Princeton, NJ: Educational Testing Service.

Carroll, B. J. (1980). *Testing communicative performance.* London: Pergamon Institute of English.

Chapelle, C. (1988). Field independence: A source of language test variance? *Language Testing, 5* (1), 62-82.

Chapelle, C., & Roberts, C. (1986). Ambiguity tolerance and field dependence as predictors of proficiency in English as a second language. *Language Learning, 36,* 27-45.

Chen, Z., & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. *Language Testing, 2* (2), 155-163.

Clifford, R. T. (1981). Convergent and discriminant validation of integrated and unitary language skills: The need for a research model. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validation of tests* of *communicative competence* (pp. 62-70). Washington, DC: TESOL.

Cohen, A. D. (1984). On taking tests: What the students report. *Language Testing, 1* (1), 70-81.

Criper, C., & Davies, A. (1988). *ELTS validation project report.* London: The British Council and the University of Cambridge Local Examinations Syndicate.

Davies, A. (1977). The construction of language tests. In J. P. B. Allen & A. Davies (Eds.), *The Edinburgh course in applied linguistics: Vol. 4. Testing and experimental methods* (pp. 38-194). London: Oxford University Press.

de Jong, J. H. A. L., & Stevenson, D. K. (1990). *Individualizing the assessment of language abilities.* Clevedon, England Multilingual Matters.

Erickson, M., & Molloy, J. (1983). ESP test development for engineering students. In J. W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 280-288). Rowley, MA: Newbury House.

Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL Quarterly, 16* (1), 43-59.

Fouly, K. A. (1985). A confirmatory multivariate study of the nature of second language proficiency and its relationship to learner variables. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign.

Hansen, L. (1984). Field dependence-independence and language testing: Evidence from six Pacific Island cultures. *TESOL Quarterly, 18* (2), 311-324.

Hansen, J., & Stansfield, C. (1981). The relationship between field dependent-independent cognitive styles and foreign language achievement. *Language Learning, 31,* 349-367.

Harrison, A. (1983). Communicative testing: Jam tomorrow? In A. Hughes & D. Porter (Eds.), *Current developments in language testing* (pp. 77-85). London: Academic Press.

Henning, G. (1987). *A guide to language testing.* Cambridge, MA: Newbury House.

Hudson, T., & Lynch, B. (1984). A criterion-referenced measurement approach to ESL. *Language Testing, 1* (2), 171-201.

Ingram, D. E. (1984). *Australian second language proficiency ratings.* Canberra: Department of Immigration and Ethnic Affairs.

Kramsch, C. (1986). From language proficiency to interfactional competence. *The Modern Language Journal, 70* (4), 366-372.

Lantolf, J. P., & Frawley, W. (1985). Oral proficiency testing: A critical analysis. *The Modern Language Journal, 69,* 337-345.

Lantolf, J. P., & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition, 10* (2), 181-195.

Long, J. S. (1983a). *Confirmatory factor analysis.* Beverly Hills, CA: Sage.

Long, J. S. (1983b). *Covariance structure models: An introduction to LISREL.* Beverly Hills, CA: Sage.

Lowe, P., Jr. (1988). The unassimilated history. In P. Lowe, Jr. & C. W. Stansfield (Eds.), *Second language proficiency assessment: Current issues.* Englewood Cliffs, NJ: Prentice Hall.

Milanovic, M. (1988). *The construction and validation* of *a performance-based battery of English language progress tests.* Unpublished doctoral dissertation, London University, Institute of Education.

Morrow, K. (1977). *Techniques of evaluation for a notional syllabus.* London: Royal Society of Arts.

Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C. J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 143-157). Oxford: Oxford University Press.

Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing, 6* (2), 199-215.

Oller, J. W., Jr. (1979). *Language tests at school: A pragmatic approach.* London: Longman.

Oller, J. W., Jr. (1983). A consensus for the eighties? In J. W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 351-356). Rowley, MA: Newbury House.

Perry, T., & Stansfield, C. W. (Eds.). (1990). *Language aptitude revisited.* Englewood Cliffs, NJ: Prentice Hall.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. In A. Valdman (Ed.), *The assessment of foreign language proficiency* [Theme issue]. *Studies in Second Language Acquisition, 10,* 121-243.

Politzer, R. L., & McGroarty, M. (1985). An exploratory study of learning behaviors and their relationship to gains in linguistic and communicative competence. *TESOL Quarterly, 19* (1), 103-123.

Purcell, E. T. (1983). Models of pronunciation accuracy. In J. W. Oller, Jr. (Ed.), *Issues* in *language testing research* (pp. 133-151). Rowley, MA: Newbury House.

Ryan, K., & Bachman, L. F. (in press). Differential item functioning on two tests of EFL proficiency. *Language Testing.*

Sang, F., Schmitz, B., Vollmer, H. J., Baumert, J., & Roeder, P. M. (1986). Models of second language competence: A structural equation approach. *Language Testing, 3* (l), 54-79.

Savignon, S. J. (1985). Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *The Modern Language Journal, 69* (2), 129-134.

Seaton, I. (1983). The English Language Testing Service (ELTS): Two issues in the design of the new "non-academic module." In A. Hughes & D. Porter (Eds.), *Current developments in language testing* (pp. 129-139). London: Academic Press.

Shohamy, E. (1983). The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning, 33,* 527-540.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing, 1* (2), 147-170.

Skehan, P. (1988). State of the art article: Language testing I. *Language Teaching, 21* (4), 211-221.

Skehan, P. (1989). State of the art article: Language Testing II. *Language Teaching, 22* (l), 1-13.

Skehan, P. (1991). Progress in language testing: The 1990s. In J. C. Alderson & B. North (Eds.), *Language testing in the 1990s: The communicative legacy* (pp. 3-21). London: Modern English Publications and The British Council.

Spolsky, B. (1989). *Conditions for second language learning.* Oxford: Oxford University Press.

Stansfield, C. W. (Ed.). (1986). *Technology and language testing.* Washington, DC: TESOL.

Stansfield, C. W., & Hansen, J. (1983). Field dependence-independence as a variable in second language cloze test performance. *TESOL Quarterly, 17* (1), 29-38.

Stansfield, C. W., & Kenyon, D. M. (1988). *Development of the Portuguese speaking test.* Washington, DC: Center for Applied Linguistics.

Stansfield, C. W., & Kenyon, D. M. (1989). *Development of the Hausa, Hebrew, and Indonesian speaking tests.* Washington, DC: Center for Applied Linguistics.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.
Sternberg, R. J. (1988). *The triarchic mind.* New York: Viking.
Swain, M. (1985). Large-scale communicative language testing: A case study. In Y. P. Lee, A. C. Y. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 35-46). Oxford: Pergamon Press.
Swinton, S., & Powers, D. E. (1980). *Factor analysis of the Test of English as a Foreign Language for several language groups* (TOEFL Research Rep. No. 6). Princeton, NJ: Educational Testing Service.
Tung, P. (1986). Computerized adaptive testing: Implications for language test developers. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 13-28). Washington, DC: TESOL.
Weir, C. J. (1983). The Associated Examining Boards Test of English for academic purposes: An exercise in content validation events. In A. Hughes & D. Porter (Eds.), *Current developments in language testing* (pp. 147-153). London: Academic Press.
Wesche, M., Canale, M., Cray, E., Jones, S., Mendelsohn, D., Tumpane, M., & Tyacke, M. (1987). *The Ontario Test* of *English as a Second Language (OESL): A report on the research.* Ottawa, Canada: Ontario Ministry of Colleges and Universities.
Widdowson, H. G. (1978). *Teaching language as communication.* Oxford: Oxford University Press.
Widdowson, H. G. (1983). *Learning purpose and language use.* London: Oxford University Press.
Wilds, C. P. (1975). The oral interview test. In R. L. Jones & B. Spolsky (Eds.), *Testing language proficiency* (pp. 29-38). Washington, DC: Center for Applied Linguistics.
Zeidner, M. (1986). Are English language aptitude tests biased towards culturally different minority groups? Some Israeli findings. *Language Testing, 3* (l), 80-95.
Zeidner, M. (1987). A comparison of ethnic, sex and age biases in the predictive validity of English language aptitude tests: Some Israeli data. *Language Testing, 4* (l), 55-71.

REVIEWS

The *TESOL Quarterly* welcomes evaluative reviews of publications relevant to TESOL professionals. In addition to textbooks and reference materials, these include computer and video software, testing instruments, and other forms of nonprint materials.

Edited by HEIDI RIGGENBACH
University of Washington

Two Reviews of Teaching and Learning Vocabulary

Following are two reviews of I. S. P. Nation's book from somewhat different perspectives.

Teaching and Learning Vocabulary.
I. S. P. Nation. New York: Newbury House, 1990. Pp. xi + 275.

■ Based on Nation's belief that a systematic, principled approach to vocabulary instruction results in better learning, this 12-chapter text introduces the ESL/EFL teacher to research findings and to pedagogical concerns in the area of vocabulary. This is important since vocabulary instruction is an area that has been neglected to varying degrees in both audiolingual and communicative language teaching.

The first five chapters serve as background: "Introduction," "The Goals of Vocabulary Learning and Vocabulary Size," "What is Involved in Learning a Word?" "Communicating Meaning," and "Assessing Vocabulary Size." Chapters 6 through 9 discuss vocabulary in relation to the four language skills (listening, speaking, reading, and writing), and the final three chapters treat related issues: learner strategies, simplification of reading materials and current directions in research on vocabulary.

The book contains an impressive bibliography, an index, and eight appendices. The appendices consist of two useful word lists: words from the *General Service List* (West, 1953) which are not likely to be well known and a university word list. There are also two passages for testing vocabulary-one with words of various frequency levels omitted and one with nonsense words for the reader to define in context. These are followed by formulas for examining the vocabulary of any textbook in terms of word frequency, lexical density, and the proportion of new to known

words. Appendix 6 is a useful list of 10 conjunction relationships (e.g., inclusion, cause-effect), and Appendix 7 illustrates vocabulary puzzles. Appendix 8 discusses the notion of a vocabulary levels test and demonstrates how a teacher might develop and score such a test. It also presents a sample levels test with words that range from the 2,000-word level to the 10,000-word level.

I have used this volume as the core text for a seminar on the role of vocabulary in language teaching. The evaluations of my nine graduate students and myself have been somewhat mixed. We agree that good points of the text are that it provides a summary of a great deal of literature; that it is a good, practical resource; and that it is clearly organized. Nation wrote the text to be easily understood by nonnative ESL/EFL teachers and, for the most part, it is very easy to read. Ironically, however, some vocabulary usages (e.g., "a ladder" [p. 104] in a stocking, for a run; I "clean my teeth" [p. 96] for I brush my teeth; "word square" [p. 254] for crossword puzzle) will strike North American readers as odd if they are not aware that the author's native dialect is New Zealand English, a variety closely related to British English.

On the negative side, some chapters lack substance. For example, chapter 9, on vocabulary and writing, emphasizes low-level concerns like spelling instead of dealing with more interesting problems like word-choice in nonnative writing; chapter 12, "Directions in Vocabulary Studies" is a scant four pages that neglects to mention, among other things, the importance of corpus-based lexical research. Also, several of the pedagogical examples (e.g., "We can find a stirrup on a horse" [p. 67] instead of on a saddle) and suggested techniques (e. g., using a nonalphabetic code to represent words) are problematic.

Yet even the students who did not particularly like the text agreed that it had generated nonstop discussion and that its shortcomings forced us to improve on the pedagogical exercises and to ask interesting research questions. Undoubtedly, the sophistication of the book notices in this issue of the *TESOL Quarterly* written by my students results in part from having read Nation's book.

Pedagogically, the text has less to offer experienced teachers than it does to novices; in addition, most of the teaching suggestions are geared to teachers of beginning and intermediate learners rather than to those of advanced learners. Word lists and the notion of frequency are promoted at the expense of communicative function. Theoretically, the text fails to fully incorporate current findings from communicative language teaching and testing as well as data-based discourse analysis. Sometimes one theory is cited where another would be more felicitous. For example, Nation's discussion

of increasing vocabulary through listening assumes simplified texts and incorporates Krashen's (1981) model of comprehensible input to account for such vocabulary acquisition; I find Bruner's (1986) notion of scaffolding (i.e., making it possible for learners to comprehend/perform beyond their level of competence because of appropriate assistance from a more advanced/expert source) a more satisfying explanation.

Despite the reservations expressed, I am positive about the book. It is an excellent resource, and when used in combination with other texts—both theoretical (e.g., Carter & McCarthy, 1988) and pedagogical (e.g., Gairns & Redman, 1986)—Nation's book is important for all language teaching professionals who are concerned with vocabulary teaching and learning.

REFERENCES

Bruner, J. (1986). *Actual minds, possible worlds.* Cambridge, MA: Harvard University Press.

Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching.* London: Longman.

Gairns, R., & Redman, S. (1986). Working *with words: A guide to teaching and learning vocabulary.* Cambridge: Cambridge University Press.

Krashen, S. (1981). The "fundamental pedagogical principle" in second language teaching. *Studia Linguistica, 35* (1 & 2), 50-70.

West, M. (1953). A *general service list of English words.* London: Longman, Green.

MARIANNE CELCE-MURCIA
University of California, Los Angeles

Although Nation teaches at Victoria University of Wellington in New Zealand, this book can be best understood if it is seen as an example of a long established tradition of British lexicography which has focused on the role of the word in language acquisition. To mention only a few of the highlights of this approach: Barnard's controlled vocabulary textbooks (e.g., Barnard, 1980), West's (1953) word count (the only one specifically focused on the needs of ESL/EFL), the simplified Longman readers, and the work by Carter (1987), Meara (1980), and McCarthy (1990), All of these approaches have assumed that vocabulary is fundamental to language learning.

In contrast, ESL in the U.S. has had minimal interest in vocabulary since the days of grammar translation. The various

philosophies of ESL pedagogy from audiolingual to communicative have consistently undervalued the importance of vocabulary in language acquisition. Pedagogical practice largely ignores vocabulary or relegates it to a position of minor emphasis.

It is obvious that Nation has the classroom very much in mind when he gives many tips, charts, and techniques for the classroom teacher. But his approach is also based on extensive scholarship, as exemplified in the 900-item bibliography. Thus it seems that this book is an excellent example of the often proposed but seldom realized synergistic relationship between research and pedagogy.

Nation divides vocabulary into low and high frequency, and specialized. It is claimed that most vocabulary learning will happen "when the learners use the language for other purposes" (p. 6) (i.e., acquisition). However, "the research on frequency counts gives clear guidelines" (p. 24) for certain goals of vocabulary learning. The main implication is that very high frequency words in English occur so often that they should be learned/taught. On the other hand, low frequency words do not deserve teaching time and instead strategies should be taught to the students for dealing with these words. On average each word is repeated between two and three times in a given text. But since approximately 40% of the words will occur only once, such material is not suited to explicit vocabulary learning.

Nation offers extensive discussion of many of the facets of vocabulary acquisition such as the difficulty of form for the learner, the amount of repetition needed for learning, the associative properties of words and semantic fields, etc. He gives pedagogical suggestions supported by the research which may be surprising to some, such as the avoidance of teaching pairs of opposites at the same time. Moreover, words which are very similar in sound or spelling should not be introduced early. This follows from the fact that vocabulary initially tends to be stored in the mind according to form, and is subsequently stored according to meaning. Similarly, Nation advocates the use of L1 in the monolingual classroom because of the speed and accuracy with which vocabulary knowledge can be taught. At the same time he urges caution in its use and, in particular, urges multiple methods of presentation, i.e., definition by abstraction, by context, and by demonstration as well as by L1.

In the area of listening there are some very intriguing techniques presented, such as a dictation test for assessing general vocabulary knowledge and activities for turning reading vocabulary into listening vocabulary.

With respect to speaking, Nation suggests that the 2,000 most frequent words of the *General Service List* (GSL) (West, 1953) are

a reasonable goal. It is best to give learners practice in saying a lot using a small number of words. He emphasizes teaching the ability to paraphrase, thereby using this small set of words to achieve fluency and independence from translation.

For reading, Nation claims that a knowledge of 3,000 headwords is needed in order to read unsimplified texts. His proposal is to learn vocabulary both by direct study and by reading. Words should be learned, he believes, in lists as well as in context. He recommends use of mnemonic techniques for remembering words and especially the reading of large amounts of simplified text. The basis of Nation's approach to vocabulary acquisition is compatible with the recommendations of Krashen and others: Although Krashen (1989) would presumably not agree with the explicit learning techniques which Nation proposes, Nation recommends extensive reading of comprehensible and enjoyable texts.

The university word list in Appendix 2 is a good example of how Nation identifies and treats different types of vocabulary. As already noted, he believes that the 2,000 words of the GSL must be learned very, very well. The middle range of vocabulary by frequency should then be acquired to some degree of (probably receptive) familiarity through extensive reading. Finally, the 40% that are highly infrequent should be acquired through contextual guessing (to be discussed later) and personal learning. But he also recommends learning some moderately frequent words generally found in academic texts which are used to signal important attributes and relationships of more technical vocabulary. This so-called academic vocabulary is represented in the 740-item university word list.

The chapter on learner strategies describes three main approaches: guessing from context, use of certain mnemonic techniques, and the use of word parts. It is claimed that guessing from context is the most important strategy for dealing with new vocabulary. In this same chapter a list of useful Latin prefixes and a list of 14 master words (whose prefixes and roots are keys to the meanings of over 14,000 words) are presented.

The chapter on simplification of reading material cites the traditional rationale for simplifying the vocabulary in texts. Unfortunately, Nation does not respond to the current criticisms of such texts as not being authentic and displaying atypical syntax, collocations, etc. In response to these criticisms there has developed a strong preference for the use of authentic texts. This remains very much an open issue.

One can criticize some of the emphases in this book. It fails to take into account a good deal of the more recent thinking about reading in a foreign language, that is, the interactive approach (Barnett, 1989). This philosophy about reading stresses that a

student must be able to read different texts with great flexibility of purpose and strategy, which certainly implies that a number of different strategies should be used in vocabulary acquisition. In this view, the question of when a learner should skip, guess, look up, etc., is largely determined by the purpose of reading. Another aspect which needs to be stressed is not only the need to read extensively but how this activity leads to acquisition, that is, the question raised by Krashen (1989). In short, from this viewpoint the reader/learner needs to be taught various metacognitive strategies which will guide the more specific strategies for vocabulary acquisition (Oxford, 1990).

However, in the final analysis this is a book which should be on the reference shelf of every ESL/EFL teacher who believes that attention must be paid to vocabulary as a significant variable in language learning.

REFERENCES

Barnard, H. (1980). *Advanced English vocabulary workbook I.* New York: Newbury House.

Barnett, M. (1989). *More than meets the eye: Foreign language reading, theory and practice.* Englewood Cliffs, NJ: Prentice Hall.

Carter, R. (1987). *Vocabulary.* London: Allen & Unwin.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73* (4), 440-464.

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning, *Language Teaching and Linguistics Abstracts,* 221-246,

McCarthy, M. (1990). *Vocabulary.* Oxford: Oxford University Press.

Oxford, R. (1990). *Language learning strategies: What every teacher should know.* New York: Newbury House.

West, M. (1953). A *general service list of English words.* London: Longman, Green.

JAMES COADY
Ohio University

***Second Language Writing: Research Insights for the Classroom* (Cambridge Applied Linguistics Series).**
Barbara Kroll (Ed.). Cambridge: Cambridge University Press, 1990, Pp. ix + 246.

■ In her preface to this book, editor Barbara Kroll tells us that the idea for it began in 1979 when she taught her first graduate course

in second language writing at the University of Hawaii and realized the need for—and lack of—an appropriate text. More than 11 years later, I taught the same graduate course at Hawaii that Kroll did, but with the advantage of the book which she and the contributors to this excellent volume have recently made available. There is no doubt in my mind that it will become the standard for texts in L2 writing instruction for some time to come.

The book is designed for teachers, teacher trainers, and researchers in the field of second language writing. It consists of 13 chapters divided into two sections: "Philosophical Underpinnings of Second Language Writing Instruction" and "Considerations for Writing Instruction." The aim of each chapter in the theoretical section is to provide "an overview perspective on one essential element in the total picture of L2 writing instruction," while that of the second section is to offer "a variety of specific studies, each focused on a different aspect of writing and/or the writing classroom" (pp. 3-4). To put it another way, the book moves from general and theoretical issues to specific and empirically based classroom studies. Kroll pairs each chapter in the theoretical first section with a corresponding empirical study in the second; and if the match is not exact, it is close enough to highlight the links between theory and practice. The intelligent design is one of the many virtues of the book.

Another is its readability. There is, alas, no guarantee that articles on writing will themselves be well written, but this collection of readings is. It is clear and unlabored, with headings and subheadings unobtrusively orienting newcomers to the field while at the same time signaling familiar territory to seasoned practitioners.

Chapters 1 and 2 are, to me, particularly welcome. In chapter 1, Tony Silva presents a brief but sorely-needed historical sketch of the teaching of second language writing. He identifies and describes the four major approaches to ESL writing since 1945 (controlled composition, current-traditional rhetoric, process, and English for academic purposes, with the last two currently competing for dominance in universities). Silva mentions, but does not develop, the relationship between L1 and L2 writing instruction, which paves the way for Ann Johns in chapter 2 to elaborate on the ways she believes L1 composition theory can "provoke thinking about theory development in ESL" (p. 33). Instead of concentrating on instructional approaches per se, L1 composition theory has focused on four components of writing: the writer, the audience, internal versus external reality, and language. Taken together, the picture that emerges from chapters 1 and 2 is that of a primarily pragmatic

view of writing in ESL as opposed to a more theoretical/philosophical (or what Johns is not afraid to call ideological) view in L1 composition.

For readers looking for a review of the research on the writing process in ESL, chapter 3 is the place to start. Not only are the studies themselves surveyed, but the research designs and findings are summarized as well, and suggestions for further research provided. My reading of this chapter tells me that at this point generalizations are premature about what goes on in the writing process of ESL students, for too few studies exist to warrant confident statements. For example, on the basis of just a handful of studies with a small number of subjects, it seems hasty to accept without serious reservation the conclusion that "a lack of competence in writing in English results more from the lack of composing competence than from the lack of linguistic competence" (p. 49).

Chapter 4 deals with ESL teachers' written responses to student writing, an area notably lacking in research. By comparison, studies abound in writing assessment, the subject of chapter 5, but, despite them, the variables involved in assessment are so numerous, shifting, and complex that, as the author Liz Hamp-Lyons notes, they have thus far escaped our control. Amid the many vexing problems, the one that stands out in my mind is that

> we do not share a construct of writing quality. . . . Since we as researchers cannot consistently agree with each other when assessing the same writing samples or even sometimes with our own judgments about the same samples made on different occasions, we cannot be looking at the same thing. (p. 80)

That the largest and best known assessment programs manage to achieve a reader reliability score of between .70 and .80 (which, while statistically significant, still leaves much variance unaccounted for) seems a tribute to the efforts of testing researchers and the programs themselves, given the great gaps in our knowledge. Clearly, however, we are a long way from answers to the myriad questions posed by writing assessment.

The first section of the book ends with a consideration of the ways in which reading affects developing proficiency in writing (chapter 6). The strength of this chapter is its clear description of the various hypotheses that have been advanced to explain the reading-writing-cognitive interaction in both L1 and L2 research. Its weakness is in its facile conclusion

> Writing teachers who are sensitive to the role that first language reading and writing abilities play in developing second language literacy skills, and also to the role that reading ability in the second language plays in the development of second language writing skills, will be better prepared to help L2 learners utilize those relationships to become proficient second language writers. (p. 100)

This is misleading. The fact is that, above all, we come away from the chapter with a respectful awareness of how much is involved in the L2 reading-writing connection (two modalities, two languages, and the cognitive processes that allow for transfer to take place between and among both) and how much is still to be explained. Sensitivity alone will not suffice to enable ESL writing teachers to help their students make use of these connections; knowledge based on research will, and the jury is still out on the crucial questions.

Of the seven empirical studies in Section 2, I would like to mention two as samples. In her nicely titled study "What does time buy? ESL student performance on home versus class compositions," Barbara Kroll reports on a study she carried out in which 100 essays were written by advanced ESL students at the university level; 50 of the essays were written in class under time pressure and 50 at home over a period of 10-14 days. The essays were then scored holistically for overall discourse and rhetorical control, coded for syntactic control, and compared. The results showed that

> while the time allowed for the preparation of an essay can contribute to some improvement for the writer both on the syntactic level, and the rhetorical level, it does not appear that additional time *in and of itself* leads to a sufficiently improved essay such that there is a statistical significance to the differences between class and home performance. (p. 150)

Time may not have proven to be a significant factor, but in Fathman and Whalley's study, teacher response was. They divided 72 college-level intermediate ESL students into four groups and asked them to write an essay based on a sequence of pictures that told a story. Teachers gave four different types of response on the essays: none at all, feedback solely on grammar (underlining errors without correcting them), solely on content, and on both grammar and content. The students were then asked to rewrite their essays. The findings showed that there was improvement on the rewritten essays for all four groups, with the greatest improvement in both grammar and content appearing in the essays of the group that received feedback in both areas. It would thus seem that "focus on grammar does not negatively affect the content of writing. . . [and] that teachers need not necessarily assign multiple drafts that

separate revision and editing stages in order to improve student writing" (pp. 186-187).

One of the risks in putting together a book in which each chapter has been written by a different author is a product that is less than the sum of its parts. It can lack unity and cohesion, and the readings may be noticeably uneven in length and quality. Thanks to Kroll's guiding hand, *Second Language Writing* avoids these pitfalls, and the result is a book that is of value not only to those interested in ESL writing but also to the field of applied linguistics as a whole.

TERRY SANTOS
Humboldt State University

BOOK NOTICES

The *TESOL Quarterly* welcomes short evaluative reviews of print and nonprint publications relevant to TESOL professionals. Book notices may not exceed 500 words and must contain some discussion of the significance of the work in the context of current theory and practice in TESOL.

The Role of Vocabulary in Language Teaching

Special-topic book notices follow on vocabulary teaching. The books discussed are of three types: volumes that present research / theoretical foundations, teacher resource texts, and student textbooks. These notices were written by my students in an Applied Linguistics seminar.

MARIANNE CELCE-MURCIA, Guest Editor
University of California, Los Angeles

***Vocabulary and Language Teaching* (Applied Linguistics and Language Study Series).** Ronald Carter and Michael McCarthy (Eds.). London: Longman, 1988. Pp. xii + 242.

This volume is devoted to exploring key issues in the role of vocabulary in pedagogical and research contexts. The author's position is that "issues in vocabulary learning cannot be divorced from the classroom teaching of vocabulary, any more than they can from theoretical and descriptive accounts of lexical structure and organization" (p. 17). *Vocabulary and Language Teaching* will benefit primarily those who seek theoretical background that can be connected to and applied to practice, such as the methodologist, materials writer, and teacher trainer.

Vocabulary and Language Teaching consists of five chapters; all authored by Carter and McCarthy except chapter 4, which includes six contributions from other researchers and practitioners who are involved in lexicology and lexicography. The volume begins with a historical survey of vocabulary studies; chapter 1 introduces important contributions of the vocabulary control movement of the 1930s. The next chapter gives a descriptive account of current research into the structure of the lexicon, that is, how semanticists have elaborated the paradigmatic aspects of lexis as in field theory and componential analysis, how the Firthian tradition in Britain has explored the syntagmatic aspects of lexis, and especially how Halliday (1966, 1985) has attempted to establish lexis as a complement to grammatical theory. The authors do not fail to point out the complexity

and shortcomings of these theories when they have been put into practice in language teaching. They cautiously imply that vocabulary learning and teaching should be pursued multidimensionally, considering even discourse-genre and sociocultural context.

The two introductory chapters establish the connection between vocabulary and other concerns: language teaching, the structure of language, and the organization of vocabulary in pedagogical contexts; this connection serves as the thread that ties together the whole volume.

In chapter 3, the authors discuss implementation of theory in pedagogical situations since 1945; they trace developments, first, with reference to vocabulary teaching in general and, second, in relation to recent progress in computer-aided lexicography, which contributes to making learners' dictionaries. The developments leading to the mid-1970s are characterized as the communicative approach, which brings the learner into focus and is concerned with learner-centered lexical strategies such as guessing meaning from context. The authors, in the first three chapters, show that they are proponents of the communicative approach to language teaching, pointing out that lexical meanings are basically negotiated in the context of unfolding discourse, which is suggested in their criticisms of field theory and componential analysis.

Chapter 4 presents an overview and specially commissioned contributions by others which address key issues in the field. Topics covered are current trends in vocabulary teaching by Nattinger, the nature of learners' mental lexicons by Channell, the relationship between vocabulary and reading by Nation and Coady, the role of the learners' dictionaries by Summers, stable and creative aspects of language use and their relevance to vocabulary teaching by Cowie, the principles and practice of designing a lexical syllabus by Sinclair and Renouf, discourse-based use of the cloze procedure by Carter, and the use of words in a systematic way for interfactional purposes in spoken discourse by McCarthy. What contributes to integrating these diverse views on vocabulary learning and teaching and putting them into perspective is the careful commentary at the end of each contribution provided by the two editors, which encourages readers to discuss related topics, explore suggestions for further research, and create pedagogical applications.

The selection of papers displays two distinctive perspectives on lexicon: first, whether a word has stable meaning (so that it can be stored in dictionaries), which reflects a distinction between the stable aspect of language on the one hand and the creative and the negotiative aspects on the other. This in turn is related to the second question of whether it is necessary to give decontextualized attention to lexicon in the face of the now popular communicative tradition, where contextualized presentation of language is emphasized. Even though the tension between the two perspectives is presented as desirable in Cowie's paper, the question of creating a methodology for vocabulary teaching which accommodates

both perspectives is unfortunately only suggested as one issue to be addressed in further research.

In the final chapter, the authors emphasize the importance of integrating the insights of discourse analysis and the communicative approach into language teaching. Their argument is that vocabulary pedagogy should concern itself with the role of vocabulary in "real" naturally-occurring contexts and with the ways in which vocabulary is used to negotiate meaning in unfolding talk (i.e., across turns and sentences). Though their attempt to incorporate the best insights of semantics, corpus-based lexicology, and discourse analysis is reasonable and persuasive, there remains the enormous problem of how to apply and adjust these findings and observations to real teaching situations, especially when learners are at the elementary level, where the imposition of lexical strategies to redefine and negotiate the lexical world of any given discourse is too difficult a task.

Since this volume provides a theoretical and conceptual foundation for vocabulary instruction, rather than presenting specific methodologies for teaching vocabulary, it serves as valuable background for those who intend to relate theory and practice.

REFERENCES

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (Eds.). *In memory of J. R. Firth* (pp. 148-162). London: Longman.

Halliday, M. A. K. (1985). An *introduction to functional grammar*. London: Edward Arnold.

KYUNG-HEE SUH
University of California, Los Angeles

The Nature of Vocabulary Acquisition. Margaret G. McKeown and Mary E. Curtis (Eds.). Hillsdale, NJ: Lawrence Erlbaum, 1987. Pp. v + 192.

Although substantial research exists in the field of first language vocabulary acquisition and instruction, which is the major focus of this book, it often fosters contradictory conclusions and pedagogical implications for the classroom. These disagreements arise because individual researchers address different issues and follow diverse approaches resulting in difficulty both in comparing results and in the establishment of a unifying framework.

Rather than providing a unique and conclusive view of vocabulary learning, the compilation of articles which constitute *The Nature of Vocabulary Acquisition* strives to and succeeds in bringing together various—and often contrasting—perspectives of different researchers. These articles address the same questions and report research on similar subjects—primarily children in elementary and secondary school. This

allows detailed examination of implications and helps to reconcile seemingly opposing results.

The editors introduce the book with a preview of the general questions addressed in all of the articles. This introduction is followed by a summary of the different perspectives and concerns that will arise in the book.

The 10 articles contained in the book have been organized into two sections. Both sections present theory based on empirical research, followed by practical implications for vocabulary teaching. The first section focuses on the nature of vocabulary acquisition, that is, on how knowledge of word meanings develops the second section deals with techniques and instructional programs for promoting knowledge of word meanings. The general questions addressed are: What does it mean to know the meaning of a word? How do individuals differ in this knowledge? HOW is vocabulary knowledge acquired and what vocabulary size do individuals have at various ages? Finally, a great deal of consideration is given to the relationship between vocabulary knowledge and reading comprehension, and its implications for vocabulary instruction.

Starting from these general concerns, each article, written by different authors, focuses on particular aspects of vocabulary acquisition. Thus, in chapter 1, Chall deals with the differences in size and instruction between "recognition" and "meaning" vocabularies; Nagy and Herman in chapter 2 also refer to vocabulary size and to the nature of the vocabulary learning task. In chapter 3, Curtis concentrates on vocabulary testing and variations in vocabulary knowledge. In chapter 4, Elshout-Morh and van Daalen-Kapteijns present us with insights into the cognitive processes in learning word meanings and how these are affected by our previous vocabulary knowledge. The following two articles, by Drum and Koropak (chapter 5) and by Stenberg (chapter 6) focus on how most vocabulary learning occurs through reading. In chapter 7, Pressley, Levin, and McDaniel distinguish between two types of learning: inference and remembering new words. Comparisons are made between mnemonic and contextual strategies. The last three chapters (8-10), by Kameenui, Dixon, and Carnine; by Beck, McKeown, and Omanson; and by Graves deal more directly with instructional programs and with the principles that should govern them.

Chapters not only present readers with qualitative and quantitative data but also with some useful guidelines for those who might be interested in replicating or conducting similar experiments on their own. The numerous references cited are a helpful resource.

Because of the variety of studies and views illustrated in all of the articles, this book represents a valuable contribution to the field of vocabulary research. Both researchers and teachers "concernecl with theory and practice of vocabulary learning" (p. 6) will benefit from the insights provided. Although findings are mostly based on L1 studies, implications for ESL/EFL research can also be inferred.

However, the book's predominant emphasis on the nature of vocabulary acquisition as relating to reading comprehension precludes it from addressing other relevant areas, such as vocabulary in speech, listening, and writing. In addition, the book does not intend to provide a range of strategies to be applied in the classroom. Although some teaching guidelines are noted, especially in the last two chapters, the rest of the book basically underlines particular problems and the general principles an instructional program in vocabulary should apply. Because of this focus, *The Nature of Vocabulary Acquisition* will be more beneficial for those who utilize it as a resource rather than as a practical teacher reference.

R. M. VICTORI
University of California, Los Angeles

Working with Words: A Guide to Teaching and Learning Vocabulary (Cambridge Handbooks for Language Teachers). Ruth Gairns and Stuart Redman. Cambridge: Cambridge University Press, 1986. Pp. vi + 200.

Working with Words makes a valuable contribution to the teaching of vocabulary, a subject that "has not received the recognition it deserves in the classroom" (p. 1). This teacher resource book is grounded in current linguistic and psychological theory relevant to vocabulary learning with many ideas for the presentation and practice of vocabulary. The step-by-step teaching activities and worksheets make it very practical as well.

Part A, the first three chapters, deals with "Words: Their Meanings and Forms," including the entire range of meaning from conceptual to affective meaning, how style, register, and dialect influence meaning; and meaning with respect to idioms, collocations, multiword verbs, etc. If we add to this list, the possible emotional or attitudinal factors that might influence the connotation of a word, and the choosing of a word appropriate to its context, then teaching vocabulary might appear to be an overwhelming task. However, in the sections on Teaching Implications following each chapter, the authors help teachers decide how class time might best address these issues.

Part B (chapters 4, 5, 6) deserves special notice for presenting theories that deal with optimum ways of teaching vocabulary with maximum retention. To help L2 learners accelerate the learning process and facilitate storage, the authors endorse "the importance of more thorough processing and systematic organisation as the basis for effective long-term retention" (p. 87). Utilizing the findings of researchers, the authors make a strong case for meaningful tasks, as well as for recycling what is learned to maximize retention.

In Part C, five chapters offer examples of meaningful tasks and practical activities, including the use of imagery, rote learning, and visuals to highlight relationships such as grids and clines. All are complete with

instructions for procedure, presentation, and practice, including several worksheets that may be legally copied for use in the classroom.

In Part D, chapter 12 demonstrates how to evaluate and adapt course books to facilitate vocabulary learning.

Other notable features are the Reader Activity sections found throughout the book, which are designed to involve the reader in a more practical exploration of the information given.

Working with Words is an excellent resource on vocabulary for the ESL/EFL teacher; it is rich in usable activities and ideas, all based on theories relevant to vocabulary learning.

SYLVIE M. CONROW
University of California, Los Angeles

***Vocabulary* (Language Teaching Series).** Michael McCarthy. Oxford University Press, 1990. Pp. x + 173.

In the field of second language teaching today there seems to be a renewed interest in overt vocabulary instruction and a movement away from methodological practices that assume vocabulary is somehow absorbed by the learner even if not explicitly taught. In keeping with this trend, several instructional texts and teacher resource manuals have recently been published. Michael McCarthy's *Vocabulary* is on the whole a welcome addition to the field for it provides instructors with a thorough introduction to the principles and theories of vocabulary learning and insights into how to put theory into practice in the classroom. While some may comment that McCarthy's style is too prescriptive and methodical, I feel his detailed advice is useful for novice teachers. In addition, he succeeds in clearly exploring all relevant issues in a straightforward and interactive manner through a task-based approach.

Vocabulary is divided into three distinct sections. The first five chapters explore current theory and provide insight into the concepts involved in lexical selection. Proceeding from a basic morphological level to a complex discussion of lexical research, this section engages the reader, assisting comprehension of the diverse and at times opaque terminology used in this field. The second section (chapters 6–10) moves from theory to classroom practice. Here instructional theory is explored and applied, ranging from semantic feature analysis to schema building, from input to what McCarthy labels uptake, or committing a word or phrase to memory. While this is the most practical section, it is not one that instructors can pick up at a moment's notice to find a useful task for the day's class, due to the interweaving of theory and practice. To use this text effectively, one is advised to read through the entire second section before going back to select appropriate teaching strategies. The last section addresses the challenges that arise in the real world of vocabulary instruction. McCarthy

provides 15 tasks that can be used to heighten student awareness and involve learners in the solution of these problem areas.

While the straightforward presentation and task-oriented approach assist readers in digesting the wealth of information, the format of the text itself limits the use of this volume as a reference book. Many of the conceptual definitions are brief and the transitions from one segment to the next are not always clearly indicated. McCarthy has carefully selected examples that illustrate the key concepts and he should also be commended on avoiding cultural bias, as the examples appear to be sensitive to the many varieties of English.

Given the structure of this text, its thorough treatment of the topic, and the practical applications found in the second and third sections, it has great value for both native- and nonnative-speaking English teachers and is written in a manner which is accessible to all.

LYN M. REPATH-MARTOS
University of California, Los Angeles

Teaching and Learning Vocabulary. Linda Taylor. New York: Prentice Hall, 1990. Pp. ix + 93.

In *Teaching and Learning Vocabulary* Linda Taylor has produced a handy guide of theoretical and practical information as a reference for instructors teaching vocabulary. Based on the belief that learner methodology must take into account not only the classroom contexts but also those of the world outside the classroom, Taylor has organized the text to reflect tasks for interpolating information, discussion, and analysis of data in order to best serve student needs for gradual lexical acquisition.

Teaching and Learning Vocabulary is divided into five chapters based on the concept posited by Richards (1976), that knowledge of a word exists on various levels. Chapter 1 discusses the importance of teaching vocabulary. Chapter 2, "The Communicative Approach to Teaching Vocabulary: Presenting New Items," is designed to help present new words by discussing them within the context of frequency, register, collocation, morphology, and semantic connotation and denotation, to name a few. Tasks and examples of small lessons are. presented with an explanation of how to set up a task along with a brief description of its theoretical background. Chapter 3, "Repetition and Interaction," deals with real problems that might arise in the classroom and discusses such critical questions as how an instructor can tell if the students have understood the explanations of the vocabulary, how an instructor can ensure that the students will be able to produce the vocabulary taught, and how the effectiveness of the instructor's teaching can be improved. Short answers to these questions, along with examples of pitfalls to avoid, provide insight into the problems posed by each question. Chapter 4, "Exercises for Consolidation," is an arrangement of follow-up exercises

based, again, around frequency, register, collocation, morphology, and semantic connotation and denotation. Finally, chapter 5, "Vocabulary in Discourse," contains a brief section on theory and many examples of vocabulary in spoken discourse. Taylor divides this chapter into rhetorical cues, modification of words or sentences to make them more acceptable to an audience, and adjectives or adverbs used as emphasizers. This chapter also contains a short section on written discourse.

While well designed and organized, direct in its presentation of the information, and supported by concise examples, this task-oriented guide has limited use as an instructor's reference text. Many of the explanations are too brief and there seems to be no overview on how an instructor might incorporate this vast body of information into a well-organized lesson. Although Taylor claims her methodology is context based and communicative in approach, this is not evident in the examples selected. Finally, lacking an introduction for the teacher, this teacher resource book makes no mention of the level for which this book is written, but it seems limited to novice teachers of beginners, and thus much of the information cannot be used at other levels without extensive adaptation.

Despite these shortcomings, I feel that Taylor has made a valuable contribution to vocabulary instruction resource books by combining both theoretical and practical examples in one slim volume. I hope that other texts might reflect Taylor's text design but be more thoroughly executed with examples representative of all levels of language learning.

REFERENCE

Richards, J. C. (1976). Vocabulary teaching and TESOL: A need for re-evaluation of existing assumptions, *TESOL Quarterly*, *10* (1): 77-90.

RACHEL GADER
University of California, Los Angeles

American Vocabulary Builder 1. Bernard Seal. New York: Longman, 1990. pp. 96.

American Vocabulary Builder 2. Bernard Seal. New York: Longman, 1990. pp. 96.

Books 1 and 2 of *American Vocabulary Builder* are U.S. adaptations of the two-volume British publication *Vocabulary Builder* (1987), written for preintermediate to intermediate level EFL/ESL students; they can be used as classroom or self-study material. Both paperback books have a similar format. They consist of three large sections divided into ten units. Each section represents a different subject area: the human body, the inner self, and the world around us in the first book; at home, at play, and at work in the second. Each of the ten units introduces a lexical set, such as parts of the house, ways of moving, noises, etc. As Seal explains in the introduction,

there are two reasons for putting words into lexical sets: It makes the learning of vocabulary more structured, and students can more easily guess the meaning from the context as all the words in a lexical set are related.

Each unit has the same structure. New vocabulary items are introduced in the first part, Words in Context, which is composed of a text or a passage for reading and studying. The target words are printed in the text in bold so that the students can easily refer to the initial text while doing the exercises. While reading, students are to guess the meanings of the new words without looking them up in the dictionary. The passages are written in such a way that, in most cases, new lexical items can be guessed from the context without any difficulty. Even though some of the texts may seem difficult for students at the intermediate level, Seal explains that this has been done intentionally. "The aim is to have students reading at a level that stretches and challenges them, provided that they are able to maintain a general sense of what the text is about" (p. 5).

Each target word from the reading passage appears again in at least one of the exercises that follow. Every unit has a set of five exercises which check students' comprehension of the target words, test their knowledge of lexical items and their usage, give learners an opportunity to expand the list of lexical items related to the topic of the unit (Dictionary Work exercises), make vocabulary learning entertaining (Just for Fun exercises), and get the students talking about the topic using new vocabulary (Think About questions).

Units can be taught in any order within each section, but Seal suggests that it is better to teach (or learn) them in groups of five: After each sequence of five units, he provides a vocabulary review section and a set of Test Yourself exercises. At the end of both books there is an index of key words along with their pronunciation and an answer key. Each unit is presented on two pages, which makes it easy for a teacher to copy; also, the students can see the initial text while doing the activities. Spaces are provided for students' answers. There are many illustrations in the books supporting the text and the exercises.

Though most passages are written in such a way that students can guess the meanings of new lexical items from the context, there are, however, a few units in which the number of new words in the text seems to be overwhelming and the meanings of some words are not clear from the context. This makes these readings difficult and may discourage students from attempting to read the passages without using their dictionaries. The total number of new vocabulary items is about 500 in each book.

In conclusion, both books are well developed and present vocabulary learning as an interesting, challenging, and enjoyable process. They are a good source of vocabulary enrichment for EFL/ESL students and can be used effectively by both native and nonnative teachers.

ELENA FROLOVA
University of California, Los Angeles/
St. Petersburg State University

The Greek and Latin Roots of English. Tamara M. Green. New York: Ardsley House, 1990. Pp. ix + 156.

This book's first chapter "A Polyglot Stew (or 'Food for Thought')" captures the reader's attention by giving a light account of the linguistic origins of many food names in U.S. cuisine. This introduction to word origins is just the appetizer for the other 13 thematic chapters in this book. Each chapter deals with the vocabulary of different disciplines that have received an infusion of Latin and Greek words. These chapters include "Latin and Greek Numbers," "Government and Politics," "Psychology," "Human Biology and Medicine," "Science and Mathematics," and "Myth, Religion and Philosophy." By studying the Latin and Greek vocabulary presented in these chapters, students can "expand their knowledge of English vocabulary" (p. vii). In addition to these field-specific chapters, others deal with the linguistic structures of Latin and Greek and with the history of language families, particularly the history of English. These chapters are "Language Families," "How Latin Works," "Latin into English," "How Greek Works," "Greek into English," and "Latin and Greek Prefixes." By studying these chapters, students can "be made aware of the functions and modes of expression in a variety of languages" (p. vii).

The material in each chapter is clear and accessible. Students should not have any difficulty in understanding the explanations of roots and affixes, or the historical context of the etymology of some of the words. Following the presentation and explanation of the vocabulary items, there are cumulative exercises that reinforce "both vocabulary already learned and analytical skills developed in previous lessons" (p. vii). Although most of the exercises do accomplish these goals, there are some whose purpose seems obscure. For instance, there are exercises that ask students to transliterate Greek words (Greek alphabet) into English (Roman alphabet).

In the Preface, the author has a section on how to use this text. This is a deceptive heading, however. Actually, this section doesn't tell us how to use the text in terms of the progression of lessons, the presentation or explanation of the material, or even the type of students (age, academic level, etc.) who could benefit from this book. Rather, this section outlines the typographical and linguistic conventions used in the book.

It would be useful to have a teacher's manual and/or an answer guide. Also one wonders how much Latin and Greek the teacher needs to know in order to use this textbook, who the appropriate audience is, and whether this book is appropriate as a course textbook or as a reference book that students might consider for self-study. These are issues that are relevant when considering any textbook for classroom adoption. With respect to the audience and use of the book, I would think that advanced ESL/EFL students, particularly in the sciences, could increase their vocabulary size and improve their recognition and understanding of Latin and Greek roots and affixes; however, I would not recommend this book for self-study since there are some exercises that require background knowledge of the

history of the Roman and Greek civilizations and of other Romance languages.

In. summary, this book presents the Latin and Greek roots of English in a way that should appeal to students who are tired of memorizing the usual lists of roots and affixes. Most of the exercises seem to be well developed for the study of roots and affixes. Nonetheless, I would be cautious about suggesting its adoption as a required textbook because almost one third of the book deals with reformation that is not essential in the learning of Latin and Greek roots. Instead, this book could well serve as a source of materials to be used with advanced students (both native and nonnative English speakers) to increase their vocabularies and to provide them with skill in recognizing and understanding the meaning of roots and affixes.

FANG-LIAN LIAO
University of California, Los Angeles

Words for Students of English: A Vocabulary Series for ESL. Pittsburgh: The University of Pittsburgh Press.

Vol. 1 (1988). Rogerson, H. D., Davis, B., Hershelman, S. J., Jasnow, C. Pp. vii + 183.
Vol. 2 (1988). Rogerson, H. D., Esarey, G., Schmandt, L. M., Smith, D. A. Pp. vii + 202.
Vol. 3 (1988). Rogerson, H. D., Hershelman, S. T., Jasnow, C., Moltz, C. Pp. vii + 216.
Vol. 4 (1988). Rogerson, H. D., Esarey, G., Jasnow, C., Hershelman, S. T., Smith, D. A., Moltz, C., Schmandt, L. M. Pp. vii + 233.
Vol.5(1989). Rogerson, H. D., Jasnow, C., Hershelman, S. T., Pp. vii + 257.
Vol. 6 (1989). Rogerson, H. D., Esarey, G., Jasnow, C., Hershelman, S. T., Smith, D. A., Snellings, C. M. Pp. vii + 301.
Vol. 7 (1990). Rogerson, H. D. Pp. vii + 231.

Words for Students of English: A Vocabulary Series for ESL is a very useful, well-written, and well-organized set of vocabulary texts designed for different purposes. The texts can serve as core texts, supplementary texts, or as texts for individual study. Volumes 1 and 2 are designed for beginners; Volumes 3 and 4, for intermediate; Volumes 5 and 6, for high-intermediate/advanced; and Volume 7 for advanced students.

The first six volumes have 25 units with the following format: word form chart, definitions and examples, introductory exercises, study exercises, and follow-up exercises. Each unit is designed around a specific topic. The exercises introduced in these volumes are sequenced from "controlled exercises through more difficult to a final phase with communicative exercises" (Vol. 1, p. xiv). These volumes, which assume prior knowledge of 600 base words, introduce approximately 300 new base words each. The selection of the 300 lexical items for each volume was based upon

"usefulness from a variety of word-frequency lists" (Vol. 1, p. xii). Most of the selected words of the first four volumes as well as the 600 base words coincide with the General Service List of English Words (West, 1953), and the Cambridge English Lexicon (Hindmarsh, 1980), whereas Volumes 5 and 6 correspond to the university word list (Nation, 1990).

Volume 7 contains 22 units divided into two categories for receptive learning and productive use; these are well designed with pertinent exercises included for both categories. The volume includes approximately 300 lexical items for production and 400 for recognition. The lexical items introduced in this volume seem to be at the 10,000-word level.

In the word form chart, the new lexical items are categorized according to their grammatical functions, such as verbs and nouns. Additional related and derived forms are included progressively. New words are introduced gradually from 10 to 32 base words in each unit as students advance to the next higher level.

In the definition and examples section, each new word introduced is defined, and example sentences are presented. "In all definitions and examples, only previously learned words are used" (Vol. 1, p. ix). In its use of unambiguous contexts, this method is consistent with current research findings, since occurrence in ambiguous contexts has proved to be an unfavorable condition for the retention of the new words (Schouten-van Parreren, 1989, p. 78). However, guessing word meanings in context and analyzing word forms are also important factors for retaining new words (Taylor, 1983, cited in Nation, 1990). Thus, giving all the definitions of the new words at the initial stage may unfavorably affect learners' retention of the new words. It may be helpful for beginning-level students to become accustomed to a bilingual dictionary, which can be a very effective tool in learning a second language (Baxter, 1980). For intermediate and advanced students, however, exercises for guessing word meanings in context should be included. This perspective aligns with the authors' suggestion in Volume 1 that the word form chart and definition and examples sections can be used for intensive individual study, followed by discussion in class of any questions raised by individual students.

The introductory exercises consist mainly of matching, short answer, and fill-in-the-blanks, which are easy to do and well organized so that learners will be able to handle them with confidence. It would also have been helpful to include association exercises such as acoustic-link exercises for beginners and imaginary-link (keyword approach) exercises for intermediate/advanced students, since generative processing of vocabulary is most helpful for learners' retention of new words (see Wittrock & Carter, 1975).

The study exercises contain a variety of activities that are well conceived and well selected for both receptive and productive learning. The association exercises, the two- and three-word verb exercises, and the analogy exercises introduced in Volumes 5 and 6 are especially helpful for productive practice. The final section of each unit presents many more exercises that encourage productive practice.

This series, consisting of a variety of exercises, is well conceived and systematically organized. It also offers clear instructions and guidance for learners and teachers. Each volume contains an index listing all the new words presented. In sum, I believe that this series, developed by ESL teachers, has much to offer both instructors and students.

REFERENCES

Baxter, J. (1980). The dictionary and vocabulary behavior: A single word or a handful? *TESOL Quarterly, 14* (3), 325-336.

Hindmarsh, R. (1980). *Cambridge English lexicon*. Cambridge: Cambridge University Press.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Schouten-van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts? *AILA Review 6*, 75-84.

West, M. (1953). A general service list of English words. Longman: Longman, Green.

Wittrock, M. C., & Carter, J. F. (1975). Generative processing of hierarchically organized words. *American]ournal of Psychology, 88* (3), 489-501.

YUICHI TODAKA
University of *California, Los* Angeles

***People Are Funny: Pictures for Practice* (Book 1).** Donald R. H. Byrd and Stanley J. Zelinski III. New York: Longman, 1987. Pp. 76.

Comics and Conversation: Using Humor to Elicit Conversation and Develop Vocabulary. Joan Ashkenas (with Sergio Aragones, illustrator). Studio City, CA: Jag Publications, 1985. Pp. iv + 30.

The prevailing justification for using cartoons in the classroom is their popularity with children, adolescents, and adults along with the ability of humor to reduce anxiety and enhance group affinity. Humor in ESL texts thus provides an excellent vehicle for teaching target language vocabulary and structure. In addition, the cultural milieu of cartoons can be the basis for developing students' experiential maps to guide vocabulary expansion.

With such benefits in mind, I review two ESL texts that use cartoons. A few words of caution, though, are in order. First, the reliance on humor might prove excessive and vitiate students' sense of the material's serious underlying purpose. Second, because they offer a wide variety of cartoon situations, these books do not provide clear opportunities for vocabulary recycling though they offer activity suggestions like role play, captioning, and retelling of stories orally or in writing.

People are Funny is designed to be a core text for a multiskills course. It is the lower level of a two-book series for high-beginning adolescent or

adult students. The uncaptioned four-frame black, white, and gray cartoons depict ironic situations, such as two men arguing after a car accident while their wives calmly exchange the necessary information, or a princess who kisses a frog and turns into a frog as a result. The stories and drawings are sophisticated and their focus on particular events makes presenting key vocabulary simple. The characters' ethnicity and ages are varied and no group is demeaned. The units' layouts are clear, with large, comfortable print. Each unit provides lists of key lexical and grammatical items, and simple instructions for activities. The exercises, too, are varied: word associations, cloze, sentence combining, and guided oral and written discourse. In addition, the first answer of each exercise is completed as an example, which is an excellent strategy for students at this level, who may be very insecure. There is also an answer key at the back of the book. The book is easy for teachers to use, too. The table of contents offers quick reference, listing the cartoons both by function and grammatical focus. Unfortunately, there is little built-in opportunity for vocabulary recycling, so teachers will need to devise additional activities to reinforce words.. However, it is easy to extend the exercises provided.

Comics and Conversation is designed to "elicit conversation and broaden vocabulary" (p. iii) of adolescent through adult learners. The authors do not provide exercises for their uncaptioned cartoons, only general suggestions for teachers to create oral or written activities for pairs or groups. The suggestions include role play, captioning, retelling the stories, and "sequencing" (recreating the story sequence from scrambled presentations of the cartoon). While rather routine, the suggested activities are very appropriate. The authors provide three- to seven-frame blackline cartoons intended to be photocopied or used with an overhead projector. The book offers a wider variety of situations and topics than *People are Funny,* ranging from everyday events to archaeology and panning for gold. The visual complexity of the cartoons also varies widely, so teachers need to gauge their purposes and class abilities carefully. While the author's goal is to practice productive vocabulary, the often visually complex cartoons might require a great deal of explanation or translation, at least for the names of the objects in the pictures. In addition, like *People are Funny,* the lack of situational redundancy requires teachers to design appropriate exercises to recycle and reinforce vocabulary and structures.

In conclusion, cartoons are a valuable adjunct for teaching and reinforcing vocabulary in its cultural context. Judiciously chosen and deployed, they can provide a concrete and memorable basis for receptive and productive vocabulary enrichment.

MICHAELA SAFADI
University of California, Los Angeles

BRIEF REPORTS AND SUMMARIES

The *TESOL Quarterly* invites readers to submit short reports and updates on their work. These summaries may address any areas of interest to *Quarterly* readers. Authors' addresses are printed with these reports to enable interested readers to contact the authors for more details.

Edited by GAIL WEINSTEIN-SHR
University of Massachusetts at Amherst

Business English and the Case Method: A Reassessment

FRANCES A. BOYD
Columbia University

■ As markets become increasingly global, the demand for managers with language skills and cultural sensitivity will grow. One way for the language teaching community to respond to this challenge is by redefining the goals of English for business purposes and rethinking the use of the case method.

Since the early 1980s, the study of English for business purposes has attracted growing interest. Business now outranks engineering as the most popular field of study among international students in the U.S. (Zikopoulos, 1990). In the 1990s, interest in international business and economics is bound to increase with the advent of the European Community, liberalization of East European economies, and expansion of U.S. commercial ties with Asia. With these dramatic developments and the opportunities they present to the language teaching community, it is important to take a fresh look at the subdiscipline of English language teaching known as English for specific purposes (ESP) for business, or ESP-B, including a reassessment of the case method—no single theme has been more prominent in discussions of curriculum and methodology for business English. Judging by professional and commercial publications, there is widespread agreement that this method borrowed from graduate business schools is the most appropriate pedagogical model for ESP-B. The consensus on this point is so strong that two basic assumptions are seldom challenged: that the purposes of ESP-B and MBA (Master of Business Administration) programs are essentially similar and therefore the same methodology and curriculum are appropriate, and that the case method is a rigidly defined procedure requiring a specific kind of case text and discussion.

This paper will examine these assumptions, first by clarifying and comparing the purposes of ESP-B and MBA programs, then by considering how the case method has been adapted by other disciplines.

On this basis, suggestions for an ideal version of the case method that satisfies the specific needs of ESP-B will be generated.

THE SPECIFIC PURPOSES OF ESP FOR BUSINESS

As a relatively new area within the field of English language teaching, ESP-B is still being defined. I have found it useful to define the purpose of ESP-B as the development of communicative competence for business settings, where communicative competence refers to Canale and Swain's (1980; see also Canale, 1983) classification of language skills into the areas of grammatical, sociolinguistic, discourse, and strategic competence. Among these, advanced students interested in business are usually most in need of sociolinguistic and discourse competence. Business settings refers to the various situations calling for language use that one might encounter on the job, for example, scheduled meetings, desk work, unscheduled meetings, telephone calls, and tours (Mintzberg, 1973). The location where these situations occur—in North America, England, or other places where English is the international language of business—may also be an important feature of the business setting.

Communicative competence for business settings is a useful definition of ESP-B because it clearly relates the field to the larger field of English as a second language. As in ESL, language is understood in its broadest sense as communication, not simply as knowledge of vocabulary and grammar. Setting, then, as the context for language use, becomes the distinguishing feature of the subfield. This definition also distinguishes ESP for business from ESP for academia, more commonly called English for academic purposes (EAP). The cognitive skills that are the domain of EAP, such as note taking, deducing, or summarizing, cannot be usefully separated by discipline.

Beyond clarifying certain necessary distinctions, the definition seeks to heal a confusing and unnecessary breach in the literature between those who see ESP-B as academic preparation (see, for example, Canseco & Byrd, 1989; Micheau & Billmyer, 1987; Westerfield, 1989) and those who see it as professional preparation (see, for example, Huckin & Olsen, 1984; Piotrowski, 1982; Richards, 1989). By stressing what these two orientations have in common, this definition recognizes that an MBA program is simply an alternate route to the business world and that many of the same communicative skills are necessary for effective performance in both business and academic settings.

This concept of ESP for business allows us to contrast ESP-B programs with MBA programs. As professional education, graduate programs in business administration have broad, long-range goals which relate to a particular definition of the nature of expertise (Kennedy, 1987). An influential contemporary view, described by Schon (1985), identifies a distinctive way of thinking—Schon calls it "reflection-in-action" (p. 54)–as the defining characteristic of the professional. Programs built on this view stress the development of an approach to problems and of a

professional self-image over the mastery of a fixed body of knowledge. In Schon's words, students learn to engage in "reflective conversations" in which they "name and frame" problems, explore "chains of inferences, " then develop courses of action (pp. 128-132).

With this background, it is now possible to identify significant differences between ESP-B and professional education for business. Simply put, the main goal of ESP-B is not to teach students how *to think* like business managers, but rather how *to communicate* like business managers in English. This narrower focus on communicative skills arises naturally from the needs of the population for business English. While nonnative-speaking students may be at least as varied in age, education, and experience as MBA candidates, they lack sociolinguistic and discourse competence in English. It is precisely these linguistic and cultural gaps that can best be addressed in specialized English courses taught by instructors whose expertise and experience lie in these areas.

THE CASE METHOD IN PROFESSIONAL EDUCATION

The case method has a long and varied history as part of the curriculum of programs in professional education. Originating in the 1870s in law, then adopted by business administration in the 1920s, it has also been applied to a lesser extent in education, social work, and other fields (Boyd, 1984; Gartner, 1976; McNair, 1954). Generally, the case method is used in conjunction with lecture and clinical approaches. Students gain theoretical knowledge from lectures and practical knowledge from clinical experience; in contrast, the case method functions as a laboratory where, in a short time and in a setting removed from the pressures of an authentic work environment, students try to solve a large number of representative problems by applying theory to practice.

Clearly the case method has persisted because it fits a major goal of professional education: providing a "sheltered situation" (Gartner, 1976, p. 215) in which students can practice reflection-in-action while playing the role of a professional. Nevertheless, each field has freely adapted the method to suit its specific purposes.

In business education, students define a problem at an actual firm, suggest courses of action, and then, taking into account all the risks and uncertainties inherent in the real business world, select a solution and justify it. In legal education, students analyze appellate court decisions using Socratic dialogue in order to discover precedents for future cases. In the field of education, the case consists of a description of a critical incident which is then systematically analyzed from various theoretical perspectives. What some English language teachers have called problems with case studies (Grosse, 1988; Westerfield, 1989)—case length, level of linguistic difficulty, assumption of cultural background knowledge—are caveats only about the particular set of curriculum materials used in business schools. The method itself is a flexible tool that can be adapted to

many educational settings and purposes, and this flexibility has not yet been fully exploited by language instructors.

A CASE METHOD FOR ESP FOR BUSINESS

How, then, should the case text and discussion be adapted to develop the specific language skills so critical to nonnative speakers in the business world?

Some helpful clues can be gleaned from linguistic and cross-cultural studies of the business setting. Negotiation, for example, has been identified as a common speech event in business with important cross-cultural variations. Neu (1986) has mapped the structure of U.S. negotiations, while Graham (1985) has cataloged the verbal and nonverbal bargaining behavior of business people of various cultures. Though their data come from business case discussions, Micheau and Billmyer's (1987) findings about turn-taking behavior and its violation by nonnative speakers are also applicable to business settings. Information from these studies can be used by case writers to create negotiating situations, select gambits, point out the significance of nonverbal language, and help nonnative speakers adjust to unfamiliar patterns of sustaining a conversation, pausing, and interrupting.

Another aspect of managerial work that has been studied from a cross-cultural perspective is letter writing. Jenkins and Hinds (1987) found that the ritualized discourse of the business letter of request in English, French, and Japanese reveals important cultural meanings. The knowledge that the form and content of these letters do not simply reflect the idiosyncratic style of the individual can be useful to international business managers as writers and readers of business letters. More research of this kind would help case writers, but even from this incomplete picture a new version of case text and case discussion for business English begins to emerge.

What follows are suggestions for the ideal case texts and case discussions for developing communicative competence for business settings. Cases of this nature might form only one part of the business English curriculum, which could also include guest lectures, on-site visits, and special lessons in pronunciation and presentation skills.

In general. the ideal case for ESP-B would borrow from both the business and education models by combining an authentic business problem with details of employees and the organizational culture. The problem-solving case discussion would maintain its central role, though neither its form nor length, and it would be framed by several other interactive exercises.

Instead of presenting one long reading, the ESP-B case would break up information about the company into smaller units and use audio and visual media as well as print. To make exploration of the cultural issues possible, cases would give a wealth of background information on undisguised companies that produce recognizable products and have interesting

histories, flesh-and-blood executives, and authentic business correspondence. They would take place in firms varying in size, corporate culture, and location, and explore problems with broad appeal. The language level of reading and listening materials would be that of the general-interest publication rather than the specialized business press, the *New York Times* rather than *Business Week*, unless familiarity with business jargon is a priority.

Some exercises would focus on language, particularly on culturally appropriate communication skills useful in negotiations. Others might highlight vocabulary, grammatical items useful for case discussions, or discourse patterns in business correspondence. The activities would culminate in a version of the problem-solving case discussion.

The case discussion would most likely take the form of a simulated business meeting. If a case can be a "virtual world" (Schon, 1985, p. 162) of work, then participants will want to practice the kinds of interactions they imagine take place in that world. Also, the empathic element required in role playing may lead to cultural insights.

The bulk of class time would be spent not in case discussion but in a series of varied interactions with classmates and instructor. In small groups and pairs, students would gather, organize, compare, and evaluate the facts and opinions relevant to the company and its problem.

In such an environment, the teacher's role would be that of process and language consultant, not discussion leader. As process managers, teachers would be responsible for the organization, pace, and quality of the interaction. As language and culture consultants, they would give written or taped feedback on language errors and act as guides on cultural matters. For feedback on sociolinguistic aspects, video is especially useful (Micheau & Billmyer, 1987; Westerfield, 1989).

In sum, the ideal ESP-13 case would preserve the spirit of the practice-centered method for professional education while focusing on the specific purposes of the business English course.

This discussion of purpose and pedagogy for business English is just one example of the debate between a focus on language or content that has been an issue in the field of ESL for a long time. About the larger question, Mohan (1986) writes: "In subject matter learning we overlook the role of language as a medium of learning. In language learning we overlook the fact that content is being communicated" (p. 1). And about business English in particular, Westerfield (1989) says: "To take a 'language teaching approach' and simplify the [business school] cases at this point would be somewhat misleading and certainly not as appealing to advanced students desiring 'real' course content" (p. 77).

I have argued that the "real" course content of business English is the development of communicative competence for business settings through specially designed cases and other activities, and that a "language teaching approach," though this is not equated with simplification, is essential. ESP-B is not a business course, but a language course.

REFERENCES

Boyd, F. A. (1984). *Educating professional educators by the case method.* Ann Arbor, MI: University Microfilms International.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt (Eds.), *Language and communication.* London: Longman.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1,* 1-47.

Canseco, G., & Byrd, P. (1989). Writing required in graduate courses in business administration. *TESOL Quarterly, 23* (2), 305-316.

Gartner, A. (1976). *The preparation of human service professionals.* New York: Human Science Press.

Graham, J. L. (1985). The influence of culture on the process of business negotiations: An exploratory study. *Journal* of *International Business Studies, 16* (1), 79-94.

Grosse, C. U. (1988). The case study approach to teaching business English. *English for Specific Purposes, 7* (2), 131-136.

Huckin, T. N., & Olsen, L. A. (1984). The need for professionally oriented ESL instruction in the United States. *TESOL Quarterly, 18* (2), 273-294.

Jenkins, S., & Hinds, J. (1987). Business letter writing: English, French, and Japanese. TESOL *Quarterly, 22* (2), 327-345.

Kennedy, M. M. (1987). Inexact sciences: Professional education and the development of expertise. In E. Z. Rothkopt (Ed.), *Review of research in education* (pp. 133-167). Washington, DC: American Educational Research Association.

McNair, M. P. (Ed.). (1954). *The case method at the Harvard Business School.* New York: McGraw-Hill.

Micheau, C., & Billmyer, K. (1987). Discourse strategies for foreign business students: Preliminary research findings. *English for Specific Purposes, 6* (2), 87-97.

Mintzberg, H. (1973). *The nature of managerial work.* New York: Harper& Row.

Mohan, B. A. (1986). *Language and content.* Reading, MA: Addison-Wesley.

Neu, J. (1986). American English business negotiations: Training for non-native speakers. *English for Specific Purposes, 5* (1), 41-57.

Piotrowski, M. V. (1982). Business as usual: Using the case method to teach ESL to executives. *TESOL Quarterly, 16* (2), 229-238.

Richards, K. (1989). Pride and prejudice: The relationship between ESP and training. *English for Specific Purposes, 8,* 207-222.

Schon, D. (1985). *The reflective practitioner.* New York: Basic Books.

Westerfield, K. (1989). Improved linguistic fluency with case studies and a video method. *English for Specific Purposes, 8* (1), 75-83.

Zikopoulos, M. (Ed.). (1990). *Open doors 1989/90: Report on international educational exchange.* New York: Institute of International Education.

Author's Address: American Language Program, 505 Lewisohn, Columbia University, New York, NY 10027

Basic Skills Tests and Academic Success of ESL College Students

MARK S. PATKOWSKI
Brooklyn College, The City University of New York

Institutions of higher education throughout the United States widely employ scores obtained on various tests of English language proficiency to make admission and placement decisions for minority language students. A fundamental rationale for such tests is that they predict "readiness" for college-level academic work; yet, the predictive value of these tests for ESL students has rarely been investigated and what research has been carried out has failed to produce consistently high correlations (Graham, 1987). Indeed, there is a growing concern that access to higher education for minority language students may be unduly restricted (Otheguy, 1990) because such students are repeatedly placed and kept in developmental or ESL courses while being denied access to other college courses.

The purpose of this paper, then, is to examine the relation between the academic success of ESL students at a large, public, urban college, and their scores at the time of admission on basic skills tests in reading and writing (as well as math); while recognizing that such a descriptive case study can only provide fragmentary evidence gathered at one particular campus, it is nevertheless felt that this evidence is relevant to the issues discussed above—issues which are being faced by an increasing number of ethnolinguistically diverse institutions nationwide.

METHOD

The entrance and academic records over a 3 year period of a cohort of 271 students in the ESL program at Brooklyn College of the City University of New York (CUNY) were examined. The following data were collected: (a) entering scores on the CUNY-wide basic skills tests in reading, writing, and math, which are administered to all incoming students and which determine placement into, as well as exit out of, the corresponding remediation or ESL programs; (b) "persistence" (i.e., continued enrollment in college) and academic performance in coursework.

The academic success of students was measured in two ways. The first consisted of student cumulative grade point averages (CPA) on the traditional "4 point" grading system (0 = F, 4 = A). The second (referred to here as "SCS rating") was based on a set of six criteria as follows: (a) completion of all developmental courses; (b) GPA = 2.0 or better; (c) completion of freshman composition; (d) completion of at least 6 credits in the Core Curriculum (which are required of all students); (e) completion of at least 50 credits; (f) a completion ratio of attempted credits of at least 70%.

Ratings of academic success (SCS) were then assigned as follows: 0 = student no longer enrolled; 1 = student currently enrolled, does not

meet at least 5 of the above criteria; 2 = student currently enrolled, meets at least 5 of above criteria.

RESULTS

Descriptive Statistics

The entering scores of the 271 ESL students on the basic skills tests were highest in math, with 167 (61.6%) receiving a passing score on this standardized, multiple-choice instrument. only 57 (21.0%) students obtained passing scores on the CRAT (CUNY Reading Assessment Test), also a standardized instrument, consisting of 15 short reading passages followed by multiple-choice items. Students who fail the WAT (Writing Assessment Test—a 50 minute essay examination, scored holistically according to CUNY-wide standards) and who indicate that English is not their first language are assigned to the ESL program. Thus, all but 2 students had failed the WAT (the other 2 had presumably enrolled themselves in ESL).

Of the original cohort of 271 students, 134 or 49.4% eventually passed the writing test (WAT). The reading test (CRAT) was eventually passed by 143 students, and a further 21 satisfied the reading requirement by means of an appeals procedure open to students under certain conditions, for a total of 60.5%. The best results were in math, with 89.4% of the original cohort eventually passing.

In terms of the academic success (SCS) ratings discussed above, 65 students obtained SCS ratings of 2 ("currently enrolled—meets 5 of 6 criteria"), 53 obtained ratings of 1 ("currently enrolled-does not meet criteria"), and 153 obtained O ratings ("no longer enrolled"). The cumulative GPAs for students in the three categories were 2.77, 1.90, and 1.40 respectively.

Correlations and ANOVAs

The two measures of academic success (GPA, SCS rating) were considered separately. First, a correlational analysis was carried out, using the WAT, CRAT, and math entering scores in turn as the independent variable, and CPA as the dependent variable. The results are presented below. (Note that because of missing scores, totals are below 271 and vary slightly on different measures.)

TABLE 1
Pearson Correlations Between Students' Entering Scores and Grade Point Averages

Independent variables		*r*	*p*
Math	($n = 263$)	.344	.000
CRAT	($n = 263$)	.255	.000
WAT	($n = 264$)	.169	.006

As can be seen, the best predictor of the grade point average of ESL students was performance on the math test; yet, even that correlation was very modest (an *r* of .344 only explains 11.8% of the variance). On the other hand, the relationship between performance on the writing test (WAT) and the ESL students' grade point average was the weakest. A multiple correlation with all three (reading, writing, math) scores as predictor variables and GPA as dependent variable did not lead to much improvement in the coefficient of multiple correlation (*R* = .370, explaining 13.7% of the variance). In other words, adding the reading and writing scores to the math score does not substantially improve the predictive value of these scores.

Next, the second measure of academic success (SCS rating) was considered by means of a series of one-way analyses of variance to test whether there were significant differences in entering math, reading and writing scores among students at the three levels of academic success. The results, which are presented below, were essentially similar to those of the correlational analysis; again, the math test yielded the strongest result and the writing test (WAT) the weakest.

TABLE 2
Analyses of Variance on Entering Scores of Students at Three Levels of Academic Success

Score	Source	*SS*	*df*	*MS*	*F*	*p*
Math	Between	1,324.9	2	662.5	9.55	.000
	Within	18,026.8	260	69.3		
CRAT (Reading)	Between	918.6	2	459.3	7.02	.001
	Within	17,007.9	260	65.4		
WAT (Writing)	Between	13.6	2	6.8	3.77	.024
	Within	469.4	261	1.8		

CONCLUSIONS

Two conclusions can be drawn from these findings: First, the entering scores of ESL students on the basic skills tests in reading, writing, and math do not appear to act as good predictors of subsequent academic success. Second, the reading and writing scores appear to have even less predictive value than the math scores; furthermore, the most difficult test for ESL students to pass (writing) is also the least predictive.

Comparable data on the predictability of these scores for native students would have been most useful, but were unavailable in this study. However, other reports have suggested that problems associated with basing educational decisions on placement instruments such as those examined here exist with native-speaking minority students (Davis, Kaiser, Boone, McGuire, 1990) and even more so with minority language students (Everson, 1990; Kleinman, 1989; Otheguy, 1990) to a greater degree than

they do with "mainstream" students. Otheguy, in particular, in a wide-ranging review of the placement procedures at CUNY concluded that the tests may foster differential passing rates among various ethnolinguistic groups irrespective of prior high school preparation or current academic readiness.

Such findings can only lead one to urge that important placement decisions which can greatly affect the academic futures of students should not be solely made on the basis of a single score on any particular test, but rather should depend on a wider, more "authentic" base of information in order to reduce barriers to the retention of minority language college students.

ACKNOWLEDGMENT

This report is drawn from a study *(Predictors of College Success,* 1989) supported by a SEEK-Funded Retention Research Project grant from the Office of Special Programs, The City University of New York.

REFERENCES

Davis, T., Kaiser, R., Boone, J., & McGuire, J. (1990). Caution in testing.*]ournal of Developmental Education, 13* (3), 2-4.

Everson, H. (1990, March). *Results of a study on the CUNY Reading Assessment Test.* Paper presented at the meeting of the CUNY Language Forum.

Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly, 21* (3), 505-521.

Kleinman, H. (1989). Does an additional 15 minutes make a difference on the CUNY Reading Assessment Test? In D. Pan & D. Rodriguez-Diaz (Eds.), *ESL Beyond ESL: Summaries of Selected Papers from the 1989 CUNY ESL Council Conference* (p. 36). New York: City University of New York, Office of Academic Affairs, CUNY ESL Council & The Instructional Resource Center.

Otheguy, R. (1990). *The condition of Latinos in the City University of New York: A report to the Vice Chancellor for Academic Affairs and to the Puerto Rican Council on Higher Education.* Unpublished manuscript, City College, The City University of New York.

Author's Address: Department of Educational Services (ESL), Brooklyn College, The City University of New York, Brooklyn, NY 11210

Language Teacher Supervision: Process Concerns

JERRY G. GEBHARD
Indiana University of Pennsylvania

■ Recently teacher educators have begun to focus on the teaching practicum as a viable part of language teacher preparation (Gebhard,

1990a; Gebhard, Gaitan, & Oprandy, 1990; Pennington, 1990; Richards & Crookes, 1988), and within the practicum, there is emphasis on "supervised teaching experiences as constituting the core of the practicum" (Richards & Crookes, 1988, p. 23). Recognizing this "core" importance, teacher educators have also discussed the processes through which supervision can be accomplished (Abbott & Carter, 1985; Fanselow, 1990; Freeman, 1982, 1990; Gaies & Bowers, 1990; Gebhard, 1990b, 1990c; Knopp, 1980). The purpose of this summary is to synthesize knowledge about the process of supervision, to address some of the problems which supervisors face, and to offer possible solutions to these problems.

SUPERVISORY GOALS AND PROCESS: PROBLEMS AND POSSIBLE SOLUTIONS

Teacher educators generally agree that the goals of supervision are to (a) improve instruction, (b) provide chances for teachers to develop decision-making skills and explore new teaching possibilities, (c) furnish ways for teachers to reflect on and work through problems in their teaching, and (d) afford teachers chances to develop a personal theory of teaching.

To reach these goals, supervisors generally follow "the cycle of supervision" (Cogan, 1973, p. 10). The supervisory cycle begins with a preobservation conference to choose aspects of the teacher's instruction on which to focus attention. Next, the supervisor observes the teacher, focusing on the agreed upon area of teaching. The supervisor and teacher then meet to discuss the class, using the supervisor's script as the center of discussion. The process then cycles back again to decisions about what areas of teaching to focus on next.

Although this supervisory process allows for the observation and possible change of the teacher's classroom behaviors in a systematic way, there are problems with the process, especially concerning how it relates to the goals of supervision. In the rest of this summary, problems and possible solutions are discussed.

Problem 1: Defining Improved Instruction

Since one of the aims of clinical supervision is to improve instruction, one problem supervisors face is in defining improvement. Teacher educators generally recognize that the relationship between teaching and learning is complex; not enough is known about how teaching behaviors result in student learning to specify improvement as it relates to student learning in all contexts. As Chaudron (1988) points out, "if any one conclusion is to be drawn from . . . a review of learning outcomes, it is that much more research remains to be conducted to determine what aids learners' target language development" (p. 179). How, then, can the goal of the supervisory process be to improve instruction?

After considering this question, Gaies and Bowers (1990) suggest that there is more than one way to perceive improvement. Rather than basing

the concept of improvement on the relationship between teaching behaviors and learning outcomes, improvement can be based on what teachers are expected to be doing in a specific teaching context, as defined by teachers and supervisors within this context, as compared to what they actually do. For example, if the teacher and supervisor within a particular teaching context value teacher movement in the classroom, and if the teacher stands in one spot for the entire class period, improvement could be realized by this teacher moving freely about the class.

Problem 2: Decision Making

A second concern is how supervisors can provide chances for teachers to cultivate decision-making skills. In the usual supervisory process, as described earlier, the supervisor observes with the intent of giving feedback on selected aspects of teaching behavior. Although this approach does focus on teachers' classroom behavior, it does not show how teachers can make their own informed teaching decisions. This is mainly because the supervisor takes on much of the responsibility for selecting, observing, and offering feedback on teachers' classroom behavior.

In order to move beyond a supervisor-dominant model and to provide chances for teachers to make their own informed teaching decisions, Fanselow (1987, 1990) advocates that the supervisor guide the teacher through a process of exploration of classroom behaviors and their consequences. He shows how teachers can do this by (a) collecting samples of classroom interaction in the form of notes, sketches, and tape recordings; (b) studying these samples to identify patterns of interaction, possibly through the use of transcriptions from tapes and observation instruments, such as Fanselow's (1987) FOCUS (Foci for Observing Communications Used in Settings), Allen, Fröhlich, and Spada's (1984) COLT (Communicative Orientation of Language Teaching), or descriptive checklists as presented by Day (1990); (c) interpreting this patterned interaction; and (d) making decisions about what to do next based on analysis and interpretations. The goal of this approach to supervision is to show teachers how to make their own informed teaching decisions by understanding the underlying rules of classroom interaction. (See Gebhard, 1990c, for an illustration of this process.)

Problem 3: Emphasis on Training

A third concern is that the supervisory process, as described earlier, focuses on training teachers to isolate, practice, and master specific behaviors such as questioning techniques, wait time, teacher talk, and use of praise behaviors. However, although training teachers to manipulate discrete behaviors is thought to be useful, teacher educators now place emphasis on teacher *development* (Fanselow, 1987, 1990; Freeman, 1989, 1990; Gebhard, 1990a; Larsen-Freeman, 1983; Richards, 1990a, 1990b; Richards & Nunan, 1990). Teacher educators see an emphasis on training

as limiting because some aspects of teaching cannot be mastered by teachers as discrete chunks. For example, as Richards (1990a) points out, classroom management cannot be taught and practiced as an isolated behavior because it includes the complex ways in which student behavior, classroom space, and language task are organized by the teacher to accomplish teaching.

Teachers need to develop such complex aspects of teaching through time and experience. Supervisors, breaking from the usual "cycle of supervision," can contribute to this development. It stands to reason that the supervisor needs to go beyond a one-to-one relationship by providing ways for teachers to participate in a wider variety of activities, including but not limited to doing action research (Nunan, 1990), observing for self-knowledge (Bailey, 1990; Bartlett, 1990; Fanselow, 1990; Porter, Goldstein, Leatherman, & Conrad, 1990), and participating in problem-solving activities (Celce-Murcia, 1983; Ellis, 1990).

CONCLUSION

Although the usual process of supervision provides the means for supervisors and teachers to focus on the teacher's actual classroom behavior, supervisors need to pay attention to how this process relates to the goals of supervision. To reach the goal of improved instruction, the focus needs to be placed on establishing and working toward acceptable ways to teach within a particular teaching context by the teacher and supervisor rather than the supervisor prescribing a "best" way to teach. While working within the process, supervisors need to show ta, 1983;g a990), as .4late 655.92 isors need to

Celce-Murcia, M. (1983). Problem solving: A bridge builder between theory and practice. In J. E. Alatis, H. H. Stern, & P. Strevens (Eds.), *Georgetown University Round Table on Languages and Linguistics 1983: Applied linguistics and the preparation of second language teachers: Toward a rationale* (pp. 97-105). Washington, DC: Georgetown University press.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning.* New York: Cambridge University Press.

Cogan, M. L. (1973). *Clinical supervision.* Boston: Houghton Mifflin.

Day, R. R. (1990). Teacher observation in second language teacher education. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 43-61). New York: Cambridge University Press.

Ellis, R. (1990). Activities and procedures for teacher preparation. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 26-36). New York: Cambridge University Press.

Fanselow, J. F. (1987). *Breaking rules: Generating and exploring alternatives in language teaching.* New York: Longman.

Fanselow, J. F. (1990). "Let's see": Contrasting conversations about teaching. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 182-199). New York: Cambridge University Press. (Reprinted from *TESOL Quarterly, 1987, 22* (l), 113-120)

Freeman, D. (1982). Observing teachers Three approaches to in-service training and development. *TESOL Quarterly, 16* (1): 21-28.

Freeman, D. (1989). Teacher training, development, and decision making: A model of teaching and related strategies for language teacher education. *TESOL Quarterly, 23* (l): 27-45.

Freeman, D. (1990). Intervening in practice teaching. In J. C. Richards and D. Nunan (Eds.), *Second language teacher education* (pp. 103-117). New York: Cambridge University Press.

Gaies, S., & Bowers, R. (1990). Clinical supervision of language teaching: The supervisor as trainer and educator. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 167-181). New York: Cambridge University Press.

Gebhard, J. G. (1990a). Interaction in a language teaching practicum. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 118-131). New York: Cambridge University Press.

Gebhard, J. G. (1990b). Models of supervision: Choices. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 156-166). New York: Cambridge University Press.

Gebhard, J. G. (1990c). Freeing the teacher: A supervisory process. *Foreign Language Annals, 23* (6), 517-525.

Gebhard, J. G., Gaitan, S., & Oprandy, R. (1990). Beyond prescription: The student-teacher as investigator. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 16-25). New York: Cambridge University Press. (Reprinted from *Foreign Language Annals,* 1987, *20* (3), 227-232).

Knopp, C. K. (1980). The supervision of foreign language teachers. In F. M. Grittner (Ed.), *Learning a second language: Seventy-ninth yearbook of the National Society for the Study of Education* (Part 2, pp. 186-207). Chicago: University of Chicago Press.

Larson-Freeman, D. (1983). Training teachers or educating a teacher. In J. E. Alatis, H. H. Stern, & P. Strevens (Eds.), *Georgetown University Round Table on Languages and Linguistics 1983: Applied linguistics and the preparation of second language teachers: Toward a rationale* (pp. 264-274). Washington, DC: Georgetown University Press.

Nunan, D. (1990). Action research in the language classroom. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 62-81). New York: Cambridge University Press.

Pennington, M. C. (1990). A professional development focus for the language teaching practicum. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 132-152). New York: Cambridge University Press.

Porter, P. A., Goldstein, L. M., Leatherman, J., & Conrad, S. (1990). An ongoing dialogue: Learning logs for teacher preparation. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 227-241). New York: Cambridge University Press.

Richards, J. C. (1990a). The dilemma of teacher education in second language teaching. In J. C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 3-15). New York: Cambridge University Press. (Reprinted from *TESOL Quarterly, 1987, 21* (2), 209-226).

Richards, J. C. (1990b). Beyond training: Approaches to teacher education.*The Language Teacher, 14* (12), 3-8.

Richards, J. C., & Crookes, G. (1988). The practicum in TESOL. *TESOL Quarterly, 22* (l), 9-27.

Richards, J. C., & Nunan, D. (Eds.). (1990). *Second language teacher education.* New York: Cambridge University Press.

Author's Address: English Department, 110 Leonard Hall, Indiana University of Pennsylvania, Indiana, PA 15705

THE FORUM

The *TESOL Quarterly* invites commentary on current trends or practices in the TESOL profession. It also welcomes responses or rebuttals to any articles or remarks published here in The Forum or elsewhere in the *Quarterly*.

Comments on Alistair Pennycook's "The Concept of Method, Interested Knowledge, and the Politics of Language Teaching"

A Reader Reacts. . .

MARTHA McCALL
Portland Public Schools

In his recent *TESOL Quarterly* article (Vol. 23, No. 4), Alastair Pennycook presents a historical and political analysis of second language teaching. Starting with the premises that "all education is political" and "all knowledge is 'interested'" (p. 590), the author questions claims of objectivity in science as a whole, in analytical views of language and education, and in empirical claims for the superiority of Method in second language teaching. His analysis demonstrates some problems with a purely political view of reality.

Paying particular attention to Krashen's (1982) theories, Pennycook wants his readers to recognize that although the natural approach is itself a reaction to linguistic structuralism, it suffers from the same prescriptivist, positivist, scientific biases. It, too, uses apparently scientific proof to promote teaching practices which are represented as different from and superior to previous approaches but are nothing more than current configurations of language teaching techniques that have been in and out of vogue for centuries. Pennycook shows how studies presented as proof of the superiority of Method over structure are often either misinterpreted or chosen only because their conclusions support the opinion of an expert, whose career is thereby bolstered.

The notion that all knowledge exists in the service of power, as set forth in the writings of Freire, Foucault, and Giroux (and cited in Pennycook), has set the terms of the debate on educational policy

and has transformed our classrooms and our minds by giving teachers a mandate to challenge all knowledge and an ideological framework for doing so. The role of the progressive teacher is to reveal the self-interested nature of the subject matter and of the educational process so that the student gains not only a competent understanding of the subject but also an active critical involvement with it. The ideas of Krashen and Terrell (1983) initially appealed to political analysts of language teaching because they minimized the presentation of grammatical structure; (whose pretense to objectivity validates elitism), they allowed student experience to determine class content (so that authority is shared), and they challenged entrenched audiolingual and cognitive theories. Natural and whole language approaches thus enjoyed a brief immunity from political criticism. But Krashen and Terrell favor these methods because they are effective, not because they support a particular position about knowledge and power. Now that their theory has become dominant in education, its authority must also be challenged and its biases revealed. Pennycook does so by analyzing the use of academic prestige to undermine the professional judgment of teachers and to stimulate publication of minimally improved textbooks. He calls attention to similarities between approaches based on Method and earlier ones organized around structure and finds the concept of Method so lacking in rigor that, under scrutiny, it collapses into incoherence.

Pennycook's history of recurring themes in language teaching provides a refreshing look at the natural approach, some of whose advocates reject all previous teaching styles with an ardor mystifying to the unconverted. While it should be noted that Krashen and Terrell are less dogmatic than some of their followers, overly narrow application of their ideas has been used to rigidly spurn traditional techniques. However, as Pennycook notes, most teachers use a more complex, supple reading of these ideas, retaining devices that have merit and declining those that do not. But when administrators, academics, and curriculum "experts" adopt a misleadingly simplistic view of whole language methods, for example, teachers and students find themselves forced to defend structural forms and the right to use them. The following examples of this come from the pages of the *TESOL Quarterly*.

Thomas Tinkham (1989), whose investigation into rote learning produced "visible shudders" (p. 695) in his colleagues, recommends that "instead of designing and teaching materials that reflect their own culturally bound strengths and attitudes, educators should take into consideration the strengths and attitudes of their students and take advantage of what those students bring to the classroom"

(p. 697). As Tinkham notes, rote learning is among both the most maligned and most commonly used techniques. A guide to its use in the service of communication might be more useful than wholesale prohibition.

Gathercole (1989) replies to a critic who fears that her recommendations will propel English language teaching back to darkest structuralism,

> We simply need to acknowledge concomitant roles for production and comprehension, for routines and rule formation and for some attention to linguistic structure alongside communicative intentions. Simply because the former in each pair was at one point identified with audiolingualism does not mean we have to throw it out like the baby with the bath water. (p. 161)

Dickens and Woods (1988) use the same image as they discuss ways conventional materials might be used in a communicative curriculum. "Too often, recently, instead of looking carefully at what is already available to us and how this may be used to reflect some of the criteria of communicative grammar tasks, 'the baby has been thrown out with the bath water'" (p. 636).

Burnaby and Sun (1989) caution Western educators that although it is important that communicative methods be introduced to China, this must be done slowly and with full awareness of cultural context. Furthermore, they assert, teachers in English-speaking countries could learn much from Chinese language teaching models rather than assuming that methods that are traditional must also be unenlightened.

Pennycook urges teachers like these, frustrated by academics who dismiss their concerns as throwbacks to some primitive era, to challenge experts and to trust their own knowledge and perceptions. His words are welcomed by teachers who need support for a critical stance toward orthodox methods, but they illustrate some limitations of political analysis.

The point here is not that the natural approach should be abandoned, but that it exemplifies general problems with an ideological preference for "Method" which is no more virtuous or disinterested than "structure," although perhaps no less useful in the classroom. Pennycook comments: "The 'new methods' seem to reduce to a constellation of techniques, which is not to say that there is nothing valuable to be gained from them" (p. 605). In other words, neither being intellectually outmoded nor academically dominant prevents a procedure from being effective. However, having demonstrated that Method is no better or worse than any other theory, the author gives us no principled way to choose what

to teach and how to teach it. This is a paradox within the political view of knowledge. All ideas are viewed as manifestation of power with no truth of their own. Since proof of validity is likely to be fruitless, personal opinion is the highest authority, and any method can be regarded as legitimate. Pennycook objects to the natural approach because of its pretense to superiority and the oppressive weight of its prestige, not because of any deficiencies it might have as a way to teach. Instead of noting that rigid proscription of memorization, phonics, and grammar is a misreading of the original theorists (who say only that such methods should not be divorced from meaning), he gives credence to those few teachers who use traditional techniques in a deadening or authoritarian way.

Of course, the irony here is that Pennycook is an expert promoting distrust of experts. As he well knows, teachers will not completely accept or reject his guidance. Most will appreciate his excellent description of the relationship between politics and learning while continuing to engage in a winnowing of truth from falsehood and right from wrong. In short, they will treat him like any other expert, valuing his fine observations even as they examine his prejudices. Their skepticism will be the triumph of a generation of writers who, like Pennycook, urge them to question truth and to challenge authority.

■

REFERENCES

Burnaby, B., & Sun, Y. (1989). Chinese teachers' views of Western language teaching: Context informs paradigms. *TESOL Quarterly, 23* (2), 219-238.

Dickens, P. M. R., & Woods, E. G. (1988). Some criteria for the development of communicative grammar tasks. *TESOL Quarterly, 22* (4), 623-646.

Gathercole, V. (1989). Comments on "Some myths you may have heard about first language acquisition": The author responds. *TESOL Quarterly, 23* (l), 157-162.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Oxford: Pergamon Press.

Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom.* Oxford: Pergamon Press.

Tinkham, T. (1989). Rote learning, attitudes, and abilities: A comparison of Japanese and American students. *TESOL Quarterly, 23* (4), 695-698.

The Author Responds. . .

ALASTAIR PENNYCOOK
The Ontario Institute for Studies in Education

In an interesting response to my article, Martha McCall has raised an issue of some importance to which I would like to take this opportunity of replying. While generously acknowledging the significance of some of the points I was trying to make, she draws attention to what she describes as "some problems with a purely political view of reality. " She argues that there is "a paradox within the political view of knowledge," namely, that "having demonstrated that Method is no better or worse than any other theory," I offer "no principled way to choose what to teach and how to teach it. " Ultimately, therefore, McCall suggests that my arguments give "credence to those few teachers who use traditional techniques in a deadening or authoritarian way. " Now while I do not quite accept parts of this analysis—rather than demonstrating that Method is no better or worse than any other theory, for example, I tried to show that the concept of Method was an inadequate and interested means of describing and evaluating teaching practices—I think McCall is pointing to a key issue. Thus, while she appears to accept much of my analysis of the conceptual incoherence of Method and the interests served by its maintenance, she suggests that we are then left with no way of distinguishing between good and bad teaching practices.

In trying to reply to this point, I would like to reiterate briefly some of the orientations of my own work so that I can relocate McCall's discussion within a broader context. I have been trying to use poststructuralist and postmodernist ways of thinking about knowledge not because they happen to be in vogue in various domains (though one would not know this from reading applied linguistics texts) but rather in order to reconstruct parts of the canon of applied linguistics, to question the "givens" of the predominant frameworks of thought on language teaching, to explore the very particular nature of and interests served by what are taken as the key concepts in applied linguistics. This is not, however, some randomly destructive project, but rather is tied to my understanding of knowledge as always partial and interested, and my belief that the claims to truth and. knowledge frequently made by applied linguists are bound up with the *disciplining* effects of discourse. As Cherryholmes (1988) argues, the modernist and structuralist metanarratives of education make very particular claims to rationality, linearity, progress, and control.

My attempts to reconstruct the applied linguistics canon have, therefore, a number of aims: First, I am interested in exploring both

now and historically the relationship between discourses of applied linguistics and other discourses such as those of linguistics, development, modernization, international relations, and so on, thus making it possible to draw connections between applied linguistics and other discursive fields; second, I am trying to investigate the interests served by the claims to truth and knowledge made in applied linguistics; third, I am interested in the effects of the growth of applied linguistics as a *discipline* (retaining the important ambivalences of that term); fourth, I want to ask with respect to applied linguistics a question similar to Foucault's (1980) "What types of knowledge do you want to disqualify in the very instant of your demand 'Is it a science'?" (p. 85); finally, and following from this last question, I want to explore the possibilities of Foucault's phrase "the insurrection of subjugated knowledges" (p. 81) and thus to look not only at the construction of applied linguistics as a scientist discourse but also at what knowledges have been left out, forgotten, submerged, and what, therefore, a new, broader, and more open applied linguistics might look like.

This last point leads me on beyond a purely reconstructive project towards a process of reconstruction. Poststructuralist and postmodernist thought has often been accused of being obsessed with difference and fragmentation because of its tendencies to reconstruct dominant discourses, leaving us at times with no obvious way to go forward. It is this issue which McCall has taken up in her response when she asks (to paraphrase) whether we are not left only with difference and fragmentation once the metanarrative of Method has been reconstructed. This dichotomy between universalist and relativism has been a common theme in structuralist and modernist thought, with many academics now dismayed as the cherished canon of universalist thought has started to crumble and apparent relativism has gained sway (literature is perhaps the best example here, but sociology, anthropology, and philosophy have also been going through major upheavals). Combined with this nervousness about the apparent relativism and fragmentation, there has also recently been an outcry from the more conservative domains that a new "political correctness" is emerging on campus. Conservative critics are alarmed that not only are forms of knowledge being challenged but that they are being replaced by new hegemonic forms that are based on political orientations especially around issues of race and gender.[1]

1 The view that "freedom of speech" is being threatened by the emergence of new, politicized voices around universities has spilled over into a massive, conservative discourse in the popular press. See, for example, "Correcting America" (1991), D'Souza (1991), Henry (1991), "Taking Offense" (1990), "A War of Words" (1991) ,& Woodward (1991).

With respect to this last issue, I think it is quite easy to see that although some challenges to the dominant epistemologies are indeed producing their own counter dogmas, by and large the reaction to the politicization of knowledge is a reaction from the Right against the growing claims of women, minorities, teachers, and so on, to having their voices heard. The question of universalist and relativism is a little more difficult, but I think it also becomes clear that to argue that without a universal or foundational theory we are left only with complete relativism is to remain bound within a circular argument. Richard Bernstein (1983) suggests that the opposition between a conviction that there must be some permanent, ahistorical framework to which we can appeal in determining the nature of rationality, truth, knowledge, reality, goodness, or rightness and a conviction that ultimately all such concepts must be understood as relative to a specific conceptual scheme, historical moment, or cultural context is the "central cultural opposition of our time" (p. 7). But he also suggests a way beyond this cultural opposition through hermeneutics and praxis. Laclau (1988) is also interesting here in his discussion of postmodernity, which can be conceived as

> the achievement of multiple awareness: epistemological awareness, insofar as scientific progress appears as a succession of paradigms whose transformation and replacement is not grounded in any algorithmic certainty; ethical awareness, insofar as the defense and assertion of values is grounded on argumentative movements. . . . which do not lead back to any absolute foundation; political movements, insofar as historical achievements appear as the product of hegemonic and contingent—and as such, always reversible—articulations and not as the result of immanent laws of history. (p. 21)

I think, therefore, that to argue that without some foundationalist discourse we are left only with relativism, is to argue from within a closed, structuralist paradigm. Moving outside this, we can start to look at how projects of reconstruction can be formed that take ethical and political questions rather than epistemological absolutes as the way forward. There are indeed reconstructive projects afoot that suggest ways of going beyond the laissez-faire relativism that McCall feels may result from a reconstruction of Method. In a response to another article of mine (Pennycook, 1990), Kanpol (1990) has suggested that my own "principled postmodernism" (p. 10) still leaves us only with a celebration of difference, and that we need in its place a "theory of similarity within difference" (p. 239). Theorizing more carefully about how we can reconstruct notions of similarity within a postmodern emphasis on difference may be one way forward. (See my response to Kanpol [Pennycook,

in press], however, for a defense of principled postmodernism and some doubts about Kanpol's ideas.) Henry Giroux (e.g., 1991) argues that we need to retain parts of the modernist concepts of agency, justice, liberty, and critical reason in order to construct a critical pedagogy to educate a critical citizenry capable of participating in a radical democracy. Moving out of the more formal domains of politics, Roger Simon (in press) has been looking at ways of reconstituting ethical practice once the metanarratives of morality have been reconstructed, a project which shares some common ground with Sharon Welch's (1985) "feminist theology of liberation." Feminist thinkers have probably taken these questions furthest as they have sought to maintain and establish a sense of solidarity once an essentialized notion of "woman" has been reconstructed. Fraser and Nicholson (1990), for example, argue that alliances can still be formed around culturally and historically specific struggles rather than around pregiven unities or assumptions of universally shared interests or identities.

What all of these thinkers have in common is both an acceptance of the challenges posed by postmodern and poststructuralist thinking and also a realization of the need to work towards reconstructive projects based around political and ethical concerns. To return to McCall's concern, then, I would say that I do not in the least want to legitimate deadening or authoritarian teaching practices but neither do I want to judge teaching practices relative to a concept of Method. As McCall discusses, the dominance of one set of ideas embodied in the concept of Method makes it very hard to discuss or use alternatives that are not currently in vogue. The concept of Method tends to discount other possibilities and becomes a dominant criterion against which teaching practices and student performances are judged. Another possible way out of this problem might appear to be by asking whether something "works," but such an appeal to functionalism and pragmatism also needs to be assessed for its assumptions about efficiency and the criteria by which one judges whether something works or not. (Thus I would question McCall's assertion that Krashen & Terrell, 1983, "favor these methods because they are effective, not because they support a particular position about knowledge and power.") What I am suggesting is that we can nevertheless make decisions about teaching according to local cultural, political, and pedagogical concerns. We can decide that some teaching practice is good not by asking, Is it communicative? or Is it O.K. to teach grammar again now? but by asking what it is we wish to accomplish with our particular students in our particular contexts. We can criticize an authoritarian teaching approach (though with caution, for what

seems authoritarian to some may not to others, and because the distinction between authoritarian and authoritative has perhaps been too often conflated) by dint of the social and political rights of the students. We as teachers can ask questions and make decisions about our teaching not according to the metanarratives of applied linguistics but according to the ethical and political projects that inform our daily lives.

A final point. McCall ends her response by pointing to an irony that I, as an "expert," am promoting distrust of experts. First, I would like to question this ascription of the title of expert since it is certainly not one I would choose myself. Perhaps, however, this status is automatically conferred by the power and prestige of the *TESOL Quarterly,* though I would hope that those of us who have had the opportunity to write here claim to do so less as experts than as people with something to say. Second, and more important, if I accept for a moment this expert status, then there is indeed a contradiction in my argument to distrust experts. But I think it is by accepting the ambivalence of such positions rather than trying to remove contradiction and doubt that we can develop a more generous understanding of language and language teaching. So I thank Martha McCall for her suggestion that despite some of the ironies and contradictions in my writing, I may have encouraged others to be skeptical and to question some of the claims made to truth and knowledge in the name of applied linguistics, including my own.

■

REFERENCES

A war of words. (1991, May 27). *Maclean's,* pp. 40-50.

Bernstein, R. (1983). *Beyond objectivism and relativism: Science, hermeneutics, and praxis.* Philadelphia: The University of Pennsylvania Press.

Cherryholmes, C. (1988). *Power and criticism. Poststructuralist investigations in education.* New York: Teachers College Press.

Correcting America. (1991, July 5). *The New York Times Higher Education Supplement;* p. *36.*

D'Souza, D. (1991, March). Illiberal education, *The Atlantic,* pp. 51-79.

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings (C.* Gordon, Ed.). New York: Pantheon.

Fraser, N., & Nicholson, L. (1990). Social criticism without philosophy: An encounter between feminism and postmodernism. In L. Nicholson (Ed.), *Feminism/postmodernism.* New York: Routledge.

Giroux, H. A. (1991). Rethinking the boundaries of educational discourse: Modernism, postmodernism, and feminism. *College Literature, 17* (2/3), 1-50.
Henry, W.. A. III. (1991, April 14). Upside down in the groves of academe. *Time,* pp. 62-64.
Kanpol, B. (1990). Political applied linguistics and postmodernism: Towards an engagement of similarity within difference. A reply to Pennycook. *Issues in Applied Linguistics, 1* (2), 238-250.
Krashen, S. D., & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom.* Oxford: Pergamon Press.
Laclau, E. (1988). Building a New Left: An interview with Ernesto Laclau. *Strategies, l* (l), 10-28.
Pennycook, A. (1990). Towards a critical applied linguistics for the 1990s. *Issues in Applied Linguistics, 1* (1), 8-28.
Pennycook, A. (in press). A reply to Kanpol. *Issues in Applied Linguistics.*
Simon, R. (in press). *Teaching against the grain: Essays for a pedagogy of possibility.* Westport, CT: Bergin & Garvey.
Taking offense. (1990, December 24). *Newsweek,* pp. 48-55.
Welch, S. (1985). *Communities of resistance and solidarity:* A *feminist theology of liberation.* Maryknoll, NY: Orbis.
Woodward, C. V. (1991, July 18). Freedom and the universities [Review of *Illiberal education: The politics of race and sex on campus]. The New York Review of Books,* pp. 32-37.

Comments on Sandra Lee McKay and Sarah Warshauer Freedman's "Language Minority Education in Great Britain: A Challenge to Current U.S. Policy"

A Reader Reacts. . .

MARJORIE HOYE
Central Oregon Community College

In their recent article in the *TESOL Quarterly* (Vol. 24, No. 3), McKay and Freedman "suggest that, as language teachers we state our goals and value standpoints on language minority programs clearly before we make any recommendations regarding particular programs for LM/LEP [language minority/limited English proficient] students" (p. 401). They list three groupings of questions whose answers would ostensibly aid language teachers in

determining whether they advocate "mainstream programs" as found in Britain, or "separate language programs''—like "ESL pull-out programs, sheltered English, or bilingual education" (p. 386)—as found in the U.S. Generally, the three groupings of questions ask teachers to clarify their assumptions and values regarding (a) social segregation in educational programs, (b) language learning, and (c) their role as language teachers. I have two major concerns with this article. One is the authors' suggestion that there are only two approaches to teaching LEP students: separate programs and mainstreaming with language support. Second is the authors' failure to raise a key issue: their questions do not ask teachers to clarify their views on academic achievement in a second language.

I will address the first concern briefly: that there exist only two approaches to teaching LM/LEP students. McKay and Freedman's view of separate programs, in which they not only conflate ESL pull-out programs, sheltered English, and bilingual education, but also all bilingual education programs, is simplistic. Such categorizing suggests, for example, that minority bilingual education programs are identical to majority bilingual programs (for a discussion of this distinction, see Williams & Snipper, 1990, pp. 51-62); or that concurrent translation programs, in which content is explained in both the native and the second languages, are equivalent to what Krashen (1981) calls ideal bilingual education, in which the subject matter is taught only in the primary language. In the United States bilingual education programs "vary along almost every conceivable dimension, as do the students they serve" (McLaughlin, 1985, p. 245), so that "bilingual education programs have proven extremely difficult to evaluate" (p. 245). This variety of bilingual programs supports McKay and Freedman's statement that both school districts in the United States and their counterparts in Britain, Local Education Authorities, enjoy "considerable autonomy in policy formulation and curriculum development" (p. 387). To suggest that minority education programs exist only as theoretically conceived at a national level is a misleading view. Before making recommendations about minority programs, language teachers need to know about programs that exist at local levels that have been designed and implemented to address very specific local problems, programs that combine mainstreaming and separation.

In presenting my second concern, I assume that all teachers share at least one goal: the academic success of their LM/LEP students. Therefore, I am disconcerted that the authors-do not ask teachers to formulate their views on academic achievement in a second language. For teachers will not be able to intelligently recommend a particular LM/LEP program unless they understand the complex

relationships among a number of variables that affect academic achievement in a second language.

McKay and Freedman's implicit assumption is that second language development is the key to language minority education and that mainstream programs that offer language support are the most effective programs in promoting language learning. They base their argument on the following syllogism. Major premise: social equality promotes language learning; minor premise: racial integration or mainstreaming will achieve social equality; conclusion: LM/LEP students should be mainstreamed. Although with Hamers and Blanc (1989) I seriously question that mainstreaming will achieve social equality, that it will ensure the valuing of a minority language in the larger community, I don't deny that language proficiency is an essential component in academic success and that social factors are important variables in language learning. But there are other variables as well that affect academic success in a second language. To fully understand the challenges that LM/LEP students face, language teachers need to understand the relationships among multiple variables that predict academic success. Again, the authors present a simplistic view to a very complex issue.

In a recent article entitled "How Long? A Synthesis of Research on Academic Achievement in a Second Language," Collier (1989) discusses the relationship among variables that research has shown affect academic success in a second language. She provides nine generalizations on optimal age of second language acquisition, cognitive development in the first language, and academic achievement in a second language and discusses their implications for specific learning situations. For example, she states that adolescents who arrive in the United States with no second language background and who are unable to continue schooling in their first language while acquiring a second language "will not have enough time left in high school to make up the lost years of academic instruction" (p. 527). This applies to students with good academic backgrounds in their native language and to students whose schooling has been "limited or interrupted" (p. 527). Without special assistance, they "may never reach the 50th NCE [Normal Curve Equivalent, read percentile]," and may drop out of school (p. 527). Collier implies that younger students have different academic needs. Those who arrive in the United States before the age of 12, with at least 2 years of schooling in their native language, will take 5–7 years to "reach the level of average performance by native speakers on L2 standardized tests in reading, social studies, and science" (p. 257) when schooled exclusively in the second language. In her final generalization, Collier (1989) reports that

"consistent, uninterrupted cognitive academic development in all subjects throughout students' schooling is more important than the number of hours of L2 instruction for successful academic achievement in a second language" (p. 527). While Collier's synthesis of research suggests that many LM/LEP students need to develop literacy skills in their native language, McKay and Freedman neglect this consideration by focusing on second (or other) language competency as a means to social equality.

From Collier's article, the following questions can be generated to aid language teachers in clarifying their views on academic achievement in a second language so that they can make intelligent recommendations about LM/LEP programs. What is the relationship between academic development in the native language and academic development in the second language? What is the relationship between language development and academic development? What is the relationship between the number of years of instruction in the second language and successful academic achievement in the second language? Once these questions are answered, as well as those posed by McKay and Freedman, language teachers will probably advocate neither a pure mainstream or a pure separate program, but

> a program that will support immigrant students' learning for the entire day, separating them only for the programs that are absolutely essential to meet unique learning needs difficult to handle within a mainstream setting, such as bilingual upgrading programs and programs providing group identity and support,. (Handscombe, 1989, p. 31)

REFERENCES

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23* (3), 509-531.

Hamers, J. F., & Blanc, M. H. A. (1989). *Bilinguality and bilingualism*. Cambridge: Cambridge University Press.

Handscombe, J. (1989). Mainstreaming: Who needs it? In J. H. Esling (Ed.), *Multicultural education and policy: ESL in the 1990s* (pp. 18-35). Toronto: OISE Press.

Krashen, S. (1981). Bilingual education and second language acquisition theory. In Evaluation, Dissemination, and Assessment Center (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 51-79). Los Angeles, CA: California State University, Los Angeles.

McLaughlin, B. (1985). *Second-language acquisition in childhood: Vol. 2. School-age children* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Williams, J. D., & Snipper, G. C. (1990). *Literacy and bilingualism.* White Plains, NY: Longman.

Response to Hoye. . .

Decision Making for Minority Education: Setting Priorities

SANDRA LEE McKAY
San Francisco State University

SARAH WARSHAUER FREEDMAN
University of California, Berkeley

In her commentary, Marjorie Hoye raises two major concerns with our article, "Minority Education in Great Britain: A Challenge to Current U.S. Policy''—that we suggest "that there are only two approaches to teaching LM/LEP students" and that we fail "to raise a key issue" by not asking "teachers to clarify their views on academic achievement in a second language." We would like to consider each of Hoye's concerns in turn.

In the introduction to our article, we point out that "U.S. policies promote separate educational programs such as ESL pull-out programs, sheltered English, or bilingual education" (386). We grouped these programs because they all potentially separate LM/LEP students. Hoye suggests that we admit to only "two approaches," that we "conflate" a number of programs, and that we "categorize" programs. Given her use of terms, we're not sure exactly where her disagreement lies. We have recognized that there are numerous approaches, like sheltered English, ESL pull-out, mainstreaming, and bilingual education. We agree with Hoye that we have categorized these approaches along two dimensions, but not that we in any way have conflated them. We created our two categories for the purpose of examining one key variable in LM/LEP programs, namely, the racial composition of classrooms. We did this in order to encourage a discussion of the social and linguistic effects of this dimension of classroom life on student learning. As suggested in our conclusion, we as a profession need to ask the question, "What are our views on social segregation in educational programs?" (p. 402). This is a difficult question, one that we did not attempt to answer, but rather one that we felt needed to be raised.

As for Hoye's second concern, we agree that the academic achievement of nonnative speakers must be a central consideration in program design. In fact, in our article we acknowledge the need for attention to academic achievement. Our third and final question focuses directly on content learning, suggesting our shared concern with Hoye on this matter.

Hoye contends that we are arguing for mainstreaming. We are not. We are rather raising three central questions for the profession to address, consider, and debate. No syllogisms of the sort Hoye puts forth underlie our article. Although social equality may well contribute to language learning, we are not so naive as to think that mainstreaming, in and of itself, will result in social equality. Social equality will be achieved only when all individuals share equal political and economic access. The school, along with other social institutions, plays a role in either supporting or undermining social equality. When social equality is attained, individuals and social groups will be able to select the language programs that best meet their needs.

We envision our article being used in the following manner. A community would set the priorities they believe important based on the questions we raised and others. If a community were to decide that its priorities included native language development, academic achievement, and social integration, then program development would proceed with these priorities in mind. Setting priorities would encourage planners to design innovative programs responsive to community needs, rather than automatically selecting a preexisting model. Our article advocated no approach. We argued rather for a careful weighing of priorities before designing or accepting any approach to minority education.

Research Issues

The *TESOL Quarterly* publishes brief commentaries on aspects of qualitative and quantitative research. For this issue, we asked two researchers to address the following question: What is the importance of power and effect size for second language research?

Edited by GRAHAM CROOKES
University of Hawaii at Manoa

Power, Effect Size, and Second Language Research

A Researcher Comments. . .

ANNE LAZARATON
The Pennsylvania State University

When we engage in testing hypotheses in our research, we hope we will be able to reject our null hypotheses (e.g., that there is no difference

between two groups) and accept our research hypotheses (e.g., that there is a difference). This decision is based on a test of significance, where some observed statistic is compared to a critical value at a prespecified level of probability. Because of the nature of significance testing, however, some of these decisions to reject the null hypothesis will be incorrect. We may incorrectly reject a true null hypothesis (H_0) —claim that an effect or relationship exists when it does not—and commit a Type I error. The likelihood of this is equal to our prespecified level of probability, alpha (typically .05). A Type II error, beta, occurs when we claim there is no effect or relationship when there is, and retain the null hypothesis when it should be rejected. Alpha and beta are inversely related: When we choose a more conservative alpha value, such as .001, to minimize the likelihood of a Type I error, at the same time we increase the chance of committing a Type II error by retaining a false H_0. The reverse is also true. As researchers, we strive for the optimal balance between the possibility of committing either type of error. The statistical test which best achieves this balance is the most powerful test we can choose.

Power, the ability of a statistical test to detect a false null hypothesis, is therefore highly desirable, because by minimizing the likelihood of making an error in evaluating a null hypothesis, we increase confidence in our findings. How do we know how powerful a test is? While it is possible to estimate power in a particular study (Cohen, 1988; Shavelson, 1988), for those of us who engage in small-scale research or who are consumers of research, it is more realistic to be aware of factors which affect power and ways we can increase the power of a statistical test. Power is a function of four factors: significance level, variability within the population, sample size, and effect size.

Conventionally, the significance level for testing the null hypothesis is often .05. But in attempting to control for Type I error by selecting a low alpha level, we increase the likelihood of Type II error. We might, therefore, consider raising the .05 level to .10 to ensure that we will be able to detect a false null hypothesis. We should also think about the trade-off between Type I and II errors, and their consequences. It may be that the conservative .01 level is appropriate if a serious decision rests on the outcome; however, in a pilot study of new materials, .10 may be reasonable. As researchers, we must choose significance level carefully. It should be selected before the study is conducted, and results reported relevant to it. This highlights the common misconception that .01 or .001 values (and the double and triple asterisks that often accompany them) indicate importance: Lower values mean only that we can be more certain that effects or relationships exist, not that they are important. For example, a statistically detectable difference of only a fraction of a point between two groups may have no practical implications for teaching.

A second factor affecting power is the variability of the population from which the sample is drawn and to which conclusions will be generalized. The less variability in the sample, the more powerful our test, because any true effect or relationship will be more easily detected when it is less obscured by random differences. One way that the effect of variability y

can be decreased is to increase sample size, a third facet of power. As more observations are made, the less variability there is in our statistical summaries because these now reflect a larger body of combined information. Also, as more observations are made, the closer our sample distribution will approach the (assumed) normal distribution of the population itself. (That it is normal is an assumption which underlies many powerful statistical procedures.) The larger the sample, the more powerful our test, and it goes without saying that we should use a large sample whenever possible. If true differences exist, they are more likely to be detected in a sample of 100, rather than one of 20. It is true, though, that we often work with data from "general learners," which tend to represent a heterogeneous population. This often leads to large within-group variance, and thus to nonsignificant differences. Consequently, even larger sample sizes are needed in these cases.

Finally, effect size is a critical component of power, although it is one of the least familiar concepts of statistical inference (Cohen, 1988, p. 10). Effect size refers to the magnitude of the difference between the population means under the null hypothesis (H_0) and the research (H_1) hypothesis. The hypothesized difference in means must be expressed exactly in the research hypothesis, not just as greater or less than the mean under the null hypothesis. The magnitude of the difference is often expressed as a z score, a "small" difference being .2 of a z score, a "medium" difference .5, and a "large" difference, .8 (Cohen, 1969, cited in Shavelson, 1988, p. 295). Unfortunately, even in well-designed experimental studies, an exact alternative hypothesis—necessary for the calculation of effect size—cannot be stated (Henkel, 1976). Though we may be unable to specify effect size in advance, we can estimate and report it after the fact, with an appropriate strength of association measure. Eta-squared, omega-squared and phi give an indication of the importance of obtained results in terms of strength of treatment effect or relationship. These measures tell us how much variability in the dependent variable can be accounted for by the independent variable, or how much information variables share in a given sample. While the measures cannot speak to strength of association in the population, they provide vital information about a study that is unavailable when just alpha levels are reported (see Hatch & Lazaraton, 1991, for more on these measures).

To summarize, power is a function of significance level, variability in the population, sample size, and effect size. Decisions about significance level and sample size, like those about hypothesis formulation, data collection, and data analysis, cannot be avoided by the researcher. It is tempting to "let the computer (or the consultant) decide," but ultimately the responsibility for our work rests with each of us, and the integrity with which we carry out this work *is* judged by our fellow applied linguists if not the larger educational community.

ACKNOWLEDGMENTS

I wish to thank Patricia Dunkel and Evelyn Hatch for insightful comments on this article.

THE AUTHOR

Anne Lazaraton is Assistant Professor of Speech Communication at The Pennsylvania State University. She publishes and does research in conversation analysis, oral proficiency testing, and research methodology in applied linguistics.

REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* New York: Newbury House.

Henkel, R. E. (1976). *Tests of significance.* (Quantitative Applications in the Social Sciences Series.) Beverly Hills, CA: Sage.

Shavelson, R. J. (1988). Statistical reasoning for the behavioral sciences (2nd ed.). Boston, MA: Allyn & Bacon.

Another Researcher Comments. . .

GRAHAM CROOKES
University of Hawaii at Manoa

The most obvious propositions which come to mind when discussing power and its companion, effect size, in a second language (SL) research context are: (a) the use or provision of power and effect size estimates is extremely desirable, and (b) they are almost never used. This may lead us to asking why there is this disuse, and what can be done about it.

Because power and effect size are unusual in SL research (the former was ranked 22 out of 23 in terms of researchers' "self-knowledge" [Lazaraton, Riggenbach, & Ediger, 1987]), I begin with a simple (and partial) definition. In a simple two-group experiment, effect size is the extent to which the mean of the experimental group differs from that of the control group, standardized in terms of the standard deviation of the two groups combined. (Several other measures exist—Murray & Dosser, 1987; Rosenthal & Rosnow, 1984—and the concept is applicable outside an experimental context, but consider this, for simplicity.)

Why is the use of this statistic desirable? First, it provides a measure which indicates whether the result of an experimental treatment was substantial or not, regardless of sample size. Second, if the likely effect size of an investigation can be estimated before the study is undertaken (via a pilot or previous inquiry) it can be used, in a "power study," to ascertain how many subjects will be needed to have a specific probability of rejecting the null hypothesis. In other words, an existing effect size estimate can tell the investigator how large a sample will be needed to obtain statistical significance at a given level of power.

Power studies are essential if we are to keep to a minimum the two main errors that can be made in statistical inference testing (Types I and II;

cf. P. Cohen, 1983): rejecting the null hypothesis when in the whole population a difference does not exist; and accepting the null hypothesis on the basis of the results obtained from the groups sampled when in the whole population a difference does exist. Most of the time, investigators use statistical testing conditions which are quite unlikely to register (by showing a statistically significant difference) an effect which actually exists and have no idea of their chances of obtaining statistically significant differences. Nevertheless, there is a persistent tendency to rely on the conventions (a) "30 subjects is sufficient," and (b) control of Type I error at 1 in 20, with no control of Type II error.

Why, one may ask, is this the case, whose fault is it, and why has nothing been done? One reason for this situation is the institutional inertia in the educational research community as a whole, and particularly in the SL research field (cf. Lazaraton et al., 1987), where until recently many senior investigators had training in linguistics rather than educational research. The situation is exacerbated by the exclusion of recent developments in statistics from statistical research texts, which present statistical inference as subject neither to debate nor change (cf. Gigerenzer & Murray, 1987; Oakes, 1986). This is indicated in the lack of reference to original works in many such texts (exemplary exceptions: Glass & Hopkins, 1984; Keppel, 1991). The one person whose fault it is not is Jacob Cohen, who with various associates has been providing ringing calls to redress this situation for several decades now, on whose work I am drawing extensively here (e.g., Cohen, 1962, 1965, 1977, 1990).

There is no good reason why nothing has been done, but the following weak excuses may be noted. (a) The SL research community cannot unaided overcome the neglect of the topic within standard statistical courses and texts. (b) Until recently, power analyses required the use of fairly complex formulae and tables (J. Cohen, 1977; Kraemer & Thiemann, 1987; but see Lipset, 1990). (c) Since so much SL research is groundbreaking, there are almost never the preliminary estimates of effect size necessary for power analyses. (d) SL research is labor-intensive in nature—no undergraduate classes of Psychology 101 with their required participation for grade credit and 200-strong *n* sizes are available to us! (e) Power and effect size programs are not to be found in the standard mainframe statistical packages.

Much of this constitutes the unfortunate history of this topic, but little can serve as justification. A simple computer program (Borenstein & Cohen, 1988; cf. Borenstein, Cohen, Rothstein, Pollack, & Kane, 1990) can handle the necessary calculations. There is a growing literature which documents the weaknesses of underpowered studies (e.g., Frieman, Chalmers, Smith, & Kuebler, 1978; Lynch, 1987; cf. Cohen & Hyman, 1979; Sedlmeier & Gigerenzer, 1989). SL researchers are increasingly able to handle original sources in quantitative methods. We now recognize that many published SL studies are no more than pilot studies, which would have been greatly improved if seen as such and followed up by doing the actual study, properly and with a decent *n* size. But in addition, in some

areas, accumulations of studies do provide the possibility of preliminary work which can be used in power studies, so that more definitive results can be obtained. Having said this, I would have to acknowledge from my experience that the labor-intensive demands of applied linguistics research still carry practical, if not logical, weight.

Nevertheless, it is essential that those of us doing quantitative work which involves the use of statistical inference commit ourselves to not publishing pieces unless they can make a substantive contribution to the field, as opposed to our resumes—we must resist the temptation of the MPU (minimal publishable unit). Just as journal editors and article readers have conventionalized (for good or bad) the .05 alpha level, it will be up to them to seek more substantive indications from those who wish to be published in the future that the results they provide are not just a chance effect of the law of small numbers. Power studies and effect size measures can help to provide this substantiation.

THE AUTHOR

Graham Crookes is Assistant Professor, Department of ESL, University of Hawaii at Manoa.

REFERENCES

Borenstein, M., & Cohen, J. (1988). *Statistical power analysis: A computer program.* Hillsdale, NJ: Lawrence Erlbaum.

Borenstein, M., Cohen, J., Rothstein, H., Pollack, S., & Kane, J. (1990). Statistical power analysis for one-way ANOVA: A computer program. *Behavior Research Methods, Instruments and Computers, 22,* 271-282.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal Social Psychology, 65,* 145-153.

Cohen, J. (1965). Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 95-121). New York: McGraw-Hill.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45(12), 1304-1312.*

Cohen, P. (1983). To be or not to be: Control and balancing of Type I and Type II errors. *Evaluation and Program Planning, 5,* 247-253.

Cohen, S. A., & Hyman, J. S. (1979). How come so many hypotheses in educational research are supported? (A modest proposal). *Educational Researcher, 8* (12), 12-16.

Frieman, J. A., Chalmers, T. C., Smith, H., & Kuebler, R. R. (1978). The importance of beta, the Type II error and sample size in the design and interpretation of the randomized control trial. *New England Journal of Medicine, 229,* 690-694.

Gigerenzer, G., & Murray, D. J. (1987), *Cognition as intuitive statistics.* Hillsdale, NJ: Lawrence Erlbaum.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Keppel, C. (1991). *Design and analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kraemer, H. C., & Thiemann, S. (1987). *How many subiects? Statistical power analysis in research.* Newbury Park, CA: Sage.

Lazaraton, A., Riggenbach, H., & Ediger, A. (1987). Forming a discipline: Applied linguists' literacy in research methodology and statistics. *TESOL Quarterly, 21* (2), 263-277.

Lipset, M. W. (1990). *Design sensitivity: Statistical power for experimental researchers.* Newbury Park, CA: Sage.

Lynch, K. B. (1987). The size of educational effects: An analysis of programs reviewed by the joint dissemination review panel. *Educational Evaluation and Policy Analysis, 9* (l), 55-61.

Murray, L. W., & Dosser, D. A. (1987). How significant is a significant difference? Problems with the measurement of magnitude of effect. *Journal of Counseling Psychology, 34* (l), 68-72.

Oakes, M. (1986). *Statistical inference:* A *commentary for the social and behavioral sciences.* New York: Wiley.

Rosenthal, R., & Rosnow, R. L. (1984). Considerations of power. In *Essentials of behavioral research* (pp. 355-365). New York: McGraw-Hill.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105,* 309-316.

INFORMATION FOR CONTRIBUTORS

EDITORIAL POLICY

The *TESOL Quarterly*, a professional, refereed journal, encourages submission of previously unpublished articles on topics of significance to individuals concerned with the teaching of English as a second or foreign language and of standard English as a second dialect. As a publication which represents a variety of cross-disciplinary interests, both theoretical and practical, the *Quarterly* invites manuscripts on a wide range of topics, especially in the following areas:

1. psychology and sociology of language learning and teaching; issues in research and research methodology
2. curriculum design and development; instructional methods, materials, and techniques
3. testing and evaluation
4. professional preparation
5. language planning
6. professional standards

Because the *Quarterly* is committed to publishing manuscripts that contribute to bridging theory and practice in our profession, it particularly welcomes submissions drawing on relevant research (e.g., in anthropology, applied and theoretical linguistics, communication, education, English education [including reading and writing theory], psycholinguistics, psychology, first and second language acquisition, sociolinguistics, and sociology) and that address implications and applications of this research to issues in our profession. The *Quarterly* prefers that all submissions be written so that their content is accessible to a broad readership, including those individuals who may not have familiarity with the subject matter addressed.

GENERAL INFORMATION FOR AUTHORS

Submission Categories

The *TESOL Quarterly* invites submissions in five categories:

Full-length articles. Manuscripts should generally be no longer than 20 double-spaced pages. Submit **three copies** plus three copies of an informative abstract of not more than 200 words. To facilitate the blind review process, authors' names should appear only on a cover sheet, not on the title page; do not use running heads. Manuscripts should be submitted to the Editor of the *TESOL Quarterly*:

Sandra Silberstein
Department of English, GN-30
University of Washington
Seattle, WA 98195 U.S.A.

Reviews. The *TESOL Quarterly* invites reviews of textbooks, scholarly works related to the profession, tests, other instructional materials (such as computer software, videotaped materials, and other nonprint materials), and other journals concerned with issues relevant to our profession. Comparative reviews, which include a discussion of more than one publication, and review articles, which discuss materials in greater depth than in a typical review, are particularly welcome. Reviews should generally be no longer than five double-spaced pages, although comparative reviews or review articles may be somewhat longer. Until further notice, submit **two copies** of reviews to the Review Editor of the *TESOL Quarterly:*

Heidi Riggenbach
Department of English, GN-30
University of Washington
Seattle, WA 98195 U.S.A.

Book Notices. The *TESOL Quarterly* also welcomes short evaluative reviews. Book notices should provide a descriptive and evaluative summary of a recent publication (see preceding section for appropriate types of publications) and a brief discussion of the significance of the work in the context of current theory and practice in the relevant area(s) of TESOL. Submissions should range between 350 and 500 words; any submission that exceeds 500 words will be returned. Submit **two copies** of book notices to Heidi Riggenbach, Review Editor, at the address given above.

Brief Reports and Summaries. The *TESOL Quarterly* also invites short descriptions of completed work or work in progress on any aspect of theory and practice in our profession. Reports of work in the areas of curriculum and materials development, methodology, teaching, testing, teacher preparation, and administration are encouraged, as are reports of research projects of a pilot nature or that focus on topics of specialized interest. In all cases, the discussion of issues should be supported by empirical evidence, collected through either formal or informal investigation, and should be grounded with references to similar or related work. Manuscripts should *summarize* key concepts and results in a manner that will make the research accessible to our diverse readership. Although all reports and summaries submitted to this section will be considered, preference will be given to manuscripts of three to seven double-spaced pages (including references and notes). *Longer articles do not appear in this section and should be submitted to the Editor of the TESOL Quarterly for review.* Send **two copies** of reports and summaries to the Editor of the Brief Reports and Summaries section:

Gail Weinstein-Shr
Reading and Writing Program
School of Education
Furcolo Hall
University of Massachusetts
Amherst, MA 01003-0096

The Forum. The *TESOL Quarterly* welcomes comments and reactions from readers regarding specific aspects or practices of our profession. Responses to published articles and reviews are also welcome. Contributions to The Forum should generally be no longer than five double-spaced pages. Submit **two copies** to the Editor of the *TESOL Quarterly* at the address given above.

Brief discussions of **Research Issues** in qualitative and quantitative research are also published in The Forum. Although these contributions are typically solicited, readers may send topic suggestions and/or make known their availability as contributors by writing directly to the Editor of Research Issues:

Graham Crookes
Department of English as a Second Language
University of Hawaii at Manoa
1890 East-West Road
Honolulu, HI 96822 U.S.A.

Guidelines for Authors

1. All submissions to the *Quarterly* should conform to the requirements of the *Publication Manual of the American Psychological Association* (Third Edition), which can be obtained from the Order Department, American Psychological Association, 1200 Seventeenth Street, NW, Washington, DC 20036. The *Publication Manual* is also available in many libraries and bookstores. Authors are responsible for the accuracy of references and reference citations, which must be in APA format.
2. All submissions to the *TESOL Quarterly* should be accompanied by a cover letter which includes a full mailing address and both a daytime and an evening telephone number. Where available, include an electronic mail address and fax number.
3. Authors of full-length articles should include **two copies** of a very brief biographical statement (in sentence form, maximum 50 words), plus any special notations or acknowledgments that they would like to have included. Double spacing should be used throughout.
4. The *TESOL Quarterly* provides 25 free reprints of published full-length articles and 10 reprints of material published in the Reviews, Brief Reports and Summaries, and The Forum sections.
5. Manuscripts submitted to the *TESOL Quarterly* cannot be returned to authors. Authors should be sure to keep a copy for themselves.
6. It is understood that manuscripts submitted to the *TESOL Quarterly* have not been previously published and are not under consideration for publication elsewhere.
7. It is the responsibility of the author(s) of a manuscript submitted to the *TESOL Quarterly* to indicate to the Editor the existence of any work already published (or under consideration for publication elsewhere) by the author(s) that is similar in content to that of the manuscript.

8. The Editor of the *TESOL Quarterly* reserves the right to make editorial changes in any manuscript accepted for publication to enhance clarity or style. The author will be consulted only if the editing has been substantial.
9. The views expressed by contributors to the *TESOL Quarterly* do not necessarily reflect those of the Editor, the Editorial Advisory Board, or TESOL. Material published in the *Quarterly* should not be construed to have the endorsement of TESOL.

PUBLICATIONS RECEIVED

Publishers are invited to send copies of their new materials to the *TESOL Quarterly* Review Editor, Heidi Riggenbach, University of Washington, at the address listed in the Information for Contributors section. Packages should be labeled REVIEW COPIES.

TESOL Quarterly readers are invited to contribute review articles and evaluative or comparative reviews for consideration for publication in the Review or Book Notice sections of the *Quarterly*. These should be sent to the *TESOL Quarterly* Review Editor, Heidi Riggenbach, University of Washington, at the address listed in the Information for Contributors section.

TESOL gratefully acknowledges receipt of the following publications.

Allwright, D., & Bailey, K. M. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers.* New York: Cambridge University Press.

Ambert, A. N. (Ed.). (1991). *Bilingual education and English as a second language.* New York: Garland.

Bell, R. T. (1991). *Translation and translating.* Harlow, England: Longman.

Bialystok, E. (Ed.). (1991). *Language processing in bilingual children.* New York: Cambridge University Press.

Borisoff, D., & Purdy, M. (Eds.). (1991). *Listening in everyday life: A Personal and professional approach.* Lanham, MD: University Press of America.

Brechtel, M. (1992). *Bringing the whole together: An integrated whole language approach for the multilingual classroom.* San Diego, CA: Dominie Press.

Carter, R., & Long, M. N. (1991). *Teaching literature.* Harlow, England: Longman.

Cheshire, J. (Ed.). (1991). *English around the world: Sociolinguistic perspectives.* New York: Cambridge University Press.

Christie, A. (1991). *Best detective stories of Agatha Christie* (The Bridge Series). Harlow, England: Longman.

Corbett, G. (1991). *Gender.* New York: Cambridge University Press.

Dickens, C. (1991). *Great expectations* (The Bridge Series) (Abridged and simplified by L. Doss). Harlow, England: Longman.

Dickens, C. (1991). *Oliver Twist* (The Bridge Series) (Abridged and simplified by L. Doss). Harlow, England: Longman.

Dogger, B. (1992). *Advanced beginner's English reader* (Teacher's Guide). Chicago, IL: National Textbook Company.

Gallico, P. (1991). *Flowers for Mrs. Harris* (The Bridge Series) (Simplified by R. H. Durham). Harlow, England: Longman.

Gorlach, M. (1991). *Introduction to Early Modern English.* Cambridge: Cambridge University Press.

Gum, L. (1991). *How to write a play.* New York: Literacy Volunteers of America.

Hale, G. A., & Courtney, R. (1991). *Note taking and listening comprehension on the Test of English as a Foreign Language* (TOEFL Research Report No. 34). Princeton, NJ: Educational Testing Service.

Henning, G. (1991). *A study of the effects of contextualization and familiarization on responses to the TOEFL vocabulary test items.* (TOEFL Research Report No. 35). Princeton, NJ: Educational Testing Service.

Henning, G. (1991). *A study of the effects of variation of short-term memory load, reading response length, and processing hierarchy on TOEFL listening comprehension item performance* (TOEFL Research Report No. 33). Princeton, NJ: Educational Testing Service.

Hernandez, H. (1989). *Multicultural education: A teacher's guide to content and process.* Columbus, OH: Merrill.

Hess, N. (1991). *Headstarts: One hundred original pre-text activities.* Harlow, England: Longman.

Holme, R. (1991). *Talking texts: Innovative recipes for intensive reading.* Harlow, England: Longman.

Hunter, J., & McRae, J. (1991). *The importance of being Wilde* (Talking Texts Series). London: England: Penguin English.

Huxley, A. (1991). *Brave new world* (The Bridge Series) (Simplified by H. A. Cartledge). New York: Longman.

Jacobs, G. M., & Power, M. A. (1991). *Putting it all together: Readings for students of English as a second language.* Ann Arbor, MI: University of Michigan Press.

Johnson, D. M. (1992). *Approaches to research in second language learning.* New York: Longman.

Jones, L. (1992). *Communicative grammar practice: Activities for intermediate students of English.* New York: Cambridge University Press.

Kenworthy, J. (1991). *Language in action: An introduction to modern linguistics.* Harlow, England: Longman.

Kitao, K., Broderick, V., Kitao, S. K., Miyamoto, H., & Miyazaki, J. (1989). *Holidays in the United States.* Tokyo, Japan: Eichosha Shinsha.

Knowles, P. L., Bailey, R., & Jillett, R. (1991). *Functioning in business* (New ed.). Harlow, England: Longman.

Loss, B. R. (1991). *Say the word! A guide to improving word recognition skills.* Syracuse, NY: New Readers Press.

Loss, B. R. (1991). *Say the word! A guide to improving word recognition skills* (Teacher's Manual). Syracuse, NY: New Readers Press.

McCarthy, M. (1991). *Discourse analysis for language teachers.* New York: Cambridge University Press.

Moss, A., & Williams, J. (1990). *Americans at school.* Harlow, England: Longman.

Nation, M. (1991). *A dictionary of modern Britain.* London: Penguin English.

Naunton, J. (1991). *Think first certificate* (Practice Exams). Harlow, England: Longman.

Nelson, V. (1991). *Listening to listen in English* (Beginning Level). Chicago, IL: National Textbook Company.

Nelson, V. (1991). *Listening to listen in English* (Beginning Level Instructor's ed.). Chicago, IL: National Textbook Company.

O'Dell, F. (1991). *Pitman exams ESOL practice tests* (Basic). Harlow, England: Longman.

O'Dell, F. (1991). *Pitman exams ESOL practice tests* (Elementary). Harlow, England: Longman.

Oltman, P. K., & Stricker, L. J. (1991). *Developing homogeneous TOEFL scales by multidimensional scaling* (TOEFL Tech. Rep. No. 1). Princeton, NJ: Educational Testing Service.

Orwell, G. (1991). *Animal farm* (The Bridge Series). Harlow, England: Longman.

Paton, A. (1989). *Cry, the beloved country* (The Bridge Series) (Simplified by G. F. Wear & R. H. Durham). Harlow, England: Longman.

Ray, S. (Ed.). (1988). *Slowly by slowly: Teaching English as a second language in Kenya's Harambee schools.* Cheshire, CT: Biblo-Moser.

Richards, J. C., Hull, J., & Proctor, S. (1991). *Interchange: English for international communication 3* (Student's Book). New York: Cambridge University Press.

Roland, G. (1992). *Communicative American English: Building fluency & cultural understanding through functional conversations & activities.* Chicago, IL: National Textbook Company.

Schoenberg, I. E. (1989). *Talk about values: Conversation skills for intermediate students.* White Plains, NY: Longman.

Scott, W. A., & Ytreberg, L. H. (1990). *Teaching English to children.* Harlow, England: Longman.

Seliger, H. W., & Vago, R. M. (Eds.). (1991). *First language attrition.* New York: Cambridge University Press.

Shulman, M. (1991). *Selected readings in business* (English for Special Purposes Series). Ann Arbor, MI: University of Michigan Press.

Smallwood, B. A. (1991). *The literature connection: A read-aloud guide for multicultural classrooms.* Reading, MA: Addison-Wesley.

Spears, R. A. (1992). *Common American phrases in everyday contexts: A detailed guide to real-life conversation and small talk.* Chicago, IL: National Textbook Company.

Stewart, E. C., & Bennett, M. J. (1991). *American cultural patterns: A cross-cultural perspective.* Yarmouth, ME: Intercultural Press.

Thornley, G. C. (Ed.). (1990). *Stories from many lands* (The Bridge Series). Harlow, England: Longman.

Wallace, M. J. (1991). *Training foreign language teachers: A reflective approach.* New York: Cambridge University Press.

Way, W. D., & McKinley, R. L. (1991). *Development of procedures for resolving irregularities in the administration of the listening comprehension section of the TOEFL test* (TOEFL Tech. Rep. No. 3). Princeton, NJ: Educational Testing Service.

Way, W. D., & McKinley, R. L. (1991). *An investigation of the use of simplified IRT models for scaling and equating the TOEFL test* (TOEFL Tech. Rep. No. 2). Princeton, NJ: Educational Testing Service.

Wennerstrom, A. (1989). *Techniques for teachers: A guide for nonnative speakers of English.* Ann Arbor, MI: University of Michigan Press.

White, R., Martin, M., Stimson, M., & Hodge, R. (1991). *Management in English language teaching.* Cambridge: Cambridge University Press.

Wigfield, J. (1988). *First steps in reading and writing* (2nd ed.). New York: Newbury House.

Wigfield, J. (1988). *Second steps in reading and writing* (2nd ed.). New York: Newbury House.

Williams, J. D., & Snipper, G. C. (1990). *Literacy and bilingualism.* White Plains, NY: Longman.

PUBLICATIONS AVAILABLE
FROM THE TESOL CENTRAL OFFICE

Publications on a wide range of topics relevant to TESOL professionals are available from the TESOL Central Office. Many of these publications are offered at reduced rates to TESOL members. An annotated list of publications can be obtained by contacting TESOL Publications Department, Suite 300, 1600 Cameron Street, Alexandria, Virginia 22314 U.S.A. A TESOL Order Form is printed in every issue of the *TESOL Quarterly* and can be used to order any of these materials. *Please note that prices inside parentheses are special discount prices for TESOL members only.* Prices are subject to change without notice.

REFERENCE GUIDES

Directory of Professional Preparation Programs in TESOL in the United States, 1989-1991. Describes 245 college and university programs in TESOL leading to doctorate, master's, or bachelor's degrees, and certificates. Overview of state certification requirements. 1989. $20.00 ($15.00). ISBN 0-939791-35-8

1987 TESOL Membership Directory. A comprehensive resource of who's who in TESOL. Both alphabetical and geographical listings of more than 11,000 institutional, commercial, and individual members. $5.00.

A World of Books: An Annotated Reading List for ESL/EFL Students. Dorothy S. Brown. More than 75 entries, including titles on various subjects and levels to assist the intermediate through advanced student in selecting reading materials. 1987. $7.50 ($6.00). ISBN 0-939791-32-3

Selected Articles From the TESOL Newsletter: 1966-1983. John F. Haskell, editor. More than 100 articles on methodology, language and culture, linguistics and grammar, language assessment, composition, ESP, reading, and vocabulary. 1986. $16.50 ($15.00). ISBN 0-939791-29-3

Annotated Bibliography of ESL Materials. Christine Aronis. Gathered from 76 publishers, the 636 ESL titles listed were designed to teach American English. Provides ESL teachers with an accurate and comprehensive list of titles available. 1983. Sale $5.00. ISBN 0-318-18069-3

Bilingual, ESOL and Foreign Language Teacher Preparation: Models, Practices, Issues. John F. Fanselow and Richard L. Light, editors. An essential reader for professional preparation. 1978. $10.00 ($8.00). ISBN 0-939791-03-X

THEMATICS

Coherence in Writing: Research and Pedagogical Perspectives. Ulla Connor and Ann M. Johns, editors. Interpretations of coherence for the ESL/EFL teacher, teacher trainer, and researcher in L1 and L2. The processes of making and understanding coherence and the manifestations of coherence in written products. Discussion questions and activities. 1990. $14.95 ($10.95). ISBN 0-939791-34-X

Students and Teachers Writing Together: Perspectives on Journal Writing. Joy Kreeft Peyton, editor. Examines journal writing as a powerful stimulus for curriculum and teacher change; as a forum for discussing the meaning and form of literary texts and the journal texts themselves; and as a bridge to other kinds of academic writing. Applications for ESL, bilingual education, composition, writing across the curriculum, writing research, literacy development, and deaf education. 1990. $11.95 ($9.95). ISBN 0-939791-36-6

Research in Reading in English as a Second Language. Joanne Devine, Patricia L. Carrell, and David E. Eskey, editors. Views of reading as an interactive process involving reader and text variables, of the interplay of reading skills and general language proficiency, and of the differences in first and second language reading. 1987. $16.50 ($15.00). ISBN 0-939791-30-7

Children and ESL: Integrating Perspectives. Pat Rigg and D. Scott Enright, editors. ESL teachers as language advocates, ESL children's writing, reading in ESL, and more. 1986. $12.50 ($10.00). ISBN 0-939791-24-2

Current Perspectives on Pronunciation: Practices Anchored in Theory. Joan Morley, editor. The pronunciation component of oral language: how it operates to transmit meaning, how it can be learned, how teachers facilitate learning. 1987. $14.00 ($12.00). ISBN 0-939791-28-5

Ending Remediation: Linking ESL and Content in Higher Education. Sarah Benesch, editor. How higher education is bridging the gap between ESL and content courses through collaboration across the curriculum. Eight faculty members describe innovative linkage programs. 1988. $12.00 ($10.00). ISBN 0-939791-33-1

TESTING

Reviews of English Language Proficiency Tests. J. Charles Alderson, Karl J. Krahnke, and Charles W. Stansfield, editors. Descriptive and evaluative information on the major ESL/EFL tests used worldwide, including more than 40 reviews. 1987. $16.50 ($15.00). ISBN 0-939791-31-5

Technology and Language Testing. Charles W. Stansfield, editor. New developments in measurement theory and test administration methodology. Papers from the 1985 Language Testing Research Colloquium. $12.50 ($9.00). ISBN 0-939791-08-0

CLASSROOM PRACTICES

Classroom Practices in Adult ESL. Donna Ilyin and Thomas Tragardh, editors. Articles on topics such as classroom organization and management, evaluation, materials, cultural considerations, literacy and reading, communicative competence, and more. 1978. $8.00 ($5.00).

Classroom Practices in ESL and Bilingual Education. Muriel Saville-Troike, editor. A collection of articles designed to acquaint classroom teachers with successful methods and materials used in the ESOL and bilingual education contexts. 1979. $5.00 ($4.00).

ON TESOL SERIES

Selected papers from TESOL Annual Conventions contained in the following volumes:

On TESOL '84. Penny Larson, Elliot L. Judd, and Dorothy S. Messerschmitt, editors. 1984 in Houston.

On TESOL '83. Jean Handscombe, Richard A. Orem, and Barry P. Taylor, editors. 1983 in Toronto.

On TESOL '82. Mark A. Clarke and Jean Handscombe, editors. 1982 in Honolulu.

On TESOL '81. Mary E. Hines and William Rutherford, editors. 1981 in Detroit.

On TESOL '76. John F. Fanselow and Ruth H. Crymes, editors. 1976 in New York City.

On TESOL '74. Ruth Crymes and William E. Norris, editors. 1974 in Denver.

Individual volumes available for $9.95 ($7.95).

Complete set of all 6 volumes of the *On TESOL* Series — $45.95 ($35.95).

TESOL PAPERS

Statement of Core Standards for Language and Professional Preparation Programs. 1985.

Guidelines for the Certification and Preparation of Teachers of English to Speakers of Other Languages in the United States. 1976.

All papers: single copies, $1.00 each; multiple copies in groups of 10 for $5.00.

New!
Resolutions Adopted by the TESOL Membership at Legislative Assemblies, 1971 to Present. 8½″ x 11″, 3-hole punched with covers, shrink-wrapped. $4.00 ($2.50).

TESOL NEWSLETTER SUPPLEMENTS

TESOL Newsletter: 21st Anniversary Issue. More than 20 articles, including an ESOL bare-bones bibliography, state of the art, early history of TESOL, facts and faces in TESOL, and more. April, 1987. $3.00.

Computer-Assisted Language Learning. 14 articles covering various aspects of CALL. February, 1986. $2.50.

Writing and Composition. 9 articles on ESL writing and composition, plus tips for writing teachers and a brief list of books and journals. February, 1984. $2.50.

TESOL CLASSICS ON SALE

The Acquisition and Use of Spanish and English as First and Second Languages. Roger W. Andersen, editor. Colloquium papers from TESOL 1978 in Mexico City. Reg. $6.50 **Sale $3.00.**

The Construct Validation of Tests of Communicative Competence. Adrian S. Palmer, Peter J. M. Groot, and George A. Trosper, editors. Papers from the 1979 Testing Colloquium in Boston. Reg. $6.50 **Sale $3.00.**

The Human Factors in ESL. James E. Alatis and Ruth Crymes, editors. Collection of papers by ESOL experts: Finocchiaro on the teacher, Paulston on communicative competence, Tucker on innovative approaches, Burt on error analysis, and more. 1977. Reg. $5.00 **Sale $3.00.**

A Handbook of Bilingual Education (Rev. Ed.). Muriel R. Saville and Rudolph C. Troike. Specific suggestions for design, instruction, and evaluation of bilingual programs. 1978. **Sale $3.00.**

NEW!

Dialogue Journal Writing with Nonnative English Speakers: A Handbook for Teachers, Joy Kreeft Peyton and Leslee Reed. This comprehensive handbook offers practical advice in an easy-to-use format free of jargon. It is recommended for teachers of nonnative English-speaking students in mainstream, bilingual, or ESL programs, from kindergarten through high school. It also has direct application to native English-speaking, gifted and talented, learning disabled, and special education students. 1990. $9.95 ($6.95). ISBN 0-939791-37-4