## CONTENTS

### ARTICLES

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension Theory and Second Language Pedagogy</td>
<td>9</td>
</tr>
<tr>
<td>Stephen J. Nagle and Sara L. Sanders</td>
<td></td>
</tr>
<tr>
<td>Computer-Assisted Language Learning as a Predictor of Success in Acquiring English as a Second Language</td>
<td>27</td>
</tr>
<tr>
<td>Carol Chapelle and Joan Jamieson</td>
<td></td>
</tr>
<tr>
<td>The Effects of Referential Questions on ESL Classroom Discourse</td>
<td>47</td>
</tr>
<tr>
<td>Cynthia A. Brock</td>
<td></td>
</tr>
<tr>
<td>Student Perceptions of Academic Language Study</td>
<td>61</td>
</tr>
<tr>
<td>Mary Ann Christison and Karl J. Krahnke</td>
<td></td>
</tr>
<tr>
<td>Salience of Feedback on Error and Its Effect on EFL Writing Quality</td>
<td>83</td>
</tr>
<tr>
<td>Thomas Robb, Steven Ross, and Ian Shortreed</td>
<td></td>
</tr>
<tr>
<td>Interrelationships Among Three Tests of Language Proficiency: Standardized ESL, Cloze, and Writing</td>
<td>97</td>
</tr>
<tr>
<td>Edith Hanania and May Shikhani</td>
<td></td>
</tr>
</tbody>
</table>

### REVIEWS

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse Strategies (Studies in Interactional Sociolinguistics 1)</td>
<td>111</td>
</tr>
<tr>
<td>John J. Gumperz</td>
<td></td>
</tr>
<tr>
<td>Reviewed by Neal Bruss</td>
<td></td>
</tr>
<tr>
<td>The Input Hypothesis: Issues and Implications</td>
<td>116</td>
</tr>
<tr>
<td>Stephen D. Krashen</td>
<td></td>
</tr>
<tr>
<td>Reviewed by Kevin R. Gregg</td>
<td></td>
</tr>
</tbody>
</table>

### BRIEF REPORTS AND SUMMARIES

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Another Look at Passage Correction Tests</td>
<td>123</td>
</tr>
<tr>
<td>Terence Odlin</td>
<td></td>
</tr>
<tr>
<td>The Effect of Induced Anxiety on the Denotative and Interpretive Content of Second Language Speech</td>
<td>131</td>
</tr>
<tr>
<td>Faith S. Steinberg and Elaine K. Horwitz</td>
<td></td>
</tr>
<tr>
<td>The Influence of Background Knowledge on Memory for Reading Passages by Native and Nonnative Readers</td>
<td>136</td>
</tr>
<tr>
<td>Helen Aron</td>
<td></td>
</tr>
</tbody>
</table>
A Major Work on Language Testing

New Technological Developments in Measurement Theory and Methods of Test Administration

Papers from the 7th Annual Language Testing Research Colloquium, ETS, Princeton, 1985


Members, $10; Non-members, $12.50 plus $1.50 postage. All orders must be prepaid.

TEACHERS OF ENGLISH TO SPEAKERS OF OTHER LANGUAGES
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 the TESOL Newsletter 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. A TESOL Membership Application is printed
in every issue of the Quarterly. For further information, contact the TESOL Central
Office, Georgetown University, Suite 205, 1118 22nd Street, N.W., Washington, D.C.
20037.

TESOL QUARTERLY is published in March, June, September, and December. Subscriptions
are not sold without membership. Communications regarding back issues, single copies, and
permission to reprint should be addressed to James E. Alatis, Executive Director, TESOL,
Georgetown University, Suite 205, 1118 22nd Street, N.W., Washington, D.C. 20037.
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 NEWSLETTER is published in February, April, June, August, October, and December.
The Newsletter contains organizational news and announcements, Affiliate and Interest Section
news and information, book reviews, conference information and reports, short articles on
current classroom practices, and general information. The Newsletter is available only through
membership in TESOL or by subscription through a regional affiliate of TESOL. All
contributions should be directed to the Editor, Alice H. Osman, F.H. LaGuardia Community
College/CUNY, 31-10 Thomson Avenue, Long Island City, NY 11101.

ANNOUNCEMENTS should be sent directly to the Editor, TESOL Newsletter, two months
prior to the month of publication and must be received by the 20th of that month (e.g., February
20 for the April issue). The TESOL Quarterly does not publish announcements.

ADVERTISING in all TESOL publications is arranged by Aaron Berman, TESOL Development
and Promotions, P.O. Box 14396, San Francisco, CA 94114.

OFFICERS AND EXECUTIVE BOARD 1985-86

President
JEAN HANDSCOMBE
North York Board of Education
Toronto, Ontario

First Vice President
JOAN MORLEY
University of Michigan
Ann Arbor, Michigan

Second Vice President
MICHELE SABINO
Houston Police Academy
Houston, Texas

Executive Director
JAMES E. ALATIS
Georgetown University
Washington, D.C.

Penelope M. Alatis
Francis C. Hammond
Junior High School
Alexandria, Virginia

Richard Allwright
University of Lancaster
Lancaster, England

Charles H. Blatchford
Fair Oaks, California

Jeffrey P. Bright
Chicago Urban Skills Institute
City Colleges of Chicago
Chicago, Illinois

Mariamne Celce-Murcia
University of California, Los Angeles
Los Angeles, California

JoAnn Crandall
Center for Applied Linguistics
Washington, D.C.

John Haskell
Northeastern Illinois University
Chicago, Illinois

Elliot L. Judel
University of Illinois at Chicago
Chicago, Illinois

Jean McConochie
Pace University
New York, New York

Carole Urruá
Lewis and Clark College
Portland, Oregon

Executive Assistant
CAROL M. LeCLAIR
Georgetown University
Washington, D.C.

TESOL QUARTERLY
Editor’s Note

Ann Fathman has for many months expressed a wish to step down from her position as Brief Reports and Summaries Editor of the Quarterly. Ann has served in this position since 1980, when she began editing what was then called the Research Notes section of the Quarterly. She has worked with four different Quarterly editors and presided over the evolution of this section into an increasingly important part of each issue.

I would like to thank Ann Fathman for her more than 6 years of service to the Quarterly. She has given freely of her time and energy and has performed a substantial service to TESOL. I am sure that previous editors of the Quarterly join with me in wishing Ann success in her future professional endeavors.

I take pleasure in announcing that Ann’s successor as Brief Reports and Summaries Editor will be Scott Enright of Georgia State University. Scott, who will complete his term as Chair of the ESOL in Elementary Education Interest Section of TESOL at TESOL ’86, brings to the Quarterly a range of professional interests and experience in professional writing that will insure the continued growth of the Brief Reports and Summaries section.

Since Scott will assume the editorship of the section beginning with the September 1986 issue, all submissions to the Brief Reports and Summaries section should be sent, effective immediately, to Scott Enright at the address listed in the Information for Contributors section of this issue.

In This Issue

The TESOL Quarterly begins its 20th year of publication with an issue that reflects very clearly the commitment in our profession to bring critical thought and empirical research to bear on issues in program design and administration and classroom practice. The contributors to this issue explore a number of topics, including computer-assisted language learning, listening comprehension, the effect of feedback on errors, ESL classroom discourse, placement testing, and the use of learner perceptions to guide curriculum design.
Stephen Nagle and Sara Sanders summarize current theories and models of second language acquisition, none of which, they claim, “directly attempts to describe linguistic production or comprehension.” Drawing on research on memory and verbal-input processing—the psycholinguistic foundations of comprehension—they offer a model of listening comprehension processing which views comprehension and learning as “interrelated, interdependent, but distinctive cognitive phenomena.” In discussing the pedagogical implications of their model, the authors reaffirm the importance of recent attempts to give listening comprehension a significant role in language instruction and argue that “listening comprehension activities facilitate the natural development of linguistic knowledge in a setting which is affectively conducive to language acquisition.”

Carol Chapelle and Joan Jamieson report the results of a study which investigated the effectiveness of computer-assisted language learning (CALL) in the acquisition of English by Arabic- and Spanish-speaking university students in an intensive ESL program. The study also explored the relationship between CALL and selected individual and cognitive/affective variables. On the basis of their findings, the authors conclude that “CALL cannot be evaluated without looking at the other student variables . . . that are important in second language acquisition.” In addition, they stress the point that in assessing the value of CALL, we cannot look at CALL as representing one form of instruction exclusively which all students need; rather, it is “necessary to assess the characteristics of students and analyze the approach taken in a particular lesson or series.”

Cynthia Brock discusses the findings of a study to determine whether the number of referential questions asked by teachers could be increased through training and whether an increase in the number of referential questions asked by teachers would have an effect on adult ESL classroom discourse. The author found that the teachers who received training in asking referential questions did increase the number of referential questions they used in their classroom teaching and that the differences in the language produced by learners in responding to referential versus display questions were “pronounced.” Brock’s study thus provides preliminary evidence that “[referential] questions may be an important tool in the language classroom, especially in those contexts in which the classroom provides learners their only opportunity to produce the target language.”

Mary Ann Christison and Karl Krahnke conducted open-ended interviews based on a structured set of topics with 80 students who had completed an intensive English program and were engaged, at the time of the investigation, in full-time study at five different U.S. universities. The interviews were done “to determine how nonnative English speakers studying in U.S. colleges and universities perceive their language learning experiences and how they use English in academic settings.” The authors found that by revealing “conflicting but genuine
underlying needs of students during the difficult process of learning and using a new language;” the interview technique can elicit “empirical data that reflect what is useful to students” and that can be the basis of “sound curriculum design in ESL programs for academic preparation.”

• To determine “the most effective and practical feedback strategy in an EFL context characterized by extremely large teacher-to-student ratios and little contact time,” Thomas Robb, Steven Ross, and Ian Shortreed designed a study which contrasted four methods of providing feedback on errors in written compositions. These methods, which “differed in the degree of salience provided to the writer in the revision process,” were used to provide feedback to Japanese college freshmen in different sections of an English composition course during an academic year. The results of the study suggest that “in general, the more direct methods of feedback do not tend to produce results commensurate with the amount of effort required of the instructor to draw the student’s attention to surface errors.”

• Edith Hanania and May Shikhani report on a study carried out at the American University of Beirut “to determine whether the addition of a cloze component to [a] standardized ESL test would improve the predictability of students’ communicative proficiency as reflected in their performance on a written test.” In their article, the authors describe the process by which cloze passages were pilot tested and how the feasibility of administering a three-instrument placement battery was determined. From their analysis, Hanania and Shikhani conclude that for placement purposes, “a cloze component can be a valuable supplement to a standardized ESL test.” Furthermore, the fact that “the cloze and writing tests appeared to measure in common some aspects of language ability” lends support to the view that cloze can be a valid and practical alternative to a writing task as a measure of communicative proficiency.

Also in this issue:

• Reviews: Neal Bruss reviews John J. Gumperz’s Discourse Strategies, and Kevin Gregg reviews Stephen Krashen’s The Input Hypothesis.

• Brief Reports and Summaries: Terence Odlin reports on a study of a passage correction test as a measure of editing skill in a second language and as a tool for second language acquisition research; Faith Steinberg and Elaine Horwitz report the results of a study of the effects on second language speech of experimentally induced anxiety; and Helen Aron discusses the implications for placement and proficiency testing of a study which investigated the effect on nonnative performance of reading passages which require culture-bound background knowledge.

• The Forum: Daniel Horowitz discusses the process-oriented approach to the teaching of writing in “Process, Not Product: Less Than Meets
the Eye,” and Alan Beretta, in “Toward a Methodology of ESL Program Evaluation,” presents the rationale for viewing program and methods evaluation as “first and foremost applied inquiry.”

Stephen J. Gaiers
Comprehension Theory and Second Language Pedagogy

STEPHEN J. NAGLE
University of South Carolina—Coastal

SARA L. SANDERS
University of South Carolina—Columbia

Second language acquisition theories and models over the past 10 years have focused primarily on learner variables, long-term language storage, and retrieval for production. This article presents a synthesis of second language acquisition research and adds interpretations of research on memory and verbal-input processing which relate to second language acquisition. From these perspectives, a theoretical model of listening comprehension in the adult language learner is developed. Implications of comprehension theory for second language teaching are then examined in light of suggestions in the pedagogical literature for increased emphasis on listening comprehension in the classroom.

In the last 10 years, much attention in second language acquisition (SLA) research has been devoted to devising theories and models which describe and explain crucial factors and processes involved in adult L2 learning. The wealth of influential variables makes modeling the adult L2 learning process quite complex, since the adult language learner’s affective makeup and conscious awareness of language rules can greatly influence, positively or negatively, L2 achievement.

In the developing literature on adult L2 learning, most models have been selective, focusing upon and emphasizing various elements, processes, or activities critical to a given theory. Curiously, there has often been a tendency in the SLA literature to use theory and model almost synonymously. A model must graphically represent theory in an economical way, but a theory does not in itself constitute a model.

In recent years SLA model building has become increasingly sophisticated, as is reflected in the trend toward representing adult language-learning theory in an information-processing framework.
A great deal of important, insightful theoretical research is being incorporated in formal models. However, the process of language comprehension, which furnishes new information to be assimilated by a language learner, is generally assumed rather than specifically examined in the theoretical literature, though some literature on second language teaching has strongly emphasized listening comprehension activities. This article first reviews theoretical foundations of current SLA models. This research as well as various studies on memory and verbal-input processing are drawn upon to present a model of adult L2 comprehension. Finally, the implications of comprehension theory for second language pedagogy are discussed.

CURRENT MODELS OF SLA

Beginning with Taylor (1974), SLA researchers have placed considerable emphasis on learner variables. Taylor has proposed that the principal difference between first language acquisition and adult second language acquisition lies in the complex affective makeup of the adult. Schumann (1976), Yorio (1976), and Strevens (1977) have presented schematizations of the interrelationships among important affective, cognitive, instructional, and other variables. None of these analyses, it may be noted, directly attempts to describe linguistic production or comprehension, but many of the crucial factors identified must be taken into account in a theory of L2 comprehension.

Krashen’s (1977) Monitor Model and accompanying theory have had the most powerful impact to date on SLA research (and on second language teaching). The Monitor Model, linked in more recent form with Dulay and Burt’s (1977) Affective Filter Hypothesis (see Figures 1 and 2), provides a persuasive scheme of processes and activities in L2 learning. Krashen’s theory rests on several fundamental principles:

1. Acquisition and learning are technical terms representing separate phenomena. Acquisition is motivated by a focus on communication and is not conscious; learning is motivated by a focus on form, is conscious, and results in metalinguistics knowledge.

2. In speech production, acquired and learned forms are generated separately, with monitoring and conscious attention to performance often modifying output; the amount of monitoring is a variable. The Monitor, as presented by its proponents, is an output component and has no effect on acquisition.
3. The conditions for optimal Monitor use are a focus on form, sufficient time, and knowledge of a pertinent rule.

4. The Monitor can be overused or misused, resulting in hesitant and/or deficient target language production.

5. In decoding L2 input, affective variables can impede acquisition and learning. This phenomenon is represented schematically by an Affective Filter.

Recent work by Krashen and others incorporates extensive observations on and recommendations for language teaching and the treatment of errors. Dulay, Burt, and Krashen (1982) elaborate on implications of the theory, as does Krashen (1982). The strong reception accorded to Dulay, Burt, and Krashen’s work has resulted from the fact that their ideas are systematically elaborated and that they fit with L2 experiences that learners undergo, such as the frustration that occurs when conscious output processing (monitoring) is inhibitive. Too much attention to form can result in an inability to communicate.
The principal objections to the theories most closely associated with the work of Krashen have dealt with the rigid separation of acquisition and learning, both as memory stores and components in speech production. Indeed, the notion that learned forms can never be transferred to acquisition is difficult if not impossible to verify, and Krashen and his co-theorists provide little objective support.

Bialystok (1978) prefers *implicit* and *explicit* to *acquired* and *learned*. Her model of second language learning (see Figure 3), which deals with strategies as well as processes, allows for the transfer of linguistic knowledge from the explicit to the implicit domain and suggests that formal practicing can motivate the shift. Similarly, Stevick (1980) argues for “seepage” (p. 276) from learning to acquisition. Adherents to the Krashen approach argue strongly that formal activities result in learning, not acquisition. Another
difference between Bialystok and Krashen is Bialystok’s emphasis on the importance of nonlinguistic knowledge in language learning and her inclusion of other knowledge as a component in her model.

Implicit support for viewing learned and acquired forms as transferable is found in Lamendella’s (1977) outline of the neurofunctional system, a system of hierarchical networks, or infra-systems, of information processing. Lamendella’s two principal hierarchies, the cognition hierarchy and the communication hierarchy, are viewed as related “neurofunctional metasystems” (p. 159) which differ in function. In adults, the cognition hierarchy is essentially a problem-solving component involved in foreign language learning, while the communication hierarchy is responsible for “primary” and “secondary” language acquisition. The important part here is that the systems are not dichotomous, as are Krashen’s acquisition and learning.

Subsequent to Lamendella (1977), Selinker and Lamendella (1978) proposed an executive component which oversees processing operations and controls the flow of information. The executive component transmits input to either hierarchy and thus is responsible for the learning or acquisition of linguistic forms. We will return to this concept shortly.

Tollefson, Jacobs, and Selipsky (1983) have presented a model (see Figure 4) which integrates components of the Bialystok, Krashen, and Lamendella models. In their view, learned and acquired knowledge, though stored separately, may be transferred to another hierarchy. The Monitor, as suggested by Bialystok (1978), affects input as well as output. Thus, the Monitor and the Affective Filter operate on input directed to the executive component, presumably influencing its processing choices. This model, which is quite advanced in both its synthesis of current theory and its representation of important theoretical constructs in an information-processing design, is a cogent example of how a model can represent a multiplicity of theoretical notions.

Like most L2 models, however, this model primarily depicts components involved in acquisition/learning and is not specifically applicable to listening comprehension. Since linguistic knowledge derives from comprehended input, which is in the learner a subset of the available raw language input, researchers and teachers alike may find an acquaintance with the psycholinguistic foundations of comprehension to be highly instructive. Therefore, we will review briefly some well-known contributions to memory and information-processing theory to see how they may support, complement, and enrich SLA theory.

COMPREHENSION THEORY AND SECOND LANGUAGE PEDAGOGY
PSYCHOLINGUISTIC FOUNDATIONS OF COMPREHENSION

Memory

Most recent models of second language acquisition have assumed discrete linguistic knowledge stores, whether learned/acquired or implicit/explicit. Major disagreements have involved the possibility of transferring knowledge from one to the other. Bialystok (1978, 1981) has stressed the importance of nonlinguistic other knowledge in processing. All three components are involved in long-term storage of information, which has been the principal concern of
SLA researchers. With the noteworthy exception of Stevick (1976), these researchers have devoted little attention to other aspects of human memory.

Most contemporary analyses of human memory have distinguished between short-term and long-term memory. Since Miller (1956), J. Brown (1958), and Peterson and Peterson (1959), short-term retention has been the subject of intensive investigation. An influential outline of memory by Atkinson and Shiffrin (1968) includes a short-term store (STS) of limited capacity and time span and a long-term store (LTS) of much greater capacity and duration. Atkinson and Shiffrin further include a sensory register of very brief duration. In auditory processing, sensory memory has also been called echoic memory (Neisser, 1967) and precategorical acoustic storage (Crowder & Morton, 1969), though a boundary between purely sensory retention (which involves no processing) and short-term retention (in which items are subject to processing) has been difficult to establish.

Input-Processing Activities

When information is temporarily stored in initial memories (sensory and short-term), activities such as scanning, searching, and comparing may relate it to other information in long-term storage, resulting in comprehension. Two factors, however, impede the processing of new information: trace decay (fading of the sensory input) and interference from newly arriving input. On the other hand, rehearsal (conscious and unconscious repetition) may strengthen an item in short-term memory. More important for second language acquisition theory, there is a consensus among researchers in memory that rehearsal is an important variable in fostering long-term retention as well. This viewpoint is reflected in the role accorded to practicing in the Bialystok (1978) and Tollefson et al. (1983) models but is overlooked in and, in fact, directly contradicts the acquisition/learning distinction propounded by Krashen.

McLaughlin (1978) and McLaughlin, Rossman, and McLeod (1983) have presented a strong theoretical challenge to the notion of dichotomous long-term language storage. Drawing upon extensive research by Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977), McLaughlin (1978) argues that the equating of learning with conscious processing is an overgeneralization.

1 Rehearsal is a natural process of which an individual may or may not be aware. Role rehearsal does not correlate well with probability of recall in lists of words held (rehearsed) for varying lengths of time (Craik & Watkins, 1973) but has been shown (Woodward, Bjork, & Jongeward, 1973) to correlate positively with probability of recognition.
Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977) identify two principal processing modes: controlled processing and automatic processing. A controlled process, according to Shiffrin and Schneider, "utilizes a temporary sequence of nodes activated under control of, and through attention by, the subject" (p. 156). Certain task demands may encourage this type of processing, and it is not necessarily conscious in all cases. An automatic process is a "sequence of nodes that nearly always becomes active in response to a particular input configuration" and is "activated without the necessity of active control or attention by the subject" (p. 155). Automatic processes require sufficient training to develop, since they depend upon a relatively permanent set of node associations. This training is provided by controlled processing.

McLaughlin et al. (1983) note that most automatic processing occurs incidentally (see Figure 5, Cell D) in normal communication activities, while most controlled processing occurs in performing new language skills (Cell A in Figure 5) which require a high degree of focal attention. They note that the development of the skills necessary to deal with complex tasks such as language processing "involves building up a set of well-learned, automatic processes so that controlled processes will be freed up for new tasks" (p. 144). Automatic processing is critical to comprehension because too much controlled processing may lead to overload and breakdown.

FIGURE 5
Performance as a Function of Information Processing and Focus of Attention
(McLaughlin, Rossman, & McLeod, 1983)


<table>
<thead>
<tr>
<th>Focus of attention</th>
<th>Information processing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controlled</td>
</tr>
<tr>
<td></td>
<td>(Cell A)</td>
</tr>
<tr>
<td>Focal</td>
<td>&quot;Intentional&quot; performance of a new skill</td>
</tr>
<tr>
<td>Peripheral</td>
<td>(Cell C)</td>
</tr>
<tr>
<td></td>
<td>&quot;Incidental&quot; performance of a new skill</td>
</tr>
</tbody>
</table>

If appropriate automatic processes are not available or are not activated in a given comprehension task, the primary resource at the individual's disposal is attention (as used here, McLaughlin et al.'s
focal attention). Attending involves the application of mental energy to processing tasks and may range from focusing on specific features of input to controlled processing for retrieval. In a recent critique of their earlier model of visual processing (LaBerge & Samuels, 1974), Samuels and LaBerge (1983) have stressed the limited amount of energy (attention) available and propose that tasks may be divided into smaller processing units when attention capacity is exceeded. Each unit can then be dealt with individually, but too much subdivision can result in slow, laborious processing. With practice, however, input that once required subdivision can be dealt with automatically, allowing attention to be held in reserve for more complex tasks.

The role of attention in input processing is similar in many respects to Krashen’s (1977, 1982) view of the Monitor’s role in language production. The Monitor focuses on form(s); that is, it analyzes (or subdivides) linguistic units into smaller components. If one views the Monitor as an input processor as well, monitoring may be described as the directing of attention to specific input (or output) items. Thus, attention (or monitoring) is an important variable, since too much subdivision can overload the attention system, filtering out other input items and causing a breakdown in processing.

A major factor in activating attention is arousal, which entails an increase of activity in the nervous system. Baddeley (1972), among others, has presented evidence that an increase in the level of arousal may lead a subject to concentrate on a smaller number of environmental cues. Hamilton, Hockey, and Quinn (1972), in testing recall of items presented in associated word pairs, found that noisy conditions (which may cause arousal) enhanced recall performance when items were elicited in the same order in which they were presented, but impaired recall when items were tested in scrambled order. Arousal, then, may foster attention to explicit matters such as order or form. Hulstijn and Hulstijn (1984) have demonstrated that L2 learners with varying degrees of explicit and implicit knowledge show increased correctness in performance when asked to pay attention to form, that is, in teacher-controlled arousal situations. Arousal may have a similar effect on comprehension tasks, activating attention and encouraging appropriate controlled processing and monitoring.

A MODEL OF ADULT SECOND LANGUAGE LISTENING COMPREHENSION

The interrelatedness among arousal, attention, monitoring, and
controlled and automatic processing suggests some sort of general control mechanism for dealing with input, such as Selinker and Lamendella’s (1978) executive component. Shiffrin (1970) has posited an executive decision maker, Craik and Lockhart (1972) and Craik (1973) have proposed a central processor in the short-term memory system, and Baddeley and Hitch (1974) have suggested that short-term memory contains a working memory component.\footnote{We are not suggesting that these notions reflect unanimity in memory theory. There are, of course, similarities in the constructs proposed by these researchers; however, Craik and Lockhart (1972) are strong proponents of a “levels of processing” view of memory which differs from the traditional dichotomous approach.}

Adams (1971) has proposed that human monitor behavior is closed-loop rather than open-loop. In a closed-loop system, information about success or errors in processing is fed back to the control center, which may then reprocess if necessary. An open-loop system has no such feedback mechanism. In language comprehension, one may continue to process input by directing attention to items not immediately comprehended; in SLA terminology, extended processing involves monitoring of input and application of explicit knowledge (learning). Viewing input processing as closed-loop also provides insight into the relationship between comprehension and acquisition/learning. The basis for meaning is the synthesis of retrieved knowledge and the individual’s judgments (inference) about unfamiliar data; processing results (even if “incorrect”) returned to the executive are available for long-term storage.

The choice made by the executive, involving activation and direction of attention and the degree to which various long-term stores will be accessed, is subject to variables such as task complexity, content, time constraints, and affective factors. If we view the executive, or working memory, as a part of the short-term memory system, we may plausibly view affective filtering and overuse of monitoring as interrelated in weakening the processing of a portion of the items in short-term storage at a given time.

In contrast with other second language models, the model in Figure 6 represents listening comprehension, not learning. Comprehension both adds to and draws upon learning, but it involves more than simple retrieval from discrete long-term storage. Not only is it influenced by various psychological and task-specific variables, it also draws upon an individual’s inferences about new data based on all types of knowledge about language and the world. From this perspective comes a sensible view that comprehension
FIGURE 6
A Model of Listening Comprehension Processing in the Adult Language Learner

Note: STS = short-term storage; LTS = long-term storage.
and learning are interrelated, interdependent, but distinctive cognitive phenomena.

Because of their interrelationship, however, theoretical learning constructs such as the Monitor and the Affective Filter derive some support, at least as broad generalizations, from psychological investigations of comprehension processes. Further, the gradual progress from controlled to automatic processing outlined by Shiffrin and Schneider (1977) underlies both comprehension and learning and supports the traditional view of teachers that practice leads to learning. Thus, theoretical positions held by second language researchers and psychologists are to a large degree complementary, except in the extreme case of Krashen’s view of linguistic memory.

PEDAGOGICAL IMPLICATIONS

Since comprehension makes material available for learning, it is reasonable to assume that comprehension is an optimal starting point of instruction in the target language and, further, that comprehension activities should be incorporated at all instructional levels. Systematic investigation of listening comprehension as a skill was not of great concern until the 1970s (Dirven & Oakeshott-Taylor, 1984, 1985); however, there is an increasing conviction among language teachers that listening comprehension is a global skill which can be taught (Byrnes, 1984, p. 325). While investigation of listening comprehension as a skill is just now coming into its own, concern with the role of listening in teaching languages is not new. Nida (1957), Asher, Kusudo, and de la Terre (1983), Postovsky (1974), Winitz (1981), Belasco (1981), Stevick (1976, 1980), and Krashen and Terrell (1983) are among those who have advocated a listening comprehension approach to language instruction and whose work reflects a heightened interest in giving listening comprehension a significant role in language instruction.

Nord (1981) proposes three progressive phases in the development of listening fluency: (a) semantic decoding; (b) listening ahead or anticipating the next word, phrase, or sentence; and (c) discrepancy detection. He notes that progressing through these stages produces a “rather complete cognitive map” (p. 98) which has a beneficial effect on the development of speaking, reading, and writing skills. Nord’s “cognitive map” might be viewed in terms of the model in Figure 6 as sets of related material in the long-term store, automatic processes for dealing with much of the retrieval, and efficient strategies for controlled processing of new information.
THE AUTHORS

Stephen J. Nagle is Assistant Professor of English at the University of South Carolina’s Coastal Carolina College, where he teaches English to native and nonnative speakers as well as courses in linguistics. His areas of interest include psycholinguistics, reading theory, and historical linguistics.

Sara L. Sanders is Director of the intensive English Program for Internationals at the University of South Carolina and Adjunct Professor of Linguistics. Her research interests include ESL teacher training, curriculum design, and oral proficiency testing.

REFERENCES


Computer-Assisted Language Learning as a Predictor of Success in Acquiring English as a Second Language

CAROL CHAPELLE
Iowa State University

JOAN JAMIESON
University of Illinois at Urbana-Champaign

This article reports the results of a study of the effectiveness of computer-assisted language learning (CALL) in the acquisition of English as a second language by Arabic- and Spanish-speaking students in an intensive program. The study also examined two student variables—time spent using and attitude toward the CALL lessons—as well as four cognitive/affective characteristics—field independence, ambiguity tolerance, motivational intensity, and English-class anxiety. English proficiency was measured by the TOEFL and an oral test of communicative competence. Results indicated that the use of CALL lessons predicted no variance on the criterion measures beyond what could be predicted by the cognitive/affective variables. In addition, it was found that time spent using and attitude toward CALL were significantly related to field independence and motivational intensity. These results indicate that (a) certain types of learners may be better suited to some CALL materials than other students and (b) it is necessary to consider many learner variables when researching the effectiveness of CALL.

Three questions are often asked about computer-assisted language learning (CALL): Do students like it? Do students use it? Does it work? These questions address practical concerns, yet they are based on two faulty assumptions. First, they assume that students think and act in a uniform manner, even though teachers and researchers alike agree that students differ in their learning styles and strategies. Second, the questions presuppose that CALL is a single method of instruction, whereas it is actually a vehicle for implementing a range of approaches representing a variety of teaching philosophies. These points do not deny the basic
importance of asking questions about the value of CALL; instead, they indicate the need to modify the questions: What kind of students like and use a particular type of CALL? Do those students who use CALL achieve greater success in the second language?

These were the questions posed in the research reported in this article, which sought to (a) characterize students who chose to use CALL when they had the option to do so and (b) discover whether students’ use of CALL accounted for variance in end-of-semester ESL performance beyond what could be explained by other variables.

COMPUTER-ASSISTED LANGUAGE LEARNING

To evaluate the effectiveness of CALL, it is important to understand the reason for having students practice ESL on the computer. Computer-assisted instruction (CAI) has evolved around three distinguishable, though interrelated, instructional ideals: individualization, record keeping, and answer judging.

Individualization in CAI refers to the fact that the computer enables students to work alone and at their own pace. Through the use of individualized instruction, poor students can attain additional practice outside of the classroom so that the teacher does not have to slow down the rest of the class. Individualization also allows the teacher to maintain the interest of good students by providing them with advanced materials. Individualized instruction provided by CAI has been used as an adjunct to classroom instruction in some cases and as the sole method of instruction in others (Chapelle & Jamieson, 1983; Otto, 1981; Smith & Sherwood, 1976; Suppes, 1981).

To provide an individualized learning environment, many developers have used a systems approach to design: A learning hierarchy is formulated, and a diagnostic mechanism is used so that either the computer program or the student can decide when the student needs to review (Bunderson, 1970; Dick & Carey, 1978; Tennyson, 1981). The difficulty, however, is in designing a diagnostic mechanism that will enable each student to proceed along a tailor-made path. Although its potential has been demonstrated, individualization has not been achieved at a sophisticated level (Hart, 1981; Kearsley, Hunter, & Seidel, 1983). To provide a student with an ideal learning path through a lesson, the lesson author must have a well-defined understanding of how students learn.

This traditional view of individualization in CAI has recently been seen in a new light. Some educators have proposed that students use the computer as a means of exploring and playing with
material (such as the target language) through group work, games, and student-initiated exchanges (Higgins & Johns, 1983; Underwood, 1984). In such an environment, students create their own learning experiences; therefore, it is difficult for the lesson designer to know what (and if) each student learns from a lesson, particularly in the case of students who have typically been unsuccessful (Steinberg, 1977).

The capability of collecting data and keeping records is a second advantage of CAI. Data on any interaction that occurs between the student and computer can be collected and subsequently analyzed. For example, students’ wrong answers in a drill can be collected and analyzed to improve the program’s error diagnosis and remediation. Record keeping is also beneficial for providing the student and/or teacher with a profile of the student’s mastery of material (Marty, 1981, 1982). Another benefit of record keeping is in the area of research; data can be collected to search for patterns in students’ learning.

Some CAI materials have incorporated research findings that indicate students learn better when (a) they have to answer questions (rather than simply read material) and (b) they receive “knowledge of the correct response” (e.g., Anderson, Kulhavey, & Andre, 1971; Sassenrath, 1975). Thus, the third advantage of CAI is embodied in answer judging. Answer judging occurs after students answer a question posed by the computer: The computer informs them whether it is right or wrong. Moreover, if the answer is wrong, the program should provide students with a meaningful explanation as to why the answer is wrong. If the program can recognize and classify students’ wrong answers, then it can save this information as student records and provide students with appropriate remedial activities (Hartley, 1974; Marty & Meyers, 1975).

Although the potential of each of these ideals has been demonstrated, their implementation on a large scale remains to be seen. In spite of the limitations of current courseware, a number of studies have been done on attitude and achievement with CAI. This research indicates that CAI is usually a popular method of instruction which is typically as effective as regular classroom instruction and may require less time on task for mastery of the target skills (e.g., Collins, 1978; Freed, 1971; J. A. Kulik, Bangert, & Williams, 1983; J. A. Kulik, C.-L.C. Kulik, & Cohen, 1980; Tsai & Pohl, 1977, 1980; Van Campen, 1981), although there are notable exceptions to this conclusion (Alderman, 1978; Murphy & Appel, 1977).

Attempts to put these ideals into practice in ESL courseware have resulted in lessons that differ from one another in a number of
relevant ways. First, ESL courseware is used to teach skill areas such as reading, writing, listening, and grammar, as well as to provide practice in using the target language by engaging the student in games or problem-solving activities. Lessons also differ with respect to the use of the target language: Some lessons use discrete elements within the language to delimit and simplify the learning task; others incorporate language in a natural context, allowing the student to practice in a more authentic L2 environment. A third difference is the kind of learning objective. Some lessons have very clearly defined objectives (e.g., the student will form the present perfect correctly); others do not (e.g., the student will interact with the program to discover its limitations). Finally, a lesson can be characterized by placing it somewhere along a continuum ranging from machine-controlled to student-controlled. In a machine-controlled lesson, the instructional decisions are made by the program; the student simply follows the program’s instructions. A student-controlled program, on the other hand, allows the student much freedom in initiating learning decisions.

METHOD

Subjects

The students enrolled in the Intensive English Institute at the University of Illinois during the Fall 1982 semester were invited to participate in the research by a letter translated into their native languages. Of the 84 students in the Institute, 28 Spanish-speaking and 20 Arabic-speaking students agreed to participate. The subjects ranged in age from 18 to 40 and had TOEFL scores ranging from 430 to 510.

Materials and Procedure

The ESL PLATO courseware is primarily a drill and practice curriculum of lessons in three skill areas: grammar, reading, and listening. Although the content differs, the lessons share many design features.

Grammar is presented in two series of lessons. The first, a series of 20 Remedial Grammar lessons, provides an intensive review of grammatical points for beginning ESL students. These lessons assume a very low vocabulary level, include a simple grammatical generalization, and provide extensive practice of specific grammar points using a wide variety of exercises. A built-in review is provided for items that are missed in each exercise.

The second series, 16 Advanced Grammar Review lessons,
provides extensive reinforcement and practice of a wide range of advanced grammar points. These lessons provide supplementary practice with minimal grammar explanations. Each of the lessons consists of at least four mechanical exercises, including substitution, transformation, question/answer, and fill-in-the-blank drills. Items answered incorrectly in these lessons are also recycled for reinforcement (see also, Stevens, 1983).

The reading lessons are also subdivided into two different series. The lower-level Vocabulary and Culture series consists of 12 lessons that simultaneously introduce and teach real-world vocabulary, familiarize the student with some important aspects of American culture, and check on the student’s command of specific grammar points, in accordance with the Remedial Grammar lessons. The lessons portray the main character, Peter Adams, in his dormitory room, at the local post office, at a restaurant, and so on.

The objectives of the higher-level Reading and Comprehension series are to (a) test comprehension of a passage, (b) increase reading speed, (c) increase active and passive vocabulary, and (d) acquaint foreign students with some aspects of American culture and history. The reading passages in each of the eight lessons consist of six or seven paragraphs that are displayed individually. While reading each paragraph, students have the option to ask for definitions of words. If a queried word was anticipated as a troublesome vocabulary item, students are given three synonyms from which to choose; otherwise, they are told that the word is not in PLATO’s dictionary. After students have read all of the paragraphs, they first answer multiple-choice comprehension questions about each paragraph, then complete a restatement or paraphrase exercise, and finally type a derivative of a keyword in the correct grammatical context.

The listening lessons are of two different types, Spelling and Dictation, each of which has two levels corresponding with the low and high levels in grammar and reading. The two Spelling series, each of which has 14 lessons, differ only in the level of difficulty of the words. The instructional exercises used in these series elicit both aural recognition and written production from the student. Each lesson consists of three lists of 10 words. Students first see a list of the words. Then they hear a word in isolation, in a sentence, and repeated in isolation. For example: “morning. John reads the newspaper in the morning. Type morning.” Some spelling errors are anticipated based on contrastive analysis of English and other languages. Incorrectly answered items are recycled at the end of each of the three segments of the lesson.

The Dictation series also has 14 lessons at each of the two levels
of difficulty. Each lesson contains two parts, a list of 10 sentences and a paragraph of 5 sentences. Students touch the screen, which in turn activates a random-access audio device, and they then hear a sentence through their headphones (much like in the Spelling lessons). Students have the option of hearing all or part of the sentence as often as necessary to complete the task, which is to type the sentence. An answer with an error is indicated to students not only by a “wrong” message but also by special symbols that indicate misspellings, inversion, errors in capitalization and punctuation, or extra words. After the correct answer has been entered, students have the option of continuing or of recording their voice and then comparing it to the model, as in a language laboratory.

All of the lessons in the eight series described above have some common design features. The lessons, which do not give students practice with global language use, employ discrete elements of language to present materials which have a clearly defined objective. For example, in the Dictation lessons students hear a sentence such as “The women asked for some instructions,” which they are directed to type. The sentence occurs in the lesson at this point to provide students with practice on past tense and quantifiers. After completing this item correctly, students can go on to the next, which may have nothing to do with what the women did with the instructions—no meaningful context is built. The PLATO lessons are more machine-controlled than learner-controlled. Although students choose from a menu the order in which they will complete the week’s lessons, the lessons themselves provide the learners with very few options.

Variables

To learn what kind of students were CALL users, it was necessary to examine a number of student variables. Affective and cognitive differences among individuals are numerous and multidimensional; however, on the basis of previous research, several variables were isolated for their importance in second language acquisition.

Field independence/dependence. Field independence/dependence (FI/D), a cognitive variable, is defined as “the extent to which a person perceives part of a field as discrete from the surrounding field as a whole, rather than embedded, or . . . the extent to which a person perceives analytically” (Witkin, Moore, Goodenough, & Cox, 1977, p. 7). A field independent (FI) person tends to approach problem solving analytically, while a field dependent (FD) person tends to approach problem solving in a more global way. In the area of intellectual problem solving, a highly FI person is able to detect
patterns and subpatterns, while an FD person tends to get lost in the totality of the stimuli. Consequently, an FI person is at an advantage in problem-solving situations in which isolating and manipulating a critical element are important, such as word problems in mathematics (Witkin et al., 1977). An FD person, on the other hand, is more capable of perceiving the total picture in a situation.

An FI person may have good analytical language skills, such as those needed in many classroom environments, while the FD person would logically be better at acquiring a second language through interaction with native speakers in social situations. However, research supports only the former claim (e.g., Bialystok & Frolich, 1978; Hansen & Stansfield, 1981; Naiman, Fröhlich, & Stern, 1975; Roberts, 1983).

The Group Embedded Figures Test (Oltman, Raskin, & Witkin, 1971), in which subjects are asked to find a given simple figure embedded in each of 18 complex figures, was used to measure FI. One point is given for each item answered correctly so subjects with high scores are considered FI.

**Ambiguity tolerance.** Ambiguity tolerance (AT) can be defined as a person’s ability to function rationally and calmly in a situation in which interpretation of all stimuli is not completely clear. People who have little or no AT perceive ambiguous situations as sources of psychological discomfort or threat (Budner, 1962). These feelings may cause them to resort to black-and-white solutions (Frenkel-Brunswik, 1949) and to refuse to consider any gray aspects of a situation. They may also strive to categorize phenomena rather than order them along a continuum (Levitt, 1953); moreover, they may arrive at premature closure (Frenkel-Brunswik, 1949) or jump to conclusions rather than take time to consider all of the essential elements of an unclear situation. People with little AT may also try to avoid ambiguous situations. Individuals who have a great deal of AT, on the other hand, enjoy being in ambiguous situations and, in fact, seek them out. They are believed to excel in the performance of ambiguous tasks (MacDonald, 1970).

Of course, L2 situations vary with respect to the amount of ambiguity present. Although ambiguity is present in any L2 situation, there is less in a formal language class in which individual elements of language are isolated for study and more in an immersion situation in which the learner has to attend to all language cues simultaneously. Research (Chapelle, 1983; Naiman et al., 1975) supports the claim of a negative relationship between AT and L2 acquisition.

AT was measured by the MAT-50 (Norton, 1975), a 62-item, Likert-type scale which consists of statements concerning work,
philosophy, art, and other topics. Subjects are to indicate agreement
or disagreement with these statements on a 7-point scale. An
example (Item 30) is given below.

A group meeting functions best with a definite agenda.
YES!  YES  yes  ?  no  NO  NO!

A subject who answers this and similar statements with a “YES!”
would get a low total score on the AT test.

**Motivational intensity.** Motivational intensity (MOT) refers to the
strength of a student’s desire to learn the L2, as reflected by the
amount of work done for classroom assignments, future plans to
make use of the language, and the effort made to acquire the
language. The logical and empirically supported hypothesis is that
MOT contributes to success in L2 acquisition (e.g., Gardner &
Lambert, 1959; Gardner, Smythe, Clement, & Gliksman, 1976).

MOT for learning English was measured by a subscale of
Gardner and Smythe’s (1979) Attitudes and Motivation Test Battery
(AMTB), which consists of 10 items, such as the one below (Item
68).

If my teacher wanted me to do an extra assignment, I would:

- a. only do it if the teacher asked me directly.
- b. definitely volunteer.
- c. definitely not volunteer.

Students who choose Alternative b in response to this and similar
questions would get the highest score for MOT; students who
choose Alternative c would get the lowest score.

**English-class anxiety.** English-class anxiety (ANX) is the degree to
which the student feels uncomfortable and nervous in the L2
classroom. Because research has found anxiety to be both positively
(e.g., Chastain, 1975; Kleinmann, 1977) and negatively (e.g.,
Gardner et al., 1976; Swain & Burnaby, 1976) related to
performance in various language situations, a distinction has been
proposed between “facilitating” and “debilitating” anxiety (Scovel,
1978). The effects of ANX on L2 acquisition appear to depend on
the amount and kind of anxiety that the learner has, as well as on the
L2 environment.

ANX was also measured by a portion of the AMTB, which
consists of five questions, one of which is given below (Item 18).

I am afraid that the other students in the class will laugh at me when I
speak English.

Students are asked to indicate their agreement or disagreement on
a 7-point scale ranging from *strongly disagree* to *strongly agree*. A
student who strongly agreed with this and similar questions would get a high score for ANX.

**Attitude toward CALL.** Students’ attitudes toward using the PLATO lessons were assessed through three items on a general student information questionnaire (Chapelle, 1983) which focused on students’ past experiences with foreign language study and current preferences in L2 study. An example of the questions used to elicit information is given below (Item 22).

Do you like to do English lessons on PLATO?

a. Yes, very much.
b. Yes.
c. It’s OK.
d. Not really.
e. No, I hate it.

**Time spent using CALL.** In addition to the self-report data, a measure of students’ actual behavior toward CALL was obtained by tabulating the number of hours each student spent working on PLATO over the course of the semester. Each student in the intensive program is routinely assigned to work 4 hours a week in the PLATO lab. Strictly speaking, however, this lab time is not required because neither lab work nor attendance is calculated as part of the student’s grade. Consequently, students who do not care to work on PLATO typically spend fewer than their scheduled hours in the lab or cease to go to the lab at all. On the other hand, those students who like to use CALL visit the lab during their scheduled time as well as during the lab’s open hours.

**English proficiency.** Students’ English proficiency was measured at the beginning and the end of the semester by the TOEFL and an oral test of communicative competence (Bachman & Palmer, 1982). The latter, which was developed and validated on the basis of Canale and Swain’s (1980) theoretical model of communicative competence, measures three general competence areas: grammatical, pragmatic, and sociolinguistic.

In addition to the tests of English proficiency administered at the beginning and the end of the semester, the subjects were given the tests of FI, AT, ANX, and MOT and the student information questionnaire in the seventh week of the semester. All of these had been translated into their native languages.

**Analysis**

The data were analyzed using SPSS (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) to perform procedures corresponding to the two questions posed in the study. A series of analyses was...
done to address the first question, What kind of student likes to use CALL? After the measures were found to have adequate reliabilities (all > .71), Pearson product-moment correlations were calculated to determine if students’ cognitive/affective characteristics were related to time spent using CALL and attitude toward CALL. Then, a multiple regression analysis was performed to determine if one predictor variable accounted for the variance in time and attitude.

The second part of the analysis focused on the question of whether those students who used PLATO more frequently got higher scores on the end-of-semester criterion measures. The correlations between end-of-semester scores and the predictor variables—beginning-of-semester language measures, student cognitive/affective characteristics, and time spent using CALL—were calculated. Multiple regression analyses, using the end-of-semester language measures as dependent variables, were then performed.

RESULTS

Time, Attitude, and Student Affective/Cognitive Factors

The first question under investigation, whether students’ cognitive/affective characteristics were related to their time spent using and attitude toward CALL, can be answered in the affirmative with respect to the subjects tested. There was a significant negative correlation between field independence and both time and attitude, indicating that highly field independent students tended not to like to work on CALL (see Table 1).

TABLE 1
Pearson Product-Moment Correlations Among Nonlanguage Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. FI</td>
<td>.205*</td>
<td>-.184</td>
<td>-.384**</td>
<td>-.394**</td>
<td>-.423**</td>
</tr>
<tr>
<td>2. AT</td>
<td></td>
<td>.021</td>
<td>-.185</td>
<td>-.115</td>
<td>-.197</td>
</tr>
<tr>
<td>3. MOT</td>
<td></td>
<td></td>
<td>.007</td>
<td>.257**</td>
<td>.511***</td>
</tr>
<tr>
<td>4. ANX</td>
<td></td>
<td></td>
<td></td>
<td>.037</td>
<td>.152</td>
</tr>
<tr>
<td>5. Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.591***</td>
</tr>
<tr>
<td>6. CATT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: FI = field independence; AT = ambiguity tolerance; MOT = motivational intensity; ANX = English-class anxiety; Time = time spent using CALL; CATT = attitude toward CALL.

* p < .05. ** p < .01. *** p < .001.
A significant positive correlation was found between motivational intensity and both time and attitude. In other words, those students who reported themselves to be working hard at learning English also tended to spend a lot of time using CALL and had a more positive attitude toward it. The relationship between motivational intensity and attitude toward CALL (what students said they liked) was stronger than that between motivational intensity and time spent on PLATO (what students actually did). The similarity of the self-report types of questions on the attitude and motivational intensity measures undoubtedly accounts for some of their shared variance. The significant \((p < .001)\) positive correlation between the time students spent using CALL and their attitude toward CALL indicates that there is a strong relationship between what students said they liked and what they actually did.

There were no significant correlations of ambiguity tolerance and English-class anxiety with time and attitude. It was expected that students who preferred a more structured environment (those with low AT) would like to work on the PLATO lessons, that is, that AT would correlate significantly, but negatively, with attitude and time. In fact, the direction of the relationship was negative, but not to a significant degree. Similarly, it was thought that students who felt nervous in English classes would like working on English at their own private terminals. The nonsignificant correlations between ANX and the CALL variables did not support this frequently made claim.

Because field independence and motivational intensity were both significantly related to time and attitude, it was necessary to determine if both variables were needed to account for the variance in time and attitude. In other words, was it simply that the motivated students liked to use CALL and that they just happened to be field independent as well? To answer this question, four multiple regression analyses were performed (see Table 2).

Using time and attitude as dependent variables, motivational intensity was entered into the equation and found to be a significant predictor for both variables. Field independence was then entered into the equation and also found to be a significant predictor for both variables. If field independence had been significantly related to time and attitude simply because it was also related to motivational intensity, it would not have been found to be a significant predictor when entered into the multiple regression analysis after motivational intensity.

The second pair of regressions addressed the question in the reverse order: Are the students who liked to use PLATO those with little field independence who just happened to be motivated? Time
and attitude were again used as the dependent variables, but this time, however, field independence was entered first. Motivational intensity was found to predict a significant amount of additional variance in attitude, but not in time. Since motivational intensity and attitude toward CALL were both self-report measures, some of their shared variance can be accounted for by this similarity. Time spent on PLATO, on the other hand, was a measure of what students did—their actual behavior. On this measure, FI alone accounted for all of the explained variance; motivational intensity was not a significant predictor.

These analyses indicate that students who are not FI show a significant preference for using CALL; moreover, FI was the exclusive predictor of time spent on PLATO. In interpreting these results, it is important to underscore the fact that the ESL lessons on the PLATO system cannot be equated with all possible CALL; instead, they represent a particular approach—one taken in many CALL lessons—but certainly not the only possible approach. The findings of this study might have been quite different if the lessons offered on the PLATO system had represented a greater variety of approaches.

It is likely that the FI students, who are capable of and accustomed to using their own internal referents, found the structured approach of the lessons in the ESL PLATO series to be inconsistent with their learning styles. They may have found it irritating to have information and exercises structured in a way different from how they would have done it for themselves. Lacking the stimulation of using their own capabilities to select and

---

### TABLE 2
Multiple Regression Analyses

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Variable entered</th>
<th>Step</th>
<th>( r )</th>
<th>( R^2 )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>MOT</td>
<td>1</td>
<td>.264</td>
<td>.070</td>
<td>2.9*</td>
</tr>
<tr>
<td></td>
<td>FI</td>
<td>2</td>
<td>-.347</td>
<td>.161</td>
<td>4.1**</td>
</tr>
<tr>
<td>CATT</td>
<td>MOT</td>
<td>1</td>
<td>.553</td>
<td>.306</td>
<td>17.2***</td>
</tr>
<tr>
<td></td>
<td>FI</td>
<td>2</td>
<td>-.447</td>
<td>.427</td>
<td>8.1**</td>
</tr>
<tr>
<td>Time</td>
<td>FI</td>
<td>1</td>
<td>-.347</td>
<td>.120</td>
<td>5.3**</td>
</tr>
<tr>
<td></td>
<td>MOT</td>
<td>2</td>
<td>.264</td>
<td>.161</td>
<td>1.8</td>
</tr>
<tr>
<td>CATT</td>
<td>FI</td>
<td>1</td>
<td>-.447</td>
<td>.200</td>
<td>9.8**</td>
</tr>
<tr>
<td></td>
<td>MOT</td>
<td>2</td>
<td>.552</td>
<td>.427</td>
<td>14.2***</td>
</tr>
</tbody>
</table>

Note:  
- FI = field independence; MOT = motivational intensity; Time = time spent using CALL; CATT = attitude toward CALL.
- *  \( p < .10 \)
- **  \( p < .01 \)
- ***  \( p < .001 \)
organize relevant language details, they may have been bored. Perhaps these qualities of the ESL PLATO lessons were unattractive to FI students.

In contrast, students with little FI may have liked being provided with a fixed set of exercises to work through. These students tend to rely on others to formulate objectives and point out important points, a role played by the PLATO lessons.

**CALL as a Predictor of Second Language Success**

The second question was whether those students who used CALL more would receive higher scores on the end-of-semester English tests than those who spent little time using CALL. If the significant negative correlations between time and end of semester scores, presented in Table 3, are seen as the answer to the CAI effectiveness question, then those students who spent the most time using CALL were those who did poorly on the end-of-semester tests. (See Chapelle & Roberts, in press, for a discussion of the negative correlations between motivational intensity and the language measures.) Before drawing that conclusion, however, it is necessary to consider simultaneously the other variables related to end-of-semester ESL proficiency.

**TABLE 3**

<table>
<thead>
<tr>
<th>Nonlanguage measures</th>
<th>FI</th>
<th>AT</th>
<th>MOT</th>
<th>ANX</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-of-semester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>language measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFL 2</td>
<td>.750***</td>
<td>.237*</td>
<td>-465***</td>
<td>-303**</td>
<td>-481***</td>
</tr>
<tr>
<td>CC 2</td>
<td>.539***</td>
<td>.062</td>
<td>-529***</td>
<td>-180</td>
<td>-336**</td>
</tr>
</tbody>
</table>

Note: FI = field independence; AT = ambiguity tolerance; MOT = motivational intensity; ANX = English-class anxiety; Time = time spent using CALL; TFL 2 = end-of-semester TOEFL; CC 2 = end-of-semester speaking test.

Several other factors must be added to predict improvement. First, because end-of-semester test scores alone do not represent the differences in progress made by students throughout the semester, beginning-of-semester English scores must also be taken into account. Second, use of the PLATO lessons cannot be considered
as a sole predictor of success because many factors come into play in L2 acquisition, among which are the affective/cognitive factors measured in this study. Thus, the question of CALL effectiveness must be posed as follows: Does time spent using CALL predict variance in end-of-semester English proficiency beyond what can be predicted by beginning-of-semester English proficiency and affective/cognitive characteristics?

To answer this question, a multiple regression analysis was performed using end-of-semester scores on the language tests as the dependent variables (see Table 4). The first variable entered into the equation was the corresponding beginning-of-semester score. Of course, the beginning-of-semester score was a significant predictor of the corresponding end-of-semester score; that is, those students who did well on the language tests at the beginning of the semester tended to be those who did well at the end of the semester. Entering the cognitive/affective variables accounted for an additional portion of the variance in end-of-semester scores. Specifically, on the TOEFL, FI and AT were found to be significant predictors of success; on the test of oral communication, FI and MOT were significant predictors. Time spent using CALL was added to the equation last to determine if this variable could account for additional variance. Time spent using CALL was not a significant predictor—either positive or negative—of end-of-semester performance on the language measures after other relevant variables had been entered.

CONCLUSIONS
Learners and Lessons

The fact that FI students tended not to like to use the CALL lessons on PLATO raises the question of what kind of instruction they might like better. As suggested, these students may prefer to use their natural abilities to structure information rather than to be presented with lessons which define the course of their learning—a suggestion consistent with the FI individual defined by Witkin et al. (1977). However, it is necessary to ask not only what kind of instruction FI students might like but also what kind of lessons they might benefit from.

There is some evidence indicating that learners are more successful when the method employed in a particular learning activity matches their cognitive style. For example, in a series of experiments (Pask, 1976) in which students were classified by cognitive type as either holist or serialist, the results showed that instruction matched to the learner’s style favors learning and that
"mismatched instruction completely disrupts it . . . and leads to specific types of misconceptions" (p. 138). In another study (Zampogna, Gentile, Papalia, & Silber, 1976), students’ conceptual level was significantly predictive of their preference and need for structure in their L2 learning environment.

When these considerations are added to the fact that cognitive/affective characteristics influence success in L2 acquisition, it is clear that there is a need for individualized instruction for students who are at a disadvantage in a typical L2 situation. Off-line activities using such an approach have been described in great detail (Birckbichler & Omaggio, 1980) for students who are, for example, too impulsive, field dependent, or intolerant of ambiguity. The purpose of such an approach is to provide students with remedial tasks that address not only the content area in which they are having problems but also the cognitive strategies that they do not naturally employ. These possibilities for individualized instruction might be greatly enhanced through the use of interactive, on-line activities for students with special problems.

Though in some sense this application of research is premature, it points toward a possibly fruitful direction for CALL to explore. Current CALL is notoriously “insensitive” to individual learner differences (Hart, 1981), as a typical lesson presents all learners with

### TABLE 4
Multiple Regression Analyses

<table>
<thead>
<tr>
<th>End-of-semester language measure</th>
<th>Step</th>
<th>Variable entered</th>
<th>( r )</th>
<th>( R^2 )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFL 2</td>
<td>1</td>
<td>TFL 1</td>
<td>.931</td>
<td>.866</td>
<td>239.3***</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>FI</td>
<td>.737</td>
<td>.907</td>
<td>15.8***</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>AT</td>
<td>.225</td>
<td>.918</td>
<td>4.9**</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>ANX</td>
<td>-.289</td>
<td>.922</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Time</td>
<td>-.449</td>
<td>.925</td>
<td>1.4</td>
</tr>
<tr>
<td>CC 2</td>
<td>1</td>
<td>CC 1</td>
<td>.796</td>
<td>.634</td>
<td>64.0***</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>FI</td>
<td>.612</td>
<td>.708</td>
<td>9.0**</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>MOT</td>
<td>-.508</td>
<td>.736</td>
<td>3.8*</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>AT</td>
<td>.034</td>
<td>.740</td>
<td>.4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>ANX</td>
<td>-.218</td>
<td>.741</td>
<td>.0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Time</td>
<td>-.322</td>
<td>.748</td>
<td>.9</td>
</tr>
</tbody>
</table>

Note: The program omitted MOT from the TOEFL regression because the  \( F \) value was too low. FI = field independence; AT = ambiguity tolerance; MOT = motivational intensity; ANX = English-class anxiety; Time = time spent using CALL; TFL 1, TFL 2 = beginning-of-semester and end-of-semester TOEFL; CC 1, CC 2 = beginning-of-semester and end-of-semester speaking test.

*  \( p < .10 \)  **  \( p < .01 \)  ***  \( p < .001 \).
the same approach, albeit each at their own speed. To lay the groundwork for more sensitive lessons, the interaction of learning style and method of instruction must continue to be researched.

**CALL Effectiveness**

The research reported here casts a new light on the question of CALL effectiveness in the context of L2 acquisition. CALL cannot be evaluated without looking at the other student variables—some of which were assessed in this study—that are important in L2 acquisition. In a study of an intact group like the one reported here, it would have appeared that use of CALL predicted low ESL proficiency scores if other variables had not been considered (see Table 3). Consideration of FI, which was negatively correlated with time using CALL and positively correlated with ESL proficiency, rendered time spent using CALL nonsignificant (Table 4). Relevant student variables must also be taken into account in a control/treatment design assessing use of CALL versus no use of CALL. In this type of experiment, unintentional placement of FI students, for example, in one of the groups would cause the results to be distorted.

Clearly, CALL effectiveness cannot be looked at as though CALL represented one form of instruction and all students were in need of that kind of instruction. Instead, effectiveness must be analyzed in terms of the effects of defined types of lessons on students with particular cognitive/affective characteristics and needs. To do this, it is necessary to assess the characteristics of students and analyze the approach taken in a particular lesson or series. Through this thoughtful observation of students and approaches, progress can be made toward successful matching of students and lessons.

This is not a new idea; instead, these results emphasize the importance of the cognitive approach in educational research, as defined by Wittrock (1979):

> It is more useful and meaningful to study, for example, how [approach] influences the attention, motivation and understanding, which in turn influence behavior, than it is useful and meaningful to study how [approach] directly influences student behavior. From this point of view, the art of instruction begins with an understanding and a diagnosis of the cognitive processes and aptitudes of the learners. (p. 5)

We have not yet scratched the surface of what CALL can provide in terms of individual instruction for language learners. Researchers and educators must continue to describe the strategies used by good language learners and to assess cognitive/affective characteristics that are important in L2 acquisition. In this way, our understanding...
of L2 acquisition can be reflected in the intelligent use of computerized lessons and ultimately in the development of more “intelligent” lessons.

THE AUTHORS

Carol Chapelle, Assistant Professor of ESL at Iowa State University in Ames, is working on an ESL courseware development project.

Joan Jamieson, Teaching Associate in the Intensive English Institute at the University of Illinois at Urbana-Champaign, coordinates the use of PLATO for ESL students.

REFERENCES


The Effects of Referential Questions on ESL Classroom Discourse

CYNTHIA A. BROCK
Houston Community College

In their examination of ESL teachers’ questions in the classroom, Long and Sato (1983) found that teachers ask significantly more display questions, which request information already known by the questioner, than referential questions. The main purpose of the study reported in this article was to determine if higher frequencies of referential questions have an effect on adult ESL classroom discourse. Four experienced ESL teachers and 24 non-native speakers (NNSs) participated. Two of the teachers were provided with training in incorporating referential questions into classroom activity; the other 2 were not provided with training. Each of the 4 teachers taught the same reading and vocabulary lesson to a group of 6 NNSs. The treatment-group teachers asked significantly more referential questions than did the control-group teachers. Student responses in the treatment-group classes were significantly longer and more syntactically complex and contained greater numbers of connective.

An abundance of questions is a hallmark of second language learners’ exposure to the target language. In informal conversation between native speakers (NSs) and beginning-level nonnative speakers (NNSs), questions are the form most frequently used by NSS to initiate topics and, as a consequence of frequent shifts in topic, the dominant form used to address NNSs (Long, 1981, 1983). NSs’ preference for questions in topic initiation in informal conversation may be due to the obligation to respond which questions generate, the assistance they provide to the NNS in the form of partially or fully preformulated responses, and the salience added by such linguistic features as rising intonation and *wh*-words (Long, 1981). In view of the observation that in many Third World societies, “conversation . . . is the context known to be capable of producing fluent sequential bilingual” (Long, 1982, p. 215), questions may be a crucial input feature fostering development of second language abilities.
Despite the growing interest in classroom processes (Long, 1980) and the apparent pervasiveness of questions in ESL classroom discourse, only two studies have examined the use of questions in ESL classrooms, and only one of these (Long & Sato, 1983) looked at the forms and functions of ESL teachers’ questions in the classroom. (The other study, White & Lightbown, 1984, counted 427 questions asked by an ESL teacher in a single 50-minute class.)

Analyzing the classroom speech of 6 teachers, as well as the speech of 36 NSs in informal conversations with NNSs, Long and Sato (1983) found significant differences in the relative proportions of two types of questions asked in the two settings. Display questions ask the respondent to provide, or to display knowledge of, information already known by the questioner, while referential questions request information not known by the questioner. Although questions predominated in both settings, the ESL teachers asked significantly more display than referential questions in the classroom. The NSs in the informal conversational setting, on the other hand, asked a majority (76%) of referential and virtually no display questions.

In contrast to the lack of studies of question types in the ESL classroom, there is a substantial body of literature about the kinds of questions teachers ask in the first language classroom. Together, several studies provide data on at least three major issues of relevance to the study reported here: (a) the intellectual level of teachers’ questions, (b) the degree to which teachers can be trained to change the types of questions they ask, and (c) the relationship between the types of questions teachers ask and certain features of their students’ responses.

The intellectual or cognitive level of questions is defined, in most first language studies, according to either Bloom’s (1956) or Gallagher and Aschner’s (1963) hierarchies. Both systems view the intellectual level of questions as ranging from those calling for the recognition or recall of factual information, which are at the lowest level of the hierarchy, to those calling for evaluation or judgment, which are at the highest. One can reasonably assume that questions at low cognitive levels, asking for factual recall or recognition, are display questions, while questions calling for evaluation or judgment are likely to be referential questions.

While research results indicate that the preponderance of teachers’ questions are at low cognitive levels, primarily at the level of factual recall or recognition (Davis & Tinsley, 1967; Gallagher, 1965; Guszak, 1967; Willson, 1973), there is evidence that they can, with training, increase the frequency in their classroom speech of questions at higher cognitive levels (Arnold, Atwood, & Rogers,
While little research has been done on the relationship between the level of the teacher’s question and features of students’ responses, results suggest that, by and large, the level of a question affects what the student says in response. For example, Willson (1973) found that an increase in the mean cognitive level of questions asked by teachers was accompanied by an increase in the mean level of students’ responses. Although mean levels of questions and responses may imply a match between particular questions and the responses to them that does not exist (Mills, Rice, Berliner, & Rosseau, 1980), Arnold et al. (1974) did find a significant one-to-one correspondence between the question level and the level of student response.

The results of other studies suggest that responses to questions calling for the recognition or recall of factual information are shorter than responses to higher-order questions calling for interpretation or opinion (Dillon, 1981; Smith, 1978). A study conducted by Cole and Williams (1973) indicated a strong positive relationship between the cognitive level of the teacher’s questions and the cognitive level, length, and syntactic complexity of the pupil’s response.

The systems used to describe the intellectual level of questions in the studies referred to above do not employ the distinction between display questions and referential questions, despite the fact that, as Mehan (1979) observes, “the use of known information questions has consequences for the knowledge that children display in the classroom” (p. 291). Mehan further observes that the use of known-information questions, which reflect the one-way flow of information from teachers to students found in most classrooms, is responsible for the fact “that conversations in classrooms have unique features, and that the demands of classroom discourse must be kept separate from the demands of everyday discourse” (p. 294).

That the use of known-information, or display, questions in the classroom generates discourse which is fundamentally different from everyday discourse is an important consideration for language teachers. An increased use by teachers of referential questions, which create a flow of information from students to teachers, may generate discourse which more nearly resembles the normal conversation learners experience outside of the classroom.

**RESEARCH QUESTIONS**

The main purpose of this study was to determine if using higher frequencies of referential questions has an effect on adult ESL
classroom discourse. It was first necessary to determine whether, with coaching, the number of referential questions asked by teachers could in fact be increased. It was hypothesized that teachers receiving a training session in the formation and use of referential questions would ask more referential questions in the classroom than teachers who did not.

If the number of referential questions asked by teachers could be increased, this increase was expected to have the following effects on classroom discourse:

1. NNSs’ responses to display questions would be shorter and syntactically less complex than their responses to referential questions.
2. A greater number of referential questions would be accompanied by a greater number of confirmation checks and clarification requests by the teacher.
3. Confirmation checks and clarification requests by the teacher would occur more frequently following referential questions than following display questions.

As Mehan (1979) observes, the use of display questions generates a variety of discourse unique to the classroom. One of its peculiarities is that because there is often only a single correct response to known information questions, and this answer is known in advance of the questions, teachers often find themselves “searching” for that answer, while students provide various “trial” responses which are in search of validation as the correct answer. (p. 291)

A consequence of interaction organized in this way may be that the teacher, who knows the answer, also provides the propositional structure into which the answer fits. In other words, the teacher may be in charge not only of the answers to the questions but also of establishing their linear coherence.

Referential questions, on the other hand, may require that a student provide, in addition to information not already possessed by the teacher, the connections between the propositions expressing that information, connections which are necessary to form linearly coherent sequences (van Dijk, 1977a). Since these “connections between propositions are typically expressed by natural connective such as and, because, yet, so, etc.” (van Dijk, 1977b, p. 5), it was hypothesized that a greater number of referential questions would be accompanied by a greater number of connective in learner speech.
METHOD

Subjects

The subjects for this study included 24 NNSs enrolled in classes in the University of Hawaii’s English Language Institute (ELI). One of the subjects was from Afghanistan; the other 23 were from East Asian countries: Korea, China, Taiwan, Cambodia, Vietnam, and Japan. Sixteen were enrolled in the ELI’s most advanced courses, for which students’ TOEFL scores typically average between 470 and 520.

Also serving as subjects for the study were 4 ESL teachers, 2 females and 2 males, all with at least 5 years of ESL teaching experience and all enrolled in the master’s program in ESL at the University of Hawaii.

Design

Four groups of 6 NNSs each were formed using a randomized block design to control for the differences in proficiency among subjects. The 4 ESL teachers were assigned to a treatment or a control group, again using a randomized block to control for gender. Each teacher was randomly assigned one of the groups of 6 students for a single class period of 40 minutes. None of the teachers was acquainted with the students before the class.

Procedures

Two separate meetings were held prior to the class: one with the teachers in the treatment group and a second with the teachers in the control group. Both groups were introduced to the reading passage to be used as the basis for a 40-minute reading and vocabulary lesson, which would be tape-recorded. The passage (DeGracia, 1983) describes the special cultural traits and habits a nurse can expect to encounter in Filipino patients.

No special instructions on the lesson’s presentation were given to the 2 teachers in the control group. They were given a list of vocabulary items taken from the passage and instructed to allow students the first 20 minutes of the period for reading. The second 20 minutes was to be spent in a discussion balanced, as the teachers thought appropriate, between the reading passage and the vocabulary items. The only stipulation was that there be interaction between the teachers and the students. The control-group teachers were told that the purpose of the study was to examine an unnamed aspect of classroom language.

The 2 teachers in the treatment group were given the same
instructions regarding the division of time in class and the balance between discussion of the reading passage and vocabulary. In addition, however, these teachers were given a 20-minute training session introducing the distinction between display and referential questions. They discussed the distinction and briefly practiced forming referential questions.

The treatment-group teachers were also given a list of vocabulary items which contained the same items as that of the control group as well as a sample referential question for each item. They were told, however, that these questions were provided only as illustrations and that they were not expected to use them during the lesson. Finally, these teachers were informed that the purpose of the study was to investigate the effect on classroom language of an increase in the number of referential questions asked by the teacher.

Analysis

Long and Sato’s (1983) adaptation of Kearsley’s (1976) taxonomy was used to categorize question types. An exemplary referential question from the study data is, Do any of you have Filipino friends? An exemplary display question is, What does temperament mean? The total number of referential questions asked by the teachers in the control group was compared with the total number asked by the teachers in the treatment group.

Mean lengths (in words) of subjects’ responses to display questions and to referential questions were calculated. For the purpose of this study, the response was considered as only that turn immediately following (and responding to) the teacher’s turn containing the question; once the teacher spoke again or another student spoke, the response was considered to have ended.

Syntactic complexity was determined by measuring the mean number of sentence-nodes (s-nodes) per communication unit. Loban (1963) defined communication units (c-units) as grammatical independent predications] or . . . answers to questions which lack only the repetition of the question elements to satisfy the criterion of independent predication. . . . “Yes” can be admitted as a whole unit of communication when it is an answer to a question such as “Have you ever been sick?” (pp. 6-7)

In this study, a segment of NNS speech was not disqualified as a c-unit because it lacked or included incorrectly the copula, the impersonal pronoun it, an auxiliary verb, prepositions, articles, or inflectional morphology. Following Freed (1978), a c-unit “may have several sentence nodes as a consequence of having several sentences, several clauses or being a run-on or compound sentence”
Infinitives and gerunds, then, as well as tensed verbs, were taken to signal an underlying s-node. Modals were not considered to be signals of underlying s-nodes.

The transcripts were coded for confirmation checks and clarification requests following the definitions in Long and Sato (1983). The total number of logical connectors used by learners in the two control-group classes was compared with the total number used by learners in the two treatment-group classes. For the purposes of this study, a word was considered a logical connector if it appeared in the extensive list compiled by Celce-Murcia and Larsen-Freeman (1983, pp. 324-329), and only those initiating a clause were counted.

A random sample from the data, containing 119 questions, was coded by a second rater for display and referential questions, confirmation checks, and clarification requests. The simple percentage nominal agreement for these five categories was .91.

In all hypothesis testing, the acceptable level of probability was set at .05.

RESULTS

The 2 control-group teachers asked a total of 141 questions, only 24 of which were referential and 117 of which were display. The treatment-group teachers, on the other hand, asked a total of 194 questions, 173 of which were referential and only 21 of which were display (see Table 1). Since the treatment-group teachers asked approximately 1.38 times as many total questions as the control-group teachers, the number of referential questions asked by the control-group teachers was weighted by a factor of 1.38. With this weighting for the unequal number of questions asked, the control-group teachers asked 33.12 referential questions. As predicted, the number of display questions asked by the control-group teachers was significantly lower than that of the treatment-group teachers.

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of referential questions asked</th>
<th>No. of display questions asked</th>
<th>Total no. of questions asked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>24</td>
<td>117</td>
<td>141</td>
</tr>
<tr>
<td>Treatment</td>
<td>173</td>
<td>21</td>
<td>194</td>
</tr>
</tbody>
</table>

\[ \chi^2(1) = 172.52, p < .001 \]
teachers who were trained in the formation of referential questions asked significantly more of them than the teachers who were not: $\chi^2(1)=93.58, \ p<.001$.

The mean length of all learner turns which were responses to referential questions was 10.00 words; the mean length of learner responses to display questions was 4.23 words. As hypothesized, this difference was significant: $t(221)=3.92, \ p<.0005$.

The mean number of s-nodes per communication unit in learner turns which were responses to referential questions was 1.19, while the mean number of s-nodes per communication unit in turns responding to display questions was 0.56. As hypothesized, this difference was significant: $t(227)=4.50, \ p<.0005$.

The total number of confirmation checks made by the 2 control-group teachers was 13; the total for the treatment-group teachers was 21. Since the treatment-group teachers took 244 turns, or about 1.73 times as many turns as the control-group teachers, who took 141 turns, the number of confirmation checks made by the control-group teachers was weighted by a factor of 1.73. With this weighting for the unequal number of turns taken, the control-group teachers made 22.49 confirmation checks. This was a slightly higher number of confirmation checks than the number made by the treatment-group teachers, but the difference between the two groups was not statistically significant: $\chi^2(1)=.005, \ n.s.$

A total of 14 confirmation checks were made by the teachers in their turns immediately following learner responses to display questions and 11 immediately following responses to referential questions. Since there were only 102 responses to display questions and 121 to referential questions, the number of confirmation checks following learner responses to display questions was weighted by a factor of 1.19. This yielded an adjusted frequency of 16.66 confirmation checks following display questions. However, the difference between the number of confirmation checks for the two types of questions was not statistically significant: $\chi^2(1)=.78, \ n.s.$

The total number of clarification requests made by the 2 control-group teachers and by the 2 treatment-group teachers was the same: 5. Again, the raw frequency for the control group was weighted by a factor of 1.73 to correct for the unequal number of turns taken by the teachers in the two groups. The adjusted frequency of 8.65 clarification requests made by the control-group

---

1 For all one-way chi-square tests, with one degree of freedom, the correction for continuity was used.

2 The total number of turns taken by the control-group teachers is distinct from the total number of display and referential questions asked by those teachers. The occurrence of the value of 141 for both measures is coincidental.
teachers was greater than the frequency for the treatment-group teachers, but the difference was not statistically significant: \( \chi^2(1) = .52, \) n.s.

Frequencies of clarification requests made by the teachers in their turns immediately following learner responses to display questions and in their turns immediately following learner responses to referential questions were too small to analyze statistically.

Learners in the treatment-group classes used a total of 71 logical connectors in all their turns during the lessons, and learners in the control-group classes 11. However, there was a significant difference—\( \chi^2(1) = 18.62, p < .001 \)—between the total number of turns taken by learners in the control-group classes (155) and those taken by learners in the treatment-group classes (242). To adjust for the unequal number of turns taken by learners in the two groups, the number of logical connectors in the control group was weighted by a factor of 1.56, resulting in an adjusted number of 17.16. As predicted, the learners in the treatment group used a significantly greater number of logical connectors: \( \chi^2(1) = 31.67, p < .001 \).

DISCUSSION

As predicted, the 2 teachers who received training were able to increase the number of referential questions they used in the classroom. The differences in the language produced by learners in response to the two question types were pronounced. Learners’ responses to referential questions were on average more than twice as long and more than twice as syntactically complex as their responses to display questions. In the two treatment classes, learners used a far greater number of connective to make explicit the links between the propositions they expressed. They also took a significantly greater number of speaking turns.

That referential questions may increase the amount of speaking learners do in the classroom is relevant to at least one current view of second language acquisition (SLA). Swain (1983), in reporting the results of a study of the acquisition of French by Canadian children in elementary school immersion classrooms, argues that output may be an important factor in successful SLA. One function she suggests output may have is to create the necessity for the learner to perform a syntactic analysis of the language. She notes that through attention to vocabulary and extralinguistic information, “it is possible to comprehend input—to get the message” (p. 249) without such an analysis. Producing one’s own messages in the target language, on the other hand, “may be the trigger that forces the learner to pay attention to the means of expression.
needed in order to successfully convey his or her intended meaning” (p. 249).

If it is true, as the study reported here suggests, that the use of referential questions increases the amount of learner output, then such questions may be an important tool in the language classroom, especially in those contexts in which the classroom provides learners their only opportunity to produce the target language.

The use of a far greater number of logical connectors by learners in the treatment classes may also have important implications. Since logical connectors are those global elements (Burt & Kiparsky, 1974) that express relationships between propositions, their effective use may be crucial to a NNS’s ability to communicate successfully. Tomiyana (1980) found that for written communication, mistakes in the use of connective linking clauses within sentences were more likely to cause breakdowns in communication than mistakes in the use of articles. If, as seems likely, connective are equally important to oral communication, then it may be useful to know that posing referential questions provides increased practice in their use.

The predicted alterations in the interaction between the teachers and the learners may not have occurred because of the generally high level of proficiency of the learners involved: There might have been more instances of unintelligible speech necessitating confirmation and clarification with students of lower proficiency.

Perhaps the most serious limitation of this pilot study was the very small number of teachers involved. Further research is necessary to investigate the effects of group size and proficiency level and to determine to what extent the effects of training persist in teachers’ questioning patterns.

While further research is necessary for verification, the results reported in this study suggest that the use of an easily implemented, cost-free teaching technique may effect substantial changes in the amount and kind of practice ESL students obtain in the classroom.

ACKNOWLEDGMENTS

This article is based on an M.A. thesis completed at the University of Hawaii at Manoa (Brock, 1984). I would like to thank the members of my thesis committee, Michael Long, the chairman, Craig Chaudron, and Richard Day, for the guidance and assistance which made this project possible. I would also like to thank the teachers and students who served as subjects and Carla Deicke and Kathryn Rulon, who provided invaluable assistance and support.
THE AUTHOR

Cynthia Brock teaches ESL at Houston Community College. She received her M.A. in ESL from the University of Hawaii at Manoa, where she was a research assistant at the Center for Second Language Classroom Research, and was an English Teaching Fellow at the Instituto Guatemalteco Americano in Guatemala.

REFERENCES


**TESOL**

Announces

SEVENTH EDITION

DIRECTORY OF
PROFESSIONAL
PREPARATION
PROGRAMS IN TESOL
IN THE UNITED
STATES: 1986–88

It lists and includes:
U.S. institutions with brief descriptions of programs leading to certificates or degrees in TESOL for 1986 through 1988
Statement of Core Standards for Language and Professional Preparation Programs
Guidelines for the Certification of Teachers of English to Speakers of Other Languages in the United States
State Certification Requirements in ESL/EFL for all fifty states
$13.00 Members, $15.00 Non-Members

1118 22nd. Street, NW, Suite 205, Washington, D.C. 20037
COMICS AND CONVERSATION
Using Humor to Elicit Conversation and Develop Vocabulary
Joan Ashkenas

For secondary and adult foreign language students.

These witty, uncaptioned COMICS with pictorial punch lines beg to be described. Highly adaptable to a variety of more structured lessons, these amusing cartoons are ideal for informal "pairing off" of students, one telling the story to a partner, with teacher assistance.

Procedures for use are suggested in this volume of twenty-two cartoons of varying length, detail and complexity. Perfect for multi-level classes. A joy to the substitute teacher.

Price: $12.95
Add 6½% sales tax; $1.00 postage and handling for up to 4 copies, $1.40 for 5 copies or more.

FOREIGN STUDENTS' GUIDE TO PRONUNCIATION
Using the Pronunciation Key of the English Dictionary
Joan Ashkenas

For adult ESL and EFL students of level 2 and above.

This additional and rewarding approach to pronunciation should be standard in the curriculum. It makes the pronunciation key of any ordinary English dictionary understandable to foreign students, focusing on teaching pronunciation of common example words, diacritical marks and respelling. Once you have presented this material you will find yourself referring to it ever afterward throughout the semester, as a tool for confirming precise pronunciation. With it, students gain insight into the variety of English speech sounds.

Affordable even by students themselves, this intentionally small-format lesson fits into their personal dictionaries to use independently after instruction.

Students' sets of 25 with one Teacher's Guide, $13.00
Separate students' sets of 10, $5.00 Sample available upon request
Add 6½% sales tax, $1.00 for postage and handling for up to four class sets; $1.40 for five class sets or more.

Send checks or money orders to:
JAG Publications, 11288 Ventura Blvd. #B-301, Studio City, CA 91604
This article reports on a study done to determine how nonnative English speakers studying in U.S. colleges and universities perceive their language learning experiences and how they use English in academic settings. Open-ended interviews, using a structured set of topics, were conducted with 80 students. Areas investigated included the value of the U.S. language training program, how the program addressed specific skill areas, how out-of-class experience contributed to language learning, what teacher qualities were valued, and how English was used in the academic setting. In general, students supported the design of most intensive ESL training, but they raised questions about some skill-area emphasis. A strong desire for more interactive instruction was expressed as well as an appreciation for personality, rather than technical, qualities of teachers. Students indicated the importance in academic work of the receptive skills of reading and listening over the productive skills of speaking and writing.

Curriculum design in ESL programs for academic preparation has, in general, failed to use the experience of students themselves as a basis for planning and decision making. This article reports on a study that attempted to discover what students believed contributed most to their language learning. The subjects had studied in intensive ESL programs in the United States and were engaged in academic study at the time of the investigation. Student attitude toward teachers and teacher behavior was also studied. Last, the study attempted to determine how these students were using English in their academic work.

OTHER STUDIES

A number of studies have been done to determine the attitudes or opinions of language students or to discover the patterns of
language use by second language speakers in academic settings. Kroll (1970) surveyed native and nonnative students who were engaged in academic study and asked them to rank a list of writing activities according to how frequently the students had to use them. The traditional personal essay did not rank as important for present, past, or future needs as did business letters of request and persuasion and reports, both survey and technical. Johns (1981) questioned academic faculty on which skills (reading, writing, speaking, listening) were most essential for nonnative speakers in the classes taught by those faculty members. They ranked the receptive skills of reading and listening as most essential for both lower and upper division classes.

Ostler (1980), surveying students who were studying in an ESL program, attempted to determine what skills ESL programs should address. Her study focused attention on what need the students believed they would have for specialized skills such as reading academic journals and papers and writing critiques and research papers. Reading texts and taking notes were ranked the highest in skills needed. Graduate students reported a greater need than undergrads for certain skills, such as writing formal papers and giving talks. All students expressed greater confidence in limited and predictable communicative encounters, such as with waiters and clerks, than in more “creative” encounters with friends and professors.

In a detailed survey of the language use patterns of nonnative speakers studying at the University of Illinois, Robertson (1982) reported quite different patterns of use among students in different disciplines but gave no overall ranking of various types of use.

Bridgeman and Carlson (1983) surveyed 190 academic departments in 34 universities to ascertain faculty views on how important writing is to academic success, what types of writing are most important in different disciplines, how faculty evaluate student writing in ESL, and how ESL and native-speaker writing differ. They concluded that faculty believed that writing was important to academic success but more important to future professional success. In the study, writing was not evaluated relative to other aspects of language use. Types of writing reportedly assigned varied widely across disciplines and academic level. The short research paper and summaries of written material were the most frequently reported, but even they were not universal, nor were written essay exams. In general, faculty felt that they evaluated ESL writing more on content than on form and that ESL and native-speaker writing differed more on sentence-level features than on organizational and other discourse-level features.
Jones, Matthews, and Rodby (1981) investigated essay-exam writing tasks presented to students in various disciplines at three different universities. After examining the actual assignments given to students, they concluded that the majority were ill-defined and did not fit into any of the traditional rhetorical types used by writing instructors to categorize academic writing.

Regarding student attitude toward language study, Horwitz (in press) has developed a student questionnaire to determine what beliefs ESL students have about language learning and has related her results to different language learning strategies.

In a questionnaire survey of 711 students in the academic intensive program at the University of Toronto, Yorio (1983) attempted to determine the strength and consistency of their beliefs about various elements of language teaching, including type of text material, teaching techniques, and skill areas. The respondents tended to support most of the learning activities on the questionnaire, including grammatical explanation. However, they did not support some activities infrequently used in their instructional program, such as translation and memorizing vocabulary. On the basis of the high number of definitive responses, Yorio concluded that students can be good sources of information on what should be included in language teaching programs and that curriculum design should take such information into account in a more systematic way.

The above studies are valuable because they provide some indication of the linguistic needs of nonnative students and an idea of what students and faculty believe is more or less valuable in learning a new language. All of these studies used some form of a questionnaire with predetermined response categories and surveyed students who were still engaged in language study.

In our opinion, these previous studies of student belief and language use suffer from problems of objectivity, sampling, and validity. Objectivity problems arise because of teacher bias and student bias. Knowledge about language and attitude toward it are subject to idealization and misconception. Linguistic, especially sociolinguistic, work over the past several decades has demonstrated that even native speakers have an unrealistic knowledge of their own language and that laymen have a limited set of concepts with which to discuss language. Much of what passes for knowledge about language is, in fact, biased and based on belief rather than fact.

Teachers have beliefs about how they want to teach and about what students need (Krahnke & Knowles, 1984). Many teachers tend to hear what students say about language learning through a
filter of personal belief. For example, the teacher who firmly believes in teaching formal grammar may hear and encourage student remarks that agree with that belief and fail to hear or seek out ones that do not. Students also have biases; they have personal and cultural expectations as to what language instruction should be. In addition, they may have a limited set of concepts for talking about language teaching and learning. Not least important, students may voice only what they think their teachers want to hear. All of this raises questions as to whether studies of learner need and preference that do not try to minimize bias can, in fact, be objective.

Sampling problems derive from the fact that the opinions reported may be those of only a few vocal students. If the range of student belief is to be accurately represented, the students who provide the opinions need to be representative of all types of students, those who freely offer opinions and those who do not.

Validity problems arise because the opinions of students and teachers about language needed in academic course work may be based on expectation or prior experience, not on current realities. Questionnaire techniques can be another source of validity problems because students may have different interpretations of categories used in the questionnaire. What a student means by “conversation” or “grammar” or “writing a paper” may be quite different from what the investigator means.

THE PRESENT STUDY

The study reported here attempted to determine (a) what types of experience former ESL students perceived as having contributed most to their language learning while they were in intensive language programs, (b) what qualities of teacher behavior former ESL students perceived as contributing most to their learning, and (c) what types of language use predominated for former ESL students in their academic work and what skills they regarded as easy or difficult for them.

Design

Because predetermined categories can be misunderstood by students, we were interested in what terms students themselves used to talk about language study. We did not want to give them factors to judge or rate or to leave them without alternatives against which to balance their opinions, since we felt that there is a tendency to rank almost everything as positive, especially
traditionally defined instructional factors. How students already engaged in academic work viewed their language needs and their previous language learning experience was also of interest, since the perspective of time and experience would make their opinions about previous language study more valuable.

To overcome problems of objectivity, sampling, and validity, this study was designed as a survey, using an individual interview technique. The objectivity problem was addressed by gathering data in a uniform way (using a standard interview schedule); by allowing students to respond to broad questions with their own terms and categories, which were clarified, when necessary, in follow-up discussion; by including a full range of views in the interview schedule; by interviewing students who had not had the interviewers as teachers; and by using two independent interviewers. The study addressed the sampling problem by interviewing a large number of students selected randomly from a variety of instructional programs. The validity problem was addressed by interviewing only students who had completed their language study and had been enrolled in full-time academic work for from one to four terms.

Subjects

The subjects surveyed in the study were 80 nonnative speakers of English who were, at the time of the interview, studying at five separate universities. They had all completed intensive English language programs and were enrolled in full-time academic study. Intensive was loosely defined as full-time language study and ranged from 4 to 6 hours a day. In some cases the intensive language programs were at the same university at which the students were doing their academic work; in other cases this was not so. All respondents had completed their intensive program within the past academic year and had studied a minimum of one term to a maximum of four terms in their intensive language programs. Twelve intensive programs, from 12 states in all parts of the United States, were represented.

Eight native language backgrounds were represented: Japanese (22%), Arabic (19%), Spanish (19%), French (12%), Chinese (9%), Thai (8%), Portuguese (6%), and Bengali (5%). Of the respondents, 72% were male and 28% were female. Ages ranged from 19 to 50 years with a mode of 19 and a mean of 24.2 (SD: 6.1). The length of time the respondents had been in the United States ranged from 6 to 32 months. The length of time they had spent studying English in the United States ranged from 2 to 12 months. Language study previous to study in the U.S. varied greatly and was not regarded as
significant in that the subject of investigation was attitudes toward intensive study in the U.S.

Undergraduates made up 71% of the sample and graduates 29%. The 21 major fields of study represented fell into the following groups: engineering (25%), business (25%), sciences (14%), social sciences (2.5%), mathematics (2.5%), humanities (12.5%), computer science (12.5%), and general education—those who had not chosen majors (6%).

In regard to relevant characteristics (age, gender, field of study, graduate/undergraduate), the sample used in this study roughly approximates the overall international student population in higher education in the United States (Boyan, 1983; Zikopolos & Barber, 1984).

Procedure

Students were interviewed using a structured questionnaire (see Appendix) containing questions about presumed topics of interest such as preference for language activities, what activities contributed most to language development, basis for instructor preference, how English is used in academic and social settings, and the contributions of out-of-class experiences to language improvement. Respondents were not presented with lists of choices but were encouraged to respond freely with their own terms and opinions.

During the interview, subjects were asked open-ended questions such as, “Which was your most difficult ESL class?” After giving an initial response, they were then encouraged to discuss the topic further, explore alternatives, and think of possibilities they had not considered. The interviewers explored terms used by the students to determine their meaning more precisely. For example, if a student, in discussing teachers, used the phrase “explains well,” the interviewer would ask a further question or questions attempting to determine what “explains well” meant to that student. Another example was, “In what ways do you use English the most for your academic classes?” After subjects gave their initial response, they were encouraged to comment on other aspects of language use in academic settings.

On teacher preference, subjects were asked questions such as, “Did you have a favorite teacher in your English program? If ‘yes,’ what did that teacher do to make you feel that way?” Responses to these questions were not directed toward mentions of specific teachers but toward the specific qualities the subjects preferred in teachers.

In questioning, subjects were not rushed to answer. They were
given as much time as possible to consider and present a response. The interviewers avoided presenting alternatives from which the subjects could choose. Every attempt was made to ascertain that the subjects understood the questions and were responding completely and candidly. When subjects’ answers did not adequately relate to the question, the interviewers used mild prompting to suggest more appropriate types of answers.

Precautions were taken to ensure that the interview information was accurate, complete, and consistent. All interviews were tape-recorded, and the information was later transcribed. Settings for the interviews varied, but only two interviewers were used. None of the respondents had had either of the interviewers as a teacher. Interviewers pursued any issues arising during the course of the interview which seemed to have a bearing on the overall concerns of the project.

The data were subjected to content analysis; that is, the subjects’ responses were evaluated to determine what specific factors they mentioned and which they regarded as most important, easy, or difficult. For example, the responses to questions on qualities of a good teacher included a wide variety of characteristics, from specific teaching techniques to general personality qualities. This information was grouped and ranked into the more general categories of Explains Well, Various Personality Characteristics, and Various Professional Characteristics.

When the information provided by subjects clearly fit into well-known categories (e.g., composition = writing), it was placed into the broader category. Some responses did not seem to fit into a well-known category (e.g., in response to a question about difficult academic uses of English, rather than mentioning reading textbooks, writing papers, or participating in seminars, one subject responded, “Figuring out what the professor wants on the tests”). Such responses were counted and categorized as miscellaneous. Almost all responses were, therefore, counted in some category. Finally, mentions in each category were totaled, giving a ranking of the relative importance of each specific type of response.

RESULTS

Obviously, in an interview, much information is expressed by individual subjects that may be of great interest but does not fit into the design of the study. The design of this study was intended to allow as much individual variation as possible in the data collection and then to reduce those data to a finite number of categories. Interpreting the results of such a process is, of course, less precise
than a purely objective survey in which subjects are required to respond in easily countable ways (yes/no, ranking 1 to 5, etc.). The following results consist of what can be reliably concluded about the subjects’ responses, based on the analytical procedures outlined above.

Table 1 shows students’ ranking of skill areas according to difficulty, interest, and whether the skill area should have been added or dropped. Skill areas included reading, listening, speaking, grammar, lab, and any other mentioned by the subjects. Skills involving written discourse ranked first in the area of difficulty. Speaking or conversation ranked first when students were asked what they would have liked to add to their intensive programs. Grammar study, which was considered the easiest and least interesting, was ranked first in classes subjects would have liked to drop.

| TABLE 1 |
| Language Skill-Area Ranking (in Percentages) |

<table>
<thead>
<tr>
<th>Most difficult:</th>
<th>Least interesting:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing</td>
<td>Grammar</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>Nothing</td>
</tr>
<tr>
<td>Listening</td>
<td>Listening lab</td>
</tr>
<tr>
<td></td>
<td>Miscellaneous</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Easiest:</th>
<th>Classes to add:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar</td>
<td>Speaking</td>
</tr>
<tr>
<td>Speaking</td>
<td>Listening</td>
</tr>
<tr>
<td>Listening</td>
<td>Miscellaneous</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most interesting:</th>
<th>Classes to drop:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking</td>
<td>Grammar</td>
</tr>
<tr>
<td>Reading</td>
<td>Nothing</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>Writing</td>
</tr>
<tr>
<td>Writing</td>
<td>Miscellaneous</td>
</tr>
</tbody>
</table>

*Miscellaneous is the sum of categories that were mentioned by fewer than 5% of the respondents.

In Table 2, out-of-class experiences are ranked according to which experiences contributed the most to language development, and subjects’ reports of amount of English used outside of the classroom are given. Social contact with native speakers, such as conversations at parties and discussions with American classmates, ranked first in experiences which contributed the most to language improvement. Listening to radio and television ranked second. It is interesting to note that in Table 1, students ranked speaking and
listening as classes they would like to add. English was spoken out of class at least 1 hour a day by 68%, but 12% said they did not speak English out of class at all.

### TABLE 2
Out-of-Class Experience Rankings (in Percentages)

<table>
<thead>
<tr>
<th>Experiences contributing most to improvement:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Social contact with native speakers</td>
<td>55</td>
</tr>
<tr>
<td>TV and radio</td>
<td>23</td>
</tr>
<tr>
<td>Shopping</td>
<td>12</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use of English outside of class:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 2 hours per day</td>
<td>35</td>
</tr>
<tr>
<td>More than 2 hours</td>
<td>33</td>
</tr>
<tr>
<td>1 hour or less</td>
<td>20</td>
</tr>
<tr>
<td>None</td>
<td>12</td>
</tr>
</tbody>
</table>

Subjects’ perceptions about the length of time they had spent in ESL programs as well as their rankings of ways they believed English language skills could have been improved are presented in Table 3. While 40% felt their English program should have been longer, only 10% felt it should have been shorter. Half of the students determined that the time they had spent in intensive programs was about right. Interacting with native speakers was considered the best way to improve their language skills by 65% of the subjects. Other ways mentioned were speaking more in class and studying harder.

### TABLE 3
Ratings of ESL Program Factors (in Percentages)

<table>
<thead>
<tr>
<th>Time spent in ESL program:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Right length of time</td>
<td>50</td>
</tr>
<tr>
<td>More time needed</td>
<td>40</td>
</tr>
<tr>
<td>Less time needed</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What would have led to faster improvement:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>More interactions with native speakers</td>
<td>65</td>
</tr>
<tr>
<td>Miscellaneous*</td>
<td>15</td>
</tr>
<tr>
<td>Speaking more in class</td>
<td>15</td>
</tr>
<tr>
<td>Studying harder</td>
<td>5</td>
</tr>
</tbody>
</table>

* Miscellaneous is the sum of categories that were mentioned by fewer than 5% of the respondents.

STUDENT PERCEPTIONS OF ACADEMIC LANGUAGE STUDY 69
Table 4 presents rankings of each skill according to frequency of use in academic settings and according to difficulty in the same settings. Subjects said that 80% of their academically related language use was spent in reading and listening (the receptive skills), with only 20% spent in speaking and writing (the productive skills). The most difficult academic uses of English were speaking and listening to lectures in class. When asked what was the easiest, 29% said nothing was easy for them, 38% said reading, and 27% said listening.

It is interesting to note that about the same number felt that listening was the most difficult (32%) as felt it was the easiest (27%). This tells us either that there are differences in learning preference (some skills are easy for some and difficult for others) or that there is a wide difference in types of listening tasks. Also, most subjects who said listening was the easiest mentioned class instructions, directions, and discussion. They saw these tasks as distinct from listening to lectures for specific information they would be tested on later.

### TABLE 4
Use of Language Skills in Academic Settings (in Percentages)

<table>
<thead>
<tr>
<th>Frequency of skill use:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>50</td>
</tr>
<tr>
<td>Reading</td>
<td>30</td>
</tr>
<tr>
<td>Speaking</td>
<td>10</td>
</tr>
<tr>
<td>Writing</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Most difficult skill:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking</td>
<td>35</td>
</tr>
<tr>
<td>Listening</td>
<td>32</td>
</tr>
<tr>
<td>Miscellaneous*</td>
<td>24</td>
</tr>
<tr>
<td>Reading</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Easiest skill:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>38</td>
</tr>
<tr>
<td>Nothing</td>
<td>29</td>
</tr>
<tr>
<td>Listening</td>
<td>27</td>
</tr>
<tr>
<td>Writing</td>
<td>6</td>
</tr>
</tbody>
</table>

*Miscellaneous is the sum of categories that were mentioned by fewer than 5% of the respondents.

Rankings for teacher preference in ESL programs appear in Table 5. Some 97% said they had favorite teachers. The nature of the questions eliciting these responses and the nature of the responses made it clear that favorite meant effective. Rather than begin this line of questioning by asking about specific teacher characteristics, the subjects were first asked to identify a preferred teacher and then to provide the characteristics that made that teacher preferable.
When asked to provide the characteristics, 40% said the teachers explained well. Various personality factors, such as whether the teacher was patient, kind, interested, caring, cooperative, enjoyable, stimulating, and helpful, were mentioned by 35%. Professional characteristics such as organization, preparation, experience, clear speech, teaching style, and fairness in grading were cited by 25%. There were also many multiple responses; most who mentioned professional characteristics also specified personality factors. Some 83% felt their ESL teachers understood what they would need in their academic class.

**TABLE 5**

<table>
<thead>
<tr>
<th>Have favorite teachers:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
</tbody>
</table>

**Reasons for Preference:**

<table>
<thead>
<tr>
<th>Explain well</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Various personality characteristics(^a)</td>
<td>35</td>
</tr>
<tr>
<td>Various professional characteristics(^b)</td>
<td>25</td>
</tr>
</tbody>
</table>

**ESL teacher knowledge of academic needs:**

<table>
<thead>
<tr>
<th>Adequate</th>
<th>83</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not adequate</td>
<td>10</td>
</tr>
<tr>
<td>Don't know</td>
<td>7</td>
</tr>
</tbody>
</table>

**Note:** There were many multiple responses. Respondents who mentioned professional characteristics almost always added personality characteristics. Personality factors were therefore almost universally mentioned. Only the first response was considered in tabulating the results.

\(^a\) Stimulating, patient, interested, enjoyable, cooperative, caring, kind, helpful.

\(^b\) Clear speech, experience, teaching style, fairness in grading, preparation, organization.

**DISCUSSION**

The following discussion includes interpretations of the quantitative findings as well as conclusions drawn from points mentioned frequently by the subjects.

Structured interviews were found to be a valuable research tool for investigating questions of belief and opinion, especially cross-culturally and in settings where affect and mode of questioning can seriously interfere with reliable data collection. Many international students come from cultures where written modes of communication call for a very different type of sincerity and candidness than is sometimes displayed in our culture. The oral mode of data collection is, therefore, much more reliable with such subjects.
The interview also allows the researchers to explore a topic at greater length, to get behind stereotyped values and expected responses to more personal beliefs and opinions. By encouraging subjects to consider their responses at greater length and from different perspectives, a fuller and richer picture of the beliefs they really do operate on emerges.

On the other hand, interviews are time-consuming and, obviously, less objective than multiple-choice questionnaires. It may be that some students feel that face-to-face interaction is, in fact, more intimidating than an impersonal questionnaire and that answering questions about learning English in English introduces some bias itself. Since this question cannot be finally resolved, we urge the use of a variety of techniques to improve our knowledge of what factors most positively affect the learning process.

Students can be valuable and reliable sources of information about what we should and should not be doing in intensive programs. Many of our subjects were quite articulate and willing to discuss their experiences in an open and objective way.

According to their responses, intensive programs are not doing a bad job. Almost all students interviewed said that their programs were beneficial and that they had good feelings about those programs. Some 97% felt they improved as a result of their ESL experience.

Most subjects felt that intensive ESL programs provided a good general preparation for academic work, but the majority did not seem to think that instruction in specific skills, such as writing specific rhetorical types or narrowly defined reading skills, addressed their later needs. Few could cite specific language skills that they were presently using in their academic work. This is not to say that there is no benefit to such specific skill instruction, only that students do not perceive it. The following responses from students illustrate this point:

Essay was hard and thinking is hard and very different from Thai. I worry so much about this my first quarter, but now I don’t use [that essay form].

Reading about interesting subjects is good for me but I don’t think all the [reading skill assignments] were good for me. Like matching, we could do it but not, you know, know the words or anything.

Students generally felt that their programs had tried to prepare them for their out-of-class social and business needs and that this had been an important part of their learning. Many felt that this aspect of language learning should be expanded, as the following comment indicates:
Yes, we do much with shopping and party and things like that. That was good, because I don’t have time for learning that now. I only have time for study. Then [during language study] is when I have time for learning about that kind of language.

The overwhelming majority of subjects preferred an active, interactional approach to language learning, at least as a central or major component of the overall program. Regardless of the type of instruction or out-of-class experiences they had had, subjects regularly expressed the opinion that natural interaction with native speakers in class (65%) and out of class (55%) was the most valuable means of learning a language. Adding more speaking to the curriculum was a wish expressed by 70%. Preferences ranged from more conversation and speaking activities in classes to more organized or personally developed opportunities to interact outside of class. Along with this was a preference for realistic learning activities—listening to real lectures or having an opportunity to participate in actual academic class work were frequently mentioned. This point is expressed in the following student responses:

What you must do is talk more with native speakers about many things. The ideal class will have a mixture of Americans and students. If they can do something together, then they will learn. I learned most of my English outside of class, from my friends.

Many subjects expressed a view that can be interpreted as a preference for learning resources rather than rigidly designed instruction. As one subject put it, “Learners must do 60% of the work on their own. The teacher should just facilitate the learning.” This is related to the desire for realistic learning activities. Beneficial examples mentioned by the subjects included Science Research Associates reading materials, read at the student’s own pace, and lectures and other listening experiences over which the student had time control.

Most subjects found that the receptive skills of listening and reading (80%) were used far more in academic work than the productive skills of speaking and writing (20%). Many subjects mentioned an initial difficulty in comprehending lectures, but most said that their skills increased rapidly with real experience. Some blamed the artificiality and lack of variety in the speech heard in their English classes for this problem. Many compensated for their listening difficulties by relying more on reading for course content. The following student remarks illustrate these points:

The first semester I understand almost nothing from lecture. That is first time I heard English like that. Professors were very hard to understand.
Mostly I read the book. . . . After one year, now I can understand most of lecture. But I still write bad notes.

Teachers in English classes speak very clear. They want us to understand everything. That is good, but my teachers now not like that. I wish I had heard English like that before, like maybe in real lecture, or on tape.

Subjects rarely felt that professors based judgments of their academic performance on the quality of their spoken and written English. Most judgments, they felt, were made on the basis of content and ideas rather than form:

Professors never care about how I speak; they only listen for whether I know.
No, they never take off for English. If I get idea right, that is all they care.

However, many subjects expressed a reluctance to engage in verbal interaction with classmates or with professors, in or out of class, because they felt that their English was inadequate:

No, I don’t talk to professor in office. I afraid he not understand me.
I want to talk with American students, but they usually not friendly. Maybe my English.

ESL teachers were positively rated more on their ability to make something comprehensible and the personality traits supporting a positive affective atmosphere than on their technical abilities.

Patience and clear pronounce are most important.
Plenty of repetitions and clear speak.
Friendly and make things clear and easy to understand are most important.
She understood what we needed and very clear in speaking.
I like it when teachers smile, know names, prepare, and try to explain to everyone.

Most subjects felt that time spent in their program was adequate, which indicates that current academic entrance standards and ESL program placement and exit criteria are about right, according to the students’ perceptions.

CONCLUSIONS

What are the implications of this study for the intensive teaching of academically specialized language and for language teaching in general? Not surprisingly, several of the conclusions we might draw from the students’ remarks may seem somewhat contradictory at
That contradiction may be symptomatic, however, of conflicting but genuine underlying needs of students during the difficult process of learning and using a new language.

First, the evidence confirms that some kind of natural interaction using the language being learned is regarded as a major means to learning the language. In this study, the students clearly felt that interaction involving real tasks, especially with native speakers, was the primary contributor to their language ability, in or outside of the classroom. Though hardly new or surprising, this conclusion seems to be ignored by the large number of academic preparation programs that adopt a reductionist view of language and language behavior. Such programs teach as though conscious performance of limited skills and routines (formal writing, grammatical judgments, outlining), governed by accuracy measures, is all or most of what prospective university students need.

Many teachers do wish to make even their academically oriented classrooms more interactive, but in discussions of students’ need for more interaction, a number of these teachers frequently mentioned the difficulty of getting students to participate in such activities willingly and consistently. Students do not happily and automatically engage in the kind of activity they later deem valuable. It seems, then, that there is a contradiction between what students say they need and what they will do.

The contradiction between students’ belief about the value of interaction and their reluctance to engage in it is probably an accurate representation of the facts. In our opinion, it is not a matter for dismay or a reason to give in to student reluctance. It is, instead, a clear indication of the challenge involved in getting students to engage in what is certainly beneficial to them but what they often are quite resistant to engage in. Real interactive activities in language learning are ego-threatening and often have little immediate measurable or observable benefit. It is difficult to measure or identify the effect of a 50-minute problem-solving activity, but it is not so difficult to conceptualize the content of a lesson on dependent clauses or one on the main ideas of a reading selection. We would like to suggest that these facts, if they are facts, present a challenge to language teachers, a challenge to try to overcome the reluctance of students to engage in interactive work, especially with native speakers.

Teachers also speak of the difficulty of devising and implementing interactive activities. Aside from presenting the rationale that academically bound students would rather study for the TOEFL, teachers accurately point out that it is much easier to teach (and test) a grammar lesson or to go over reading comprehension
questions. But the difficulty does not alleviate the responsibility to make the effort. The task for language teachers is not simply to devise or use interactive activities, but to overcome student avoidance and discomfort when these activities are used. Much of the work we have done with teaching activities over the past years indicates that this reluctance can be overcome and that students are satisfied with the results when they do so. But the effort is somewhat greater than that required to teach a lesson in language form.

Second, this study suggests that the receptive skills of listening and reading may have greater importance than are usually attached to them. The students in our survey strongly indicated that they relied on these skills much more heavily than on productive skills. It may seem that there is a contradiction here between the importance of interaction and speaking, identified above, and the academic importance of reading and listening. However, there is not really a contradiction: The subjects clearly indicated that speaking and interaction were valuable for learning the language but that listening and reading were more important in helping them survive in the academic arena. Receptive skills have received less attention in instruction than have productive ones, probably because they are harder to measure and identify—how do we know if someone has made a mistake in comprehending a lecture? It is much easier to count spelling errors.

In academic settings, we certainly have a responsibility to teach to the most accurate and appropriate language skills we can. However, if students are not being called on to write accurately or to interact verbally in classroom settings but are instead being asked to read lengthy texts and listen to hours of lectures, then our best efforts may be off the mark. Some previous studies have also shown what we discovered, that our idealization of what second language students need in an academic setting may be quite different from what they actually do need (and we believe that the present study is a step in determining what those actual needs are). Balancing realistic and responsible evaluations of academic linguistic needs with the survival tactics that students may engage in, and the allowances that instructors frequently make, is a difficult matter. But it is not solved by ignoring the reality of academic life and shortchanging students by, for example, spending the majority of our instructional time trying to eliminate error in their spoken and written production while robbing them of valuable extensive reading experiences.

Third, this study provides more evidence for something we are becoming increasingly convinced of, that one of the most important
qualities in a language teacher is comprehensibility. The teacher who provides students with a rich, but understood language experience is at least perceived as the one who contributes most to language development. There is steadily accumulating evidence that classrooms in which students are attentive and active are ones in which the teacher uses the natural conversational techniques of repetition, restatement, and clarification to ensure that the maximum number of students in the class get the maximum understanding and exposure to the language being used (Hu-Pei Au & Mason, 1981; Pica & Doughty, in press).

This contradicts somewhat the claim of students that the language of the classroom is unnatural, that it is too comprehensible. Once again, both sides of this dilemma are probably true in their own way. To develop language ability, a high degree of clarity and comprehensibility are desirable. But to develop the learners’ ability to process natural speech outside of the classroom, some experiences, possibly with some initial support from instruction, are also desirable. Attending academic lectures, then listening more intensively to tapes of them and building comprehensibility would be one way to address this issue.

Fourth, our subjects reminded us regularly that they have or want lives outside the classroom, that they need and want to interact with Americans, and that they generally experience great difficulty in doing so. They felt that language teaching should help address this need, not only for its social benefits, but for the language learning experience that it would provide. Certainly, most of our intensive language teaching programs provide some “social” experiences—parties, field trips, films. But we often regard them as fluff, as something we do to keep the students happy, and we do not invest the energy in them that we do in our lesson plans. If we view such experiences as necessary to provide social contacts with native speakers and to provide intensive language experiences, our planning may be somewhat different. In one program we were involved with, for example, joint activities with a native-speaker speech communication class were regularly used to supplement classroom activity. Though difficult to arrange, the experience was an intensive and valuable one.

Fifth, we need to remind ourselves regularly that students can be valuable sources of information on the language learning, socialization, and academic preparation experience. They may not do this in the most direct way, and we may have to interpret individual statements in light of context, occasion, and who is saying what to whom. Many student remarks may be expressions of a need or desire for the security that Stevick (1982) speaks of. For example,
in our opinion, much of grammar instruction is just that, a need to feel understanding and knowledge. Such understanding is, of course, rare in the learners’ own language but is felt to be desirable in a new language. That is not surprising and, as such, gives us a rationale for including a limited amount of such instruction to meet the need for security. But we also have to take as meaningful the regular statements of a need for opportunity to use the language, and we must interpret these statements as challenges to provide what may be difficult and threatening, but still valuable. In sum, if the good teacher is a good communicator, then we have to pay attention to both ends of the communication channel and listen carefully to what the students have to say as well as to what we say to them.

In conclusion, the study reported here should be regarded as one more way to understand the language learning process. Along with Ostler (1980), we believe that sound curriculum design in ESL programs for academic preparation should be based on empirical data that reflect what is really useful to students and not only on the intuitions and experience of the teaching personnel. In combination with observation and analysis of interlanguage behavior, classroom activity, and teacher behavior, studies of learner belief and attitude are valuable sources of insight into language learning. Questionnaires are one way to do this; interviews are another. A carefully conducted survey such as the one reported here can reveal much about what language learners believe is useful in language learning and also about what they really have to do in a second language once they begin to live and study in it.

THE AUTHORS
Mary Ann Christison, Assistant Professor and Director of the English Training Center at Snow College in Utah, is the author of several reference resource books for ESL teachers. She is also on the TESOL Newsletter editorial staff and edits the Affiliate/Interest Section page.

Karl J. Krahnke is Assistant Professor of English and Director of ESL Teacher Training at Colorado State University. He has taught and directed ESL programs in Afghanistan, Iran, Utah, and Washington. He is currently coordinating the development of an annotated bibliography of ESL texts.
REFERENCES


STUDENT PERCEPTIONS OF ACADEMIC LANGUAGE STUDY
Personal Information

Name _________________________________________________________________

Native language _______________________________________________________

Age ________  Sex:  M _______  F ________

School attending ______________________________________________________

Major __________________________________________________________________

Graduate ________  Undergraduate ________

Time in U.S. ___________________________________________________________________

Time studying English:  In U.S. ________  Overall ________

Program  Concerns

1. Do you feel that your English skills improved as a result of your English program in the U. S.?

2. How could you have improved faster?

3. If you could have dropped a class from your ESL program, what would you have dropped?

4. If you could have added a class to your ESL program, what would you have added?

5. Would you have made any other changes?

6. Do you feel your ESL program helped you meet your academic needs?

7. Do you have good feelings about your language program?

8. Do you feel your ESL program helped you meet your social/practical needs in the U. S.?

9. Would you have studied English longer or for a shorter time, or was the length of your study about right?

ESL Classes and Language Skill Areas

1. Which was your most difficult ESL class? Why?

2. Which was your easiest class? Why?

3. Which was your most interesting class? Why?

4. Which class was least interesting? Why?
5. What activities did you prefer to do in your ESL classes?
6. What activities do you feel contributed most to your improvement in English?
7. What activities do you feel contributed least to your improvement?

**Out-of-Class Experiences**
1. What out-of-class experiences helped you improve your language skills the most?
2. How much English do you speak socially out of school?
3. In what ways do you use English the most away from school?

**Academic Classes**
1. In what ways do you use English the most for your academic classes?
2. What are the most difficult things for you to do in English in your academic classes?
3. What are the easiest things for you to do in English in your academic classes?

**Language Teachers**
1. Did you have a favorite teacher in your English program?
2. If “yes,” what did that teacher do to make you feel that way?
3. Do you think teachers in your ESL program knew what English you would need in your academic classes?
from Houghton Mifflin:

The text for students who mean business

Joseph F. Buschini and Richard R. Reynolds

Comprehensive, thoroughly researched text teaches students to improve business communication skills. Provides step-by-step instruction in writing and revising letters, memos, reports, and documents. Style, punctuation, and grammar guidelines in each chapter reinforce the basics of good writing.

• comprehensive coverage of essential business topics
• authoritative style guides
• developmental writing assignments and hundreds of exercises
• chapter on world trade communications
• up-to-date computer terminology
• simulations of international business transactions and communications

ENGLISH ALFA
The complete English program that teaches students how to use the language and how the language works

For adoption consideration, request an examination copy from

Houghton Mifflin
One Beacon Street
Boston, Massachusetts 02108, U.S.A.
Salience of Feedback on Error and Its Effect on EFL Writing Quality

THOMAS ROBB
Kyoto Sangyo University

STEVEN ROSS
Baika Junior College

IAN SHORTREED
Kansai University of Foreign Studies

To date, few empirical studies have been designed to evaluate the effects of different types of feedback on error in the written work of second language writers. The study reported in this article contrasted four methods of providing feedback on written error. These methods differed in the degree of salience provided to the writer in the revision process. In the study, a factor analysis was used to reduce an initial set of 19 measures of writing skill to a subset of 7. Each of the 7 measures in the subset was then used as a dependent variable in an analysis of covariance design which contrasted the effects of the feedback methods on subsequent narrative compositions. Evidence against direct correction of error in written work is discussed.

Over the past decade, considerable attention has been given to the treatment of error in the written work of second language learners. There is still no consensus, however, on how teachers can best react to student error or at what stage in the composing process such feedback should be given. Krashen (1984), for instance, advocates delaying feedback on errors until the final stage of editing and offers intensive reading practice as a long-range cure for the immediate problems of surface error. Research on the composing processes of native English speakers has reflected a similar orientation toward error correction by proposing that teachers respond to more global problems of planning and content in student writing (Griffin, 1982).

Reports from the classroom, on the other hand, indicate that teachers still respond most frequently to mechanical errors. In a study of writing in the secondary schools, Applebee (1981) found that 80% of foreign language teachers ranked mechanical errors as the most important criterion for responding to student writing. A recent study by Zamel (1985) shows that ESL teachers approach
student writing with a similar attitude. When she compared ESL and content teachers’ feedback on the same samples of writing, Zamel found that language teachers focused primarily on mechanics, whereas teachers from other disciplines responded most frequently to the students’ presentation of facts and concepts. Another study, which concurs with Zamel’s findings, reveals that content-area teachers’ perception of error gravity varies with the age of the instructor and the amount of exposure to nonnative writers (Vann, Meyer, & Lorenz, 1984).

A second and equally important finding of these classroom studies is that teachers often provide indiscriminate feedback to students. Such feedback as Cohen and Robbins (1976) report negates any positive effects of error correction. They found in their study of three ESL writers that the kinds of verb-form errors in each learner’s writing reflected a systematic aspect of the learner’s interlanguage. Since the teachers did not keep a record of the types of errors each learner made, it was impossible to provide suitable remedial work. The studies of Greenbaum and Taylor (1982) and Fearn (1982) report a similar phenomenon among college composition teachers. In both studies, teachers were presented with sentences containing specific types of errors and were asked to classify them and provide the corrected form. While most teachers were able to perform the latter task, almost 35% of the errors were categorized incorrectly.

Other studies have been designed to provide a more systematic approach to error feedback. Stiff (1967) examined the effect of terminal and marginal corrections but found that neither type of correction was significantly related to writing quality. Hillocks (1982) investigated the effect of long and short comments in conjunction with instructional variables such as pre-writing, but owing to the complexity of the design, the results were difficult to interpret. Hendrickson (1978) attempted to control for error gravity by employing Burt and Kiparsky’s (1972) global and local error taxonomy, but both treatments (direct and selective corrections) resulted in insignificant reduction of errors. As Hendrickson (1981) has pointed out, overt correction of both groups’ compositions may have negated the effect of selective feedback.

Two recent studies provide empirical support for Hendrickson’s conclusions. In a carefully designed experiment, Lalande (1982) found that students who used an error code when revising their compositions made significantly greater gains than a group whose compositions were corrected directly by the instructor. In a similar study, Semke (1984) found that overt correction of student writing tended to have negative side effects on both the quality of
subsequent compositions and on student attitudes toward writing in the foreign language.

The findings of these studies support Corder (1981) and Brumfit (1980), who have hypothesized that learners will retain feedback only if they are forced to approach error correction as a problem-solving activity. Brumfit identifies six different methods of providing indirect feedback, ranging from locating an error by using an error code (the most salient) to simply asking students to revise their compositions without any feedback at all (the least salient form). The essential question to be asked about feedback on error, however, is concerned with the appropriate degree of salience necessary before students can effectively revise a composition. What is the most effective and practical feedback strategy in an EFL context characterized by extremely large teacher-to-student ratios and little contact time?

The study reported here investigated the relative merits of indirect and direct feedback by comparing four types of error treatment, each of which provided EFL writers with progressively less salient information for making revisions in their compositions. The investigators tested the hypothesis that more salient error-feedback treatments would have a significant effect on improving the student's overall writing quality. Thus, the study was designed to verify the findings of Lalande (1982), Hendrickson (1978), and Semke (1984) in an EFL context.

METHOD

A total of 134 Japanese college freshmen were alphabetically assigned to four sections of English composition. A cloze test administered during the first class meeting indicated no significant differences between the groups at the onset of the study ($F=.250$, n.s.). Students also wrote a narrative composition on an assigned topic in the second class meeting, which provided an additional baseline measure. Although the cloze test did not indicate any between-group differences, the first composition did, and it was therefore used as the covariate in the subsequent analyses.

Students attended a total of 23 classes over the academic year, from mid-April until mid-January, with a summer vacation of 2

A finding ancillary to this study relates to first and second language research which suggests that the mode of writing has a significant effect on syntactic complexity. Matched-Pair t tests were used here to determine if the 19 original measures of writing quality varied across expository and narrative modes for these elementary-level EFL writers. No significant differences emerged, however, which contrasts with research suggesting that text unit length and other objective measures vary considerably between narrative and persuasive writing samples (Crowhurst, 1983; Sinclair, 1983).
months and a winter vacation of 2 weeks. Each class meeting lasted for an hour and a half, for a total of 34.5 hours of classroom instruction. An attempt was made to offset any teacher-style variable by rotating the two instructors between the four classes approximately one third of the total instruction time.

Classroom activities for the four groups were identical: 40% of class time was spent on editing grammatical errors produced by freshmen writers on the same topic the previous year, and 40% of the time was spent on sentence-combining exercises. The remainder of the class time was spent in preparation for the next week’s assigned composition. All composition assignments were identical for the four sections and included a selection of expository, narrative, and descriptive essays. Learners in all four sections were required to revise their weekly essays, based on the feedback provided by the instructor. The revisions were returned to the instructor during the following class meeting and were then checked for accuracy.

The variable manipulated by the investigators was the type of feedback learners in each group received. The correction group \((n = 30)\) papers were completely corrected by the instructor, with the corrections covering all categories of lexical, syntactic, and stylistic errors. Substantive errors in content or organization were not corrected. Once the papers were returned, the students in the correction group needed only to copy their original compositions, carefully incorporating the instructor’s corrections.

The compositions of the coded feedback group \((n = 37)\) were marked in an abbreviated code system in which the type of error was indicated on the student’s paper. Students in this group revised their compositions by using a guide to decipher the instructor’s markings on their papers.

The compositions of the uncoded feedback group \((n= 37)\) were marked over with a yellow text-marking pen. The uncoded feedback differed from the coded feedback in the salience of the marking: The former specified the location of the places in need of editing or revision but did not indicate specifically why the instructor chose to mark any given part of the composition.

The compositions of the marginal feedback group \((n = 30)\) were marked with the least salient method. The number of errors per line was totaled and written in the margins of the student’s paper. Students were requested to reread each line of their composition to search for the places in need of revision. Once an error was located, the students had to correct it. Figure 1 summarizes the four feedback methods.

The students wrote five narrative test compositions at equal intervals during the academic year. These five compositions were
FIGURE 1
Feedback Methods

<table>
<thead>
<tr>
<th>Group</th>
<th>Location of error</th>
<th>Content of error</th>
<th>Model by teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correction</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Coded</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Uncoded</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Marginal</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

analyzed and graded using 1 subjective and 18 objective measures of writing ability:

1. A holistic range of writing ability (A, B, C, D)
2. The Usage Correctness Score (Brodkey & Young, 1981)
3. The total of words written
4. The total number of additional clauses embedded in T-units
5. The total number of error-free T-units
6. The total number of T-units written
7. The ratio of words in error-free T-units to total error-free T-units
8. The ratio of error-free T-units to total T-units
9. The ratio of error-free T-units to total clauses
10. The ratio of error-free T-units to total words written
11. The ratio of words in error-free T-units to total words
12. The number of words per T-unit
13. The ratio of words in error-free T-units to total T-units
14. The number of error-free clauses
15. The ratio of total clauses to total words written
16. The ratio of additional clauses to total words written
17. The total number of clauses written
18. The total number of words in error-free T-units
19. The ratio of error-free clauses to total clauses

Three raters graded the five sets of narrative compositions after masking over student identification markings and assigning the papers at random. Interrater reliability estimates (Kendall’s coefficient of concordance) calculated at the start of the study were sufficient at .87 for the objective scoring and .81 for the holistic ratings.
Each narrative test composition was factor-analyzed separately so that the original battery of 19 measures could be reduced to a less redundant subset. In all, 676 compositions were included in the analysis in five sets, each of which was rotated to varimax solution yielding three composite factors labeled accuracy, fluency, and syntactic complexity. The factors so labeled encompassed clusters of variables that incorporated the “error-free” criterion (accuracy), the gross number of total clauses and words (fluency), and “additional” clauses (complexity).

A subset of 7 variables was selected from the original 19 by narrowing the list down to those variables that loaded highly and consistently on one of the factors (see Appendix). The four progress test compositions were analyzed separately in an analysis of covariance using data derived by using the pretest composition as the covariate. Analysis of covariance permits comparisons to be made among nonequivalent groups, when, as in this study, the assumptions of a truly randomized experimental design are not met. Although the initial between-group differences were small on the pretest, all scores reported in the tables are adjusted for the initial differences on the covariate.

RESULTS

Two of the measures of accuracy, the ratio of error-free T-units to total T-units and the ratio of error-free T-units to total clauses, showed a trend toward a difference between the groups by the third test composition (see Table 1). However, an examination of the mean scores for each of the feedback groups suggests that the assumption underlying overt correction—that more correction results in more accuracy—was not convincingly demonstrated. The apparent trend indicating that accuracy improved in the correction group’s writing in fact did not continue from the third composition to the fourth, which was written after the winter vacation.

The results of the analysis of the accuracy criterion concur with the aforementioned research on error correction: In general, the more direct methods of feedback do not tend to produce results commensurate with the amount of effort required of the instructor to draw the student’s attention to surface errors. Rather, as Table 1

2 The scores of 6 sophomore repeaters were entered in the factor analyses. The scores for these students were not included in the data base on which analyses of covariance were performed.

3 Given the fact that a number of variables loading on the fluency factor also load substantially on both accuracy and complexity as well, it could be argued that an oblique method of rotation would provide a more accurate factor structure than does the varimax rotation used here.
suggests, practice in writing over time resulted in gradual increases in the mean scores of all four groups when compared with the initial pretest scores, regardless of the method of feedback they received.

On the fluency measures, initial differences among the groups on the first two tests gradually diminished (see Table 2). The results for these measures provide some counterevidence to the claim that overt correction “causes” foreign language writers to be overly concerned with surface structure to the extent that fluent writing is constrained. Whatever negative influences corrective feedback might have produced seem to have been completely offset by the practice effect arising from writing weekly assignments.

On the first three narrative compositions, no statistically significant differences were found on the complexity measures (see Table 3). There is some reason to believe that the kind of correction given to the correction group was still too obscure for the students to untangle as they compared their original compositions with the rewritten passages and corrected structures provided by the instructors. This finding suggests that EFL writers can assimilate only a small proportion of corrective feedback into their current grammatical system, especially when the corrections are not detailed enough to be applied to the more complex and

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Composition</th>
<th>Correction</th>
<th>Coded</th>
<th>Uncoded</th>
<th>Marginal</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>.343</td>
<td>.367</td>
<td>.362</td>
<td>.312</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.375</td>
<td>.387</td>
<td>.380</td>
<td>.365</td>
<td>1.47</td>
</tr>
<tr>
<td>2</td>
<td>.405</td>
<td>.385</td>
<td>.389</td>
<td>.363</td>
<td>0.24</td>
</tr>
<tr>
<td>3</td>
<td>.474</td>
<td>.414</td>
<td>.432</td>
<td>.432</td>
<td>2.25</td>
</tr>
<tr>
<td>4</td>
<td>.436</td>
<td>.404</td>
<td>.435</td>
<td>.419</td>
<td>0.29</td>
</tr>
</tbody>
</table>

**TABLE 2**

<table>
<thead>
<tr>
<th>Ratio of error-free T-units to total clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

**TABLE 3**

<table>
<thead>
<tr>
<th>Ratio of words in error-free T-units to total words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>.330</td>
</tr>
<tr>
<td>.313</td>
</tr>
<tr>
<td>.310</td>
</tr>
<tr>
<td>.390</td>
</tr>
<tr>
<td>.341</td>
</tr>
</tbody>
</table>
### TABLE 2
Fluency

<table>
<thead>
<tr>
<th>Composition</th>
<th>Correction</th>
<th>Coded</th>
<th>Uncoded</th>
<th>Marginal</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>143.0</td>
<td>127.7</td>
<td>155.2</td>
<td>131.8</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>252.2</td>
<td>232.5</td>
<td>253.3</td>
<td>247.9</td>
<td>3.70*</td>
</tr>
<tr>
<td>2</td>
<td>258.1</td>
<td>239.5</td>
<td>255.3</td>
<td>251.3</td>
<td>2.23</td>
</tr>
<tr>
<td>3</td>
<td>260.7</td>
<td>259.8</td>
<td>275.8</td>
<td>264.9</td>
<td>1.15</td>
</tr>
<tr>
<td>4</td>
<td>273.2</td>
<td>264.1</td>
<td>267.4</td>
<td>288.7</td>
<td>0.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

*p < .05.

### TABLE 3
Complexity

<table>
<thead>
<tr>
<th>Composition</th>
<th>Correction</th>
<th>Coded</th>
<th>Uncoded</th>
<th>Marginal</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>0.023</td>
<td>0.027</td>
<td>0.021</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.030</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>0.029</td>
<td>0.027</td>
<td>0.028</td>
<td>0.028</td>
<td>0.24</td>
</tr>
<tr>
<td>3</td>
<td>0.038</td>
<td>0.037</td>
<td>0.037</td>
<td>0.038</td>
<td>0.98</td>
</tr>
<tr>
<td>4</td>
<td>0.037</td>
<td>0.041</td>
<td>0.032</td>
<td>0.036</td>
<td>1.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

*p = <.05.

problematic aspects of word order and syntax (Ross, 1982). This of course says nothing about the value of feedback on aspects of coherence or textual cohesion.

For additional clauses, however, there is a significant difference
from Composition 3 to Composition 4, with the coded group improving by almost four extra clauses.\(^4\)

DISCUSSION

Teachers of English as a foreign language often spend a great deal of time responding to the mechanics of student writing. This study, however, does not support the practice of direct correction of surface error. Since negligible differences were found among the groups on most of the criterion measures, the results suggest that less time-consuming methods of directing student attention to surface error may suffice. While well-intentioned teachers may provide elaborate forms of corrective feedback, time might be more profitably spent in responding to more important aspects of student writing.

This latter observation relates directly to some recent studies on the responding strategies of ESL teachers (see Zamel, 1985). Corrective feedback exclusively on sentence-level errors addresses only one aspect of overall student writing ability. Indeed, if teachers consider their students in need of some form of corrective feedback at the editing stage of writing, then, as Eskey (1983) argues, focus on form is justified. However, teachers should not assume, as they often do, that such feedback directly affects other aspects of composing ability. The fact that students in all of the groups in this study wrote more complex structures as the course progressed indicates that improvement was independent of type of feedback.\(^5\)

The implications of this study extend well beyond classroom practice, to basic issues in the teaching of writing. The results suggest that highly detailed feedback on sentence-level mechanics may not be worth the instructors’ time and effort even if, as Cohen (in press) suggests, students claim to need and use it. Alternatively, teachers can respond to student writing with comments that force the writer back to the initial stages of composing, or what Sommers (1982) refers to as the “chaos,” “back to the point where they are shaping and restructuring their meaning” (p. 154).

\(^4\) The suddenness of this improvement seems to indicate that some extraneous influence, such as recent work in a concurrent class, may have brought about the increase.

\(^5\) These gains more likely resulted from the combined effects of systematic sentence-combining practice and the writing of weekly compositions, not to mention the effect of six other language courses that the students were concurrently enrolled in. A follow-up study (now in progress) is investigating the extent to which sentence combining and journal writing influence differential development of EFL writing skills.
ACKNOWLEDGMENTS

The study reported in this article was partially funded by a research grant from the Japan Association of Language Teachers.

THE AUTHORS

Thomas Robb is an Assistant Professor of English at Kyoto Sangyo University. He is currently Executive Secretary of the Japan Association of Language Teachers.

Steven Ross is a Lecturer in English at Baika Junior College, Kyoto Sangyo University, and Osaka University. He is a member of the Japan Association of Language Teachers and the Japan Association of College English Teachers.

Ian Shortreed is a Lecturer in the Department of English at Kansai University of Foreign Studies, where he has taught for the past 5 years.

REFERENCES


## APPENDIX

Factor Structure of 19 Measures of EFL Writing (N = 676)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Composition</th>
<th>Commu-</th>
<th>Factor 1 (accuracy)</th>
<th>Factor 2 (fluency)</th>
<th>Factor 3 (complexity)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>nality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holistic</td>
<td>1</td>
<td>.326</td>
<td>.476</td>
<td>.468</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.493</td>
<td>.511</td>
<td>.468</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.214</td>
<td>.392</td>
<td>.537</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.384</td>
<td>.557</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.380</td>
<td>.547</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total words</td>
<td>1</td>
<td>.863</td>
<td>.914</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.469</td>
<td>.665</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.441</td>
<td>.588</td>
<td>.306</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.373</td>
<td>.579</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.295</td>
<td>.512</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>1</td>
<td>.478</td>
<td>.666</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.665</td>
<td>.669</td>
<td>.405</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.617</td>
<td>.749</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.696</td>
<td>.805</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.658</td>
<td>.796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional clauses</td>
<td>1</td>
<td>.813</td>
<td>.474</td>
<td>.706</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.898</td>
<td>.826</td>
<td>.450</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.806</td>
<td>.868</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.962</td>
<td>.954</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.943</td>
<td>.963</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total clauses</td>
<td>1</td>
<td>.957</td>
<td>.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.941</td>
<td>.968</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.954</td>
<td>.967</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.874</td>
<td>.877</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.892</td>
<td>.834</td>
<td>.443</td>
<td></td>
</tr>
<tr>
<td>Error-free T-units</td>
<td>1</td>
<td>.921</td>
<td>.721</td>
<td>.600</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.940</td>
<td>.825</td>
<td>.492</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.953</td>
<td>.801</td>
<td>.535</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.942</td>
<td>.708</td>
<td>.894</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.968</td>
<td>.706</td>
<td>.638</td>
<td></td>
</tr>
<tr>
<td>Words in error-free</td>
<td>1</td>
<td>.859</td>
<td>.736</td>
<td>.535</td>
<td></td>
</tr>
<tr>
<td>T-units</td>
<td>2</td>
<td>.904</td>
<td>.888</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.900</td>
<td>.889</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.906</td>
<td>.849</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.866</td>
<td>.885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-units</td>
<td>1</td>
<td>.939</td>
<td>.875</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.919</td>
<td>.878</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.889</td>
<td>.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.836</td>
<td>.888</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.868</td>
<td>.908</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words in error-free</td>
<td>1</td>
<td>.932</td>
<td>.847</td>
<td>.460</td>
<td></td>
</tr>
<tr>
<td>T-units</td>
<td>2</td>
<td>.942</td>
<td>.862</td>
<td>.402</td>
<td></td>
</tr>
<tr>
<td>total T-units</td>
<td>3</td>
<td>.922</td>
<td>.847</td>
<td>.369</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.911</td>
<td>.892</td>
<td>.319</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.891</td>
<td>.911</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>Composition</td>
<td>Communitivity</td>
<td>Factor 1 (accuracy)</td>
<td>Factor 2 (fluency)</td>
<td>Factor 3 (complexity)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Error-free</td>
<td>1</td>
<td>.938</td>
<td>.959</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-units/total T-units</td>
<td>2</td>
<td>.957</td>
<td>.973</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.909</td>
<td>.952</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.952</td>
<td>.926</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.922</td>
<td>.918</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>1</td>
<td>.940</td>
<td>.957</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-units/total clauses</td>
<td>2</td>
<td>.970</td>
<td>.971</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.874</td>
<td>.908</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.960</td>
<td>.893</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.960</td>
<td>.812</td>
<td>−.483</td>
<td></td>
</tr>
<tr>
<td>Error-free</td>
<td>1</td>
<td>.917</td>
<td>.882</td>
<td>−.323</td>
<td></td>
</tr>
<tr>
<td>T-units/total words</td>
<td>2</td>
<td>.891</td>
<td>.912</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.907</td>
<td>.849</td>
<td>.342</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.935</td>
<td>.783</td>
<td>−.559</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.944</td>
<td>.765</td>
<td>.407</td>
<td></td>
</tr>
<tr>
<td>Words in error-free T-units/total words</td>
<td>1</td>
<td>.889</td>
<td>.926</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.924</td>
<td>.941</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.916</td>
<td>.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.934</td>
<td>.913</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.885</td>
<td>.920</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words per T-unit</td>
<td>1</td>
<td>.662</td>
<td></td>
<td>.806</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.574</td>
<td></td>
<td>.614</td>
<td>−.420</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.769</td>
<td></td>
<td>−.681</td>
<td>.536</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.750</td>
<td></td>
<td>.839</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.777</td>
<td></td>
<td>−.787</td>
<td>.385</td>
</tr>
<tr>
<td>Words in error-free T-units/total words</td>
<td>1</td>
<td>.514</td>
<td>.378</td>
<td></td>
<td>.599</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.468</td>
<td></td>
<td>.609</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.624</td>
<td></td>
<td>−.468</td>
<td>.586</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.505</td>
<td></td>
<td>.540</td>
<td>.451</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.552</td>
<td></td>
<td>−.686</td>
<td></td>
</tr>
<tr>
<td>Error-free clauses</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.739</td>
<td></td>
<td>.465</td>
<td>.711</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.756</td>
<td></td>
<td>.441</td>
<td>.747</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.730</td>
<td></td>
<td>.470</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.703</td>
<td></td>
<td>.669</td>
<td>.505</td>
</tr>
<tr>
<td>Error-free clauses/total clauses</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.757</td>
<td></td>
<td>.515</td>
<td>.722</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.742</td>
<td></td>
<td>.444</td>
<td>.673</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.615</td>
<td></td>
<td>.762</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.718</td>
<td></td>
<td>.702</td>
<td>.410</td>
</tr>
<tr>
<td>Total clauses/total words</td>
<td>1</td>
<td>.147</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.301</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.593</td>
<td></td>
<td></td>
<td>.491</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.779</td>
<td></td>
<td></td>
<td>.459</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.681</td>
<td></td>
<td></td>
<td>.459</td>
</tr>
<tr>
<td>Additional clauses/total words</td>
<td>1</td>
<td>.624</td>
<td></td>
<td></td>
<td>.785</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.771</td>
<td></td>
<td></td>
<td>.861</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.577</td>
<td></td>
<td></td>
<td>.756</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.919</td>
<td></td>
<td></td>
<td>.935</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.879</td>
<td></td>
<td></td>
<td>.931</td>
</tr>
</tbody>
</table>

*Not tallied.*
ESL texts that teach more than A, B, C...

INDEPENDENT WRITING
Teresa D. O'Donnell and Judy L. Paiva
Advanced Composition
This text helps ESL learners attain college-level writing proficiency by emphasizing a process approach to writing—from prewriting strategies, revising, multiple drafting, and independent editing through to the final draft. Structural reviews reflect the interaction between the writing mode and the structures necessary for each type of essay.
paper/220 pages/1986/#630012

GRAMMAR IN CONTEXT
Sandra N. Elbaum
Book One: high beginning to low intermediate levels
Book Two: high intermediate to low advanced levels
In this two-book series, grammar is presented through short, nonfiction readings that impart information on American life that is of interest to ESL learners. Some exercises elicit factual response; while others encourage students’ contributions about their own experiences and peer interaction. Linguistically and culturally, GRAMMAR IN CONTEXT eases the transition to American life.
Book One: paper/350 pages/1986/#231371
Book Two: paper/350 pages/1986/#231398

ENGLISH INTEGRATED
Henrietta C. Dunham and Catherine Vaden Summers
Advanced-level ESL
This combination reader and grammar contains mature, engrossing readings by such notable authors as W.H. Auden, Henry Fairlie, and Susan Sontag. Grammar is taught using examples from the preceding reading. Each selection includes vocabulary, pre-reading questions, author biographies, discussion questions, and exercises.
paper/420 pages/1986/#197122

...from
Little, Brown
and Company
College Division
34 Beacon Street, Boston, MA 02106
Cloze, which combines the advantages of integrative testing and objective scoring, was investigated as a supplement to a standardized ESL test and as an alternative to a written composition test. Three tests of English language proficiency were given to a large group of students applying for admission to the American University of Beirut (AUB): the AUB English Test, a cloze test, and a written composition. The tests were taken by the same group of students at the same examination session, enabling direct study of interrelationships among the three measures. Regression analysis of the test scores showed that pairwise correlations were all high and that a combination of cloze test scores and AUB English Test scores significantly improved the prediction of communicative language proficiency, as measured by the composition test scores. In addition, there was a substantial residual correlation between the cloze and writing tests, which suggests that these tests may measure in common some aspects of language ability beyond those that they share with the AUB English Test, a standardized ESL test. These results indicate that a cloze component can serve as a valuable supplement in language proficiency testing. Further implications of the findings of the study are discussed.

This article reports on some recent research on ESL proficiency testing which was carried out at the American University of Beirut (AUB). The main purpose of the work was to study interrelationships among three types of measures: a standardized ESL test, a cloze test, and a written composition test. More specifically, the study sought to determine whether the addition of a cloze component to the standardized ESL test would improve the predictability of students’ communicative proficiency as reflected in their performance on a writing test.

At the American University of Beirut, where English is the medium of instruction, a standardized ESL test is taken by large
numbers of applicants to the University and to other institutions of higher education in the area. This instrument, the AUB English Test, was developed in the 1960s by the AUB Office of Tests and Measurements and is similar in purpose and content to other known ESL proficiency tests, such as TOEFL.

The 2-hour test consists of 200 multiple-choice items distributed over four sections: structure, vocabulary, reading comprehension, and miscellaneous abilities. The items are designed to sample a wide range of abilities required in a program of university-level education in which English is the medium of instruction. The test has several equivalent versions and is standardized on a College Entrance Examination Board (CEEB) scale (Harris, 1969, p. 127). Test items are analyzed and revised on a continuing basis. Test reliability coefficients (KR-20 or KR-21) are reported to range from .94 to .98, and two studies of criterion-related validity have yielded correlation coefficients of .78 and .88 between AUB English Test and TOEFL scores (Baroudi, 1983; Miller, 1983).

The AUB English Test is offered at several scheduled sessions and is taken by over 8,000 candidates a year. A minimum score of 500 is required for admission to the University. The scores are also used for placement into the appropriate level of English communication skills courses.

As is the case in other institutions where standardized tests are given, concern has been expressed at the American University of Beirut that performance on a discrete-point test may not accurately reflect ability to function in the language (Moller, 1981, p. 59). Indeed, some teachers hold the view that knowledge of English is assessed better and more directly through students’ writing. Support for these views comes from language acquisition research, which has found that a learner’s performance varies according to the task required. On formal tasks which focus on one item at a time, the language learner can bring to bear a conscious knowledge of rules that have been formally learned but have not yet become part of the learner’s productive grammar system, or acquired knowledge (Dulay, Burt, & Krashen, 1982, p. 62). Thus, in many cases, the score on a discrete-point test may be indicative not of overall language ability, but of intensive learning of isolated items and grammar rules. In fact, a number of language schools have sprung up in recent years which aim at giving direct training for the AUB test, the TOEFL, and other tests.

The advantage of adding an essay-writing component to the test was previously investigated in a study carried out by the AUB Office of Tests and Measurements, in collaboration with the University’s Communication Skills Program. The results indicated
that the addition of an essay-writing component only marginally improved the prediction of students’ English course grades and their grade point averages. This small gain was considered to be offset by the additional burden involved in administering and scoring the writing test (Miller, 1978).

As an alternative approach to this problem, we have explored the use of a written cloze test to supplement the AUB English Test. A cloze test typically consists of a passage of about 300 words, from which 50 words have been deleted at regular intervals. The first sentence is usually left intact to help establish the context. A person taking the test has to fill in each blank with the word which best fits the meaning. Scoring is objective and can be done by the exact-word method, for which only the word given in the original text is considered correct, or by the acceptable-word method, for which acceptable alternatives are also marked correct.

Cloze is considered an integrative rather than a discrete-point test because it draws at once on the overall grammatical, semantic, and rhetorical knowledge of the language. To reconstruct the textual message, students have to understand key ideas and perceive interrelationships within a stretch of continuous discourse, and they have to produce, rather than simply recognize, an appropriate word for each blank. The focus of the task involved is more communicative than formal in nature, and it is therefore considered to reflect a person’s ability to function in the language.

Cloze procedure, which was first applied as a reliability measure with native speakers (Taylor, 1953), has since been demonstrated in many studies to have substantial concurrent validity as an integrative test of overall proficiency in English as a second language (Hinofotis, 1980; Irvine, Atai, & Oller, 1974; Oller, 1972; Oller & Conrad, 1971; Stubbs & Tucker, 1974). In these studies, high correlations were obtained between cloze scores and corresponding scores on an established measure of language proficiency, such as the UCLA English as a Second Language Placement Examination. In one of the studies (Stubbs & Tucker, 1974), which was carried out at the AUB, the correlation between cloze scores and scores on the AUB English Test for a sample of 211 students was found to be .71. A very high correlation of .97 between scoring by the exact-word and by the acceptable-word method was also reported.

In the research reported in this article, the cloze test was investigated not as an independent proficiency measure, but as a supplementary component in ESL testing. It seemed possible that a cloze test, by combining the advantages of integrative testing and objective scoring, might serve a purpose similar to essay writing,
while avoiding the problem of subjectivity and the additional work involved in scoring writing samples. Two-way relations had previously been considered between the AUB English Test and cloze (Stubbs & Tucker, 1974) and between the AUB English Test and writing (Miller, 1978). To make possible a direct study of interrelationships among these three proficiency measures, a three-part proficiency test battery, comprised of the AUB English Test, a cloze test, and a written composition, was given concurrently to a large group of students at the same examination session.

METHODOLOGY

The study was carried out in three stages, the first of which involved selecting cloze passages and composition topics. Since previous research (particularly Alderson, 1979) had indicated that cloze procedure does not automatically produce valid proficiency tests, special attention was given to the selection and validation of cloze passages. Several criteria guided the preliminary choice. Passages were picked from written material in English textbooks designed for upper high school to college sophomore level. The passages had to be on a general topic not requiring specialized knowledge and had to form coherent self-contained units of discourse.

Each text was to yield 50 systematically spaced blanks (fifth or seventh word). The deleted items were then examined to ascertain that they covered a variety of syntactic and cohesive functions and that they were adequately cued in the text. Fifteen cloze passages which were prepared in this way were informally tried out; the six which appeared to differentiate well among ability levels were selected for a trial run.

The topics chosen for the composition test, which were related to the students’ general experience, were ones that the students would be motivated to write about and that would be likely to elicit a variety of linguistic structures. Topics requiring argumentation or definition of abstract concepts were excluded. Ten composition topics were selected; these were paired to provide a choice of two at each examination sitting. The students were expected to write about 250 words on one of the two given topics.

The second stage of the study involved a preliminary run of the tests. The purpose was twofold: to make a final selection of the cloze versions to be used and to examine the feasibility of administering a triple test. A group of 400 students taking the English entrance examination was given the regular AUB English Test and a cloze test. The six cloze versions were randomly
distributed in equal numbers. The results of this group (as explained below) served as a basis for the selection of four cloze versions for the final run. At a later sitting of the same examination, a group of 200 students was given the complete triple test: cloze (1/2 hour), AUB English Test (2 hours), and written composition (1/2 hour), in that order. This arrangement was found to be successful and was maintained throughout. As had been anticipated, the cloze task served as a warm-up activity preceding the long multiple-choice test, and the writing task provided a welcome opportunity for free self-expression at the end.

Cloze selection was made as follows. For each of the cloze versions, the scores were used to obtain a frequency distribution, the mean, median, standard deviation, interquartile range, a phi-coefficient (based on cloze and AUB test means), and a difficulty coefficient (relative to the mean of medians for all the cloze versions). The four cloze passages which showed large spread, the highest discrimination, and a difficulty coefficient of about .50 were chosen for the final run. The estimated (KR-21) reliability coefficients for the cloze tests ranged from .92 to .98, and their correlation with the AUB English Test was .76.

In the third and major stage of the study, the three tests (cloze, AUB test, composition, in that order) were administered to all 1,572 students taking the AUB English Test at the next examination session. The students took the examination in 10 sittings, with about 150 students at each sitting. AUB test scores were, as usual, computer graded and converted to standardized scores, using the CEEB scale, which has a mean of 500 and a standard deviation of 100. From this population, a sample and a subsample were selected for analysis. The sample ($N = 337$), about 20% of the population, consisted of students whose mean score was near the population mean; the subsample ($n = 159$) consisted of students whose mean score was somewhat above the mean, that is, whose score was above the minimum score of 500 required for admission to the University. All subjects came from a variety of schools in the area.

The cloze tests were scored by the exact-word method. Since the four cloze versions had different means and standard deviations, the raw scores were converted to the same CEEB scale. The conversions were based on the AUB test mean (484.6) and standard deviation (95.0) for the sample of 337 students used in the study (see Table 1).

The written compositions were each graded independently by two experienced teachers, who, using the general-impression (holistic) method, considered grammar, mechanics, and rhetorical aspects. Grading, which was set on a percent scale at 5-point
intervals, was calibrated to correspond to proficiency levels for remedial, regular, and advanced communication skills courses. The scores given by the two graders were averaged; if the difference exceeded 10 points, a third reader graded the composition, and the three scores were averaged. The reliability of scoring was high; linear correlations between the scores of the graders averaged .78.

The scores for all three tests, along with additional information on student background, were computerized for data analysis using the SPSS program (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). Data were analyzed first for the overall sample of 337 students and then for the subsample of 159 whose AUB English Test scores were 500 or above. Multiple regression analysis was carried out on the three measures for all permutations, yielding linear, multiple, and partial correlation coefficients, regression equations, and F ratios.

RESULTS

The primary aim of the analysis was to answer two questions: (a) How do the cloze and writing tests relate to each other and to the AUB test? (b) Do AUB test and cloze scores together predict overall ability, as indicated by the composition scores, better than AUB test scores alone? The results of the quantitative analysis of data are presented below, first for the whole sample and then for the restricted sample.

Overall Sample

As Table 2 demonstrates, the tests correlate highly with each other: .79 for AUB English Test and cloze, .73 for AUB test and
composition, and .68 for cloze and composition. The high correlations indicate a degree of commonality among the three tests and confirm their validity as tests of language proficiency.

<table>
<thead>
<tr>
<th>Tests</th>
<th>Overall sample</th>
<th>Restricted sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUB test/cloze</td>
<td>.79</td>
<td>.57</td>
</tr>
<tr>
<td>AUB test/composition</td>
<td>.73</td>
<td>.51</td>
</tr>
<tr>
<td>Cloze/composition</td>
<td>.68</td>
<td>.51</td>
</tr>
</tbody>
</table>

Next, a multiple regression analysis was done to determine the correlation between composition scores and combined scores on the AUB test and the cloze test (see Table 3). Composition scores were treated as the dependent variable, and AUB test and cloze test scores as independent variables. Whereas AUB test scores alone account for 53% of the variation in composition scores, the combination of AUB test and cloze scores accounts for 56% of the variation, a small but distinct improvement. The extent of this improvement can be assessed by the F-ratio statistic, which measures the additional effect of a variable when the contributions

<table>
<thead>
<tr>
<th>Tests</th>
<th>Simple $r$ Overall sample</th>
<th>Simple $r$ Restricted sample</th>
<th>$R^2$ Overall sample</th>
<th>$R^2$ Restricted sample</th>
<th>$F$ Overall sample</th>
<th>$F$ Restricted sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composition/ AUB test</td>
<td>.73</td>
<td>.51</td>
<td>.53</td>
<td>.26</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Composition/ AUB test and cloze</td>
<td>.75</td>
<td>.58</td>
<td>.56</td>
<td>.33</td>
<td>23.5</td>
<td>16.5</td>
</tr>
</tbody>
</table>
of other variables have been accounted for. In this case, $F$ was found to be 23.5, which is significant at the .01 level (critical $F$ value is 6.7 at $p = .01$). On this basis, it may be concluded that the incorporation of cloze test scores with AUB test scores significantly improves the predictability of composition scores.

To arrive at a more complete understanding of the interrelationships among the three measures, multiple regression analysis was also done with AUB test scores and then with cloze test scores as the dependent variable, the other two measures being permuted in each case. The results of both sets of analysis showed the same general pattern—that a combination of two measures increases the predictability of the third, again significantly at the .01 level. Comparison of the $F$ ratios for the various permutations indicates that the additional effect of AUB test scores is the largest, that of cloze test scores next, and that of composition test scores the least. These results imply that the AUB English Test is the most comprehensive of the three, which is not surprising in view of its length and varied content.

**Restricted Sample**

Since the questions being investigated in the study were particularly relevant to students who secure admission to university, the above numerical analysis was repeated on the restricted sample of students whose AUB English Test scores were 500 or above ($n= 159$). This group’s mean scores on the three tests are of course higher than those of the overall sample, which included weaker students. The correlation coefficients, on the other hand, are lower, as is expected for a truncated sample (see Tables 1 and 2).

Multiple regression analysis again shows a small improvement in the predictability of composition scores when cloze test and AUB test scores are combined, $R^2$ rising from .26 to .33 (see Table 3). The $F$-ratio value of 16.5 indicates that this improvement is again significant at the .01 level. The regression equation, with composition scores as the dependent variable, shows a larger relative contribution from cloze test scores (.047) than was the case for the overall sample. The same pattern emerges when the role of dependent variable is permuted among the three tests.

---

1 The appropriate combination for best fit is given by the regression equation: composition scores $= 8.27 + .057$ (AUB test scores) + .033 (cloze test scores). This equation shows that the contribution of AUB test scores is almost twice that of cloze test scores (the coefficients being .057 and .033 respectively).

2 Composition scores $= -1.10 + .061$ (AUB test scores) + .047 (cloze test scores).
Partial Correlation Analysis

The above analysis demonstrates that there is substantial overlap among these three tests and that the combination of any two sets of scores increases the predictability of the third. These interrelationships can also be examined by determining the extent to which two of the tests overlap, independently of what they have in common with the third. The extent of this residual overlap is indicated by the partial correlation coefficient, partial $r$, which measures the association (correlation) between two variables while controlling (eliminating) the effect of the third. Multiple regression analysis enables the computation of partial $r$ values.

The three sets of two-way relationships involved here are represented in Figure 1 for both the overall and the restricted samples. The three language proficiency measures are placed at the apexes of the triangles; linear correlation coefficient values ($r$) appear outside and partial-$r$ values within the sides of the triangles. For any two language tests, the difference between the values of $r$ and partial-$r$ reflects the extent of the commonality (overlap) between these two tests which is also shared with the third, that is, the strength of the effect of the third. The partial $r$ values reflect what is common between the two tests and at the same time distinctive from the third.

FIGURE 1
Two-Way Correlations Among Three Tests

<table>
<thead>
<tr>
<th>Overall sample</th>
<th>Restricted sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloze</td>
<td>.68</td>
</tr>
<tr>
<td>Composition</td>
<td>.26</td>
</tr>
<tr>
<td>AUB test</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>.73</td>
</tr>
<tr>
<td></td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>.31</td>
</tr>
</tbody>
</table>

Note: Linear correlation coefficients are shown outside the triangles; partial correlation coefficients are shown inside the triangles.

Figure 1 indicates that for the overall sample, the AUB test has the largest effect, followed by the cloze and then the composition test.
Correspondingly, the partial correlation is largest between the AUB test and the cloze followed by the AUB test and composition, and then cloze and composition. For the restricted sample, a similar pattern emerges, but the effects of the AUB test and of the cloze test are comparable, and the values for the residual correlations are closer together.

In the context of this study, the partial correlation between cloze and composition scores is of particular interest. The data show that although there is strong overlap between the AUB test and the cloze and between the AUB test and the composition test, there remains a partial correlation between cloze and composition of .26 for the overall sample and of .31 for the restricted sample. This partial correlation points to a residual commonality between cloze and composition independent of the AUB test. The cloze and writing tests therefore appear to provide additional information about proficiency beyond that provided by the AUB English Test, information which is common to both of them. It should be noted that these findings do not preclude the possibility that either the cloze or the writing test measures aspects of language ability other than those which they measure in common with each other or with the AUB English Test.

DISCUSSION

The above results reveal an interesting pattern of interrelationships among the AUB English Test, cloze test, and writing test. The high correlations observed reflect the validity of these measures. A combination of cloze test and AUB test improved the predictability of language ability as indicated by writing scores. Indeed, the combination of any two of the three tests improved the predictability of the third. Furthermore, the cloze and writing tests appeared to measure in common some aspects of language ability beyond those that they share with the AUB test. These results have a number of theoretical and practical implications.

As already noted, the interrelationships indicate a degree of commonality among the tests as well as an independent contribution from each. Although this study was not concerned with factor analysis, this observation is in accord with a growing consensus regarding language testing, namely that models of language proficiency which combine general and specific factors provide the best explanation for language test data (Bachman & Palmer, 1982; Oller, 1983).

With respect to the residual association between the cloze and writing tests, the question arises as to the basis for this commonality.
One interpretation could be related to the predominantly integrative nature of the cloze and writing tests. From this viewpoint, cloze and writing require the student to draw upon several language skills simultaneously and involve complex processing of language while the focus is on content. Both tests also require production of language rather than mere recognition of correct items, although writing may be considered to include the communicative dimension more directly (Moller, 1981).

The commonality between cloze and writing may also be related to the testing of higher-order language abilities, which include the discourse-level factors of cohesion and organization. There has been some disagreement in the literature concerning this matter. Alderson (1979) in particular has claimed that cloze tests provide a measure of core linguistic skills of a relatively low order. Other researchers, however (for example, Bachman, 1982; Brown, 1983; Chihara, Oller, Weaver, & Chavez-Oller, 1977), have demonstrated that the cloze procedure can test not only lower-order linguistic skills, but also higher-level ability involving discourse constraints across sentences. In his study on cloze testing, Bachman (1982) used a rational deletion procedure (rather than systematic deletion at regular intervals) to ensure the inclusion of cohesive items. In the study reported in this article, cloze passages were carefully chosen so that blanks systematically included both syntactic and cohesive factors and therefore probably covered higher-order skills. The greater residual commonality between cloze scores and writing scores for the restricted sample of more advanced students, relative to the overall sample, lends further support to this conclusion.

At the practical level, the results of this study have three important implications. First, of the three tests, the AUB English Test was the most comprehensive, which was perhaps to be expected in view of its length and varied content. Thus, despite its shortcomings, a multiple-choice standardized test remains a valid, reliable, and comprehensive measure of language proficiency.

Second, since a combination of cloze and standardized ESL tests significantly increases information about a learner’s level of proficiency, a cloze component can be a valuable supplement to a standardized ESL test. Of course, composition writing can also be used as a supplement, but for large numbers of candidates, cloze clearly has the practical advantage of objective scoring. In either case, the combination of different types of tests may serve not merely to tap different aspects of proficiency but also to reduce bias which may arise from learner characteristics (Farhady, 1982).

Third, the inclusion of a cloze component in proficiency testing...
has a favorable by-product for language teaching. Since the task involved in a cloze test is integrative in nature, tapping overall abilities similar to those required in communicative language use, the incorporation of a cloze procedure can be expected to promote communicative language teaching in the classroom.

Although this study has demonstrated the effectiveness of cloze as a supplement to ESL testing, it is important to recognize that the cloze technique does not necessarily produce valid proficiency tests. Cloze passages must be carefully prepared, tried out, and validated. Furthermore, our experience with several cloze versions suggests that validity is related more to the individual items of the test than to general variables, such as rate of systematic deletion, a point which requires further study. We are now extending our analysis of the data to identify distinctive item characteristics which make for successful cloze testing.

ACKNOWLEDGMENTS
This work was supported in part by an Arts and Sciences research grant from the American University of Beirut. We are grateful to Mr. G. Miller, Director of the AUB Office of Tests and Measurements, for his kind cooperation and help with the project.

THE AUTHORS
Edith Hanania (Ph. D., Indiana University, 1974) is an Associate Professor in the Departments of English and Education at the American University of Beirut. Her research interests include language acquisition and language testing.

May Shikhani (M.A. in TEFL, American University of Beirut, 1983) teaches English in Lebanon.

REFERENCES


For professionals...
from professionals

Images and Options in the Language Classroom
Earl W. Stevick
An accessible discussion of how teaching can be found on what we know of learning and memory. Stevick defines options available to language teachers and presents criteria for evaluating them. Each chapter is filled with exercises that ask readers to draw on their own past experiences and do their own thinking before they come to the author's ideas.

Hardcover: 32150-6, $24.95
Paperback: 31281-7, $8.95

Approaches and Methods in Language Teaching
A description and analysis
Jack C. Richards and Theodore S. Rodgers
An overview and analysis of the major approaches and methods used in second and foreign language teaching, including Grammar Translation, the Direct Method, Situational Language Teaching, Audiolingualism, Communicative Language Teaching, the Silent Way, Community Language Learning, Total Physical Response, the Natural Approach, and Suggestopedia. The authors examine each method at the levels of approach, design, and procedure, which highlights the differences and similarities among the methods.

Hardcover: 32093-3, $24.95
Paperback: 31255-8, $8.95
REVIEWS

The TESOL Quarterly welcomes evaluative reviews of publications of relevance 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 VIVIAN ZAMEL
University of Massachusetts/Boston

Discourse Strategies (Studies in Interactional Sociolinguistics 1)

Work in theoretical linguistics in the past half century has largely amounted to defining successive levels of language structure, each one more abstract than the one before—phonetic, phonemic, syntactic, semantic, pragmatic, and, as in John Gumperz’s work, discourse-strategic levels of language. Applied linguists in turn have explored how linguistic or cultural differences at each new level create barriers to learning.

For empirical linguistics, the increased abstraction from phonetics to discourse strategy has meant that the method of observing structure must be supplemented by inferring tacit attitudes about language use within linguistic communities. For applied linguistics, the increasing abstractness of each newly defined level of structure has meant an increasingly sober estimate of what it will take to overcome, in Labov’s (1970) words, “the teacher’s ignorance of the student’s language as well as the student’s ignorance of the teacher’s language” (p. 1). When teaching across dialect or language differences is understood as requiring only an understanding of alternate pronunciations, it is easy to be optimistic. But when language difference is understood in terms of attitudes toward communication emanating from an unconsciously held ethnic identity, the prospects of pedagogy cannot seem as good.

John Gumperz’s work centers on the values carried by different codes in code-switching situations, which puts it in the purview of ESL instruction. In Discourse Strategies, Gumperz extends this study to conflicts of social attitudes toward communication in situations more subtle than those involving the switch from one
language (or dialect) to another. Gumperz examines ethnic conventions regarding language use itself: Who can speak when? What does it mean for one speaker to speak while another is concluding? What does a shift to the vernacular mean? What does a particular vernacular form mean (like check out)? When is refusal to cooperate a gesture of independence or etiquette rather than a blunt refusal? When is a literal threat a metaphor? Because the speakers Gumperz observes seem to have much in common and because their differences are so far from their explicit understanding, their failure to communicate may leave behind greater enmity than that raised in overt code conflict.

The differences which Gumperz describes cannot be “decoded” from the language itself: There is no one-to-one correspondence between sound and meaning. “Exclusive interaction with individuals of similar background leads to reliance on unverbalized and context-bound presuppositions in communication” (p. 71), Gumperz states. “Speakers who have little experience to the contrary often fail to account for the fact that others who do not share their communicative experience may also not have the background knowledge to interpret their speech as they themselves do” (p. 71).

For example, a black college student phones to explain missing a course deadline to a white instructor in whose office he had worked for several years. When the instructor answers, “Hello?” the student asks, “How’s the family?” As Gumperz describes it, more than the normal interval elapses before the instructor answers, “Fine.” Although the instructor responds to the news of the late paper, “That’s o.k., I can wait,” he will refuse to give the student a grade without seeing the finished paper. The student will be annoyed. Gumperz states: “He claimed that the telephone call had led him to hope he would be given special consideration” (p. 136).

Gumperz presents an instance of hurt feeling which emanated from a difference between British and American English in a flattery convention. The meaning of intonational stress in two codes plays the key role:

When a house painter arrived at the home of a middle class couple in California, he was taken around the house to survey the job he was about to perform. When he entered a spacious living room area with numerous framed original paintings on the walls, he asked in a friendly way, “Who’s the artist?” The wife, who was British, replied, “The painter’s not too well known. He’s a modern London painter named ——.” The house painter hesitated and then, looking puzzled, said, “I was wondering if someone in the family was an artist.” (p. 144)
Gumperz comments:

“Who’s the artist?” is a formulaic comment that fits a paradigm often uttered by Americans being escorted around a house. . . . Such formulas are often a conventionalized way of fulfilling the expectation that a complimentary comment be made upon seeing someone’s house for the first time. . . . The British wife in the above example was not familiar with this paradigm and its attendant routine, and therefore took the house painter’s question to reflect an objective interest in the paintings. . . . we need to know how the formulaic nature of utterances is signalled. In the example given here, there are both extralinguistic and linguistic cues. The extralinguistic signals lie in the setting and the participant’s knowledge of what preceded the interaction. There are at least three linguistic signals: first, the semantic content; second, the syntactic paradigm; and third, the contextualization cues such as prosody (e.g., the stress and high pitch on the first syllable of “artist,” and its marked high falling intonation). . . . Formulaic use of language is always a problem for non-native speakers. It is perhaps even more of a danger, however, between people who ostensibly speak the same language but come from different social or regional backgrounds. Since they assume that they understand each other, they are less likely to question interpretations. (pp. 144-145)

The phenomena Gumperz describes are particularly sobering because they intensify with the sophistication, the education and sensitivity, the good intention and exasperation of the speakers. As an example of problems of interethnic communication, Gumperz presents a conversational tangle between a British staff member of an agency (B) and a British-educated Pakistani teacher of mathematics (A). The following exchange took place well into their conversation:

A: Um, may I first of all request for the introduction please
B: Oh yes sorry (pause, about 1 second)
A: I am sorry
B: I am E.
A: Oh yes (breathy) I see . . . oh yes . . . very nice
B: and I am a teacher here in the Center
A: very nice
B: and we run (pleased to meet you (laughs) different
A: classes (A laughs) yes, and you are Mr. A?
B: N. A.
What is most evident is the interlocutors’ lack of agreement on the direction and pace of the conversation. Gumperz notes the oddity of A’s final contribution to this exchange, “Would it be enough introduction?” and relates it to the lack of “coordination of speaker’s messages with backchannel cues such as ‘um’ [and] ‘yes’ “ (p. 176).

For Gumperz, the lack of conversational synchrony evident in odd messages and in failures of cueing, gaps, and overlaps expresses the speakers’ discomfort and failure to cooperate. Gumperz notes the contributions of ethnic speech mechanisms to the conflict. For example, “very nice” is a translation of Urdu *buhut uccha*, a backchannel sign of interest similar to our “O.K, go on.” It is not equivalent to “very nice” as a Western speaker of English would use it, as a “response to children who behave properly” (p. 179).

It is difficult to imagine more well-intentioned and well-educated humans speaking to each other; if conflicting discourse strategies undermine their communication, the prospects will not improve for speakers who have never imagined at all that different cultures have different customs of language use. Rather than trying harder and making the gaps and overlaps in conversation more acute, these naive speakers would simply walk away in anger or bewilderment. Such communicative failures can create impossible classroom situations. But there is no reason to assume that abstract, largely tacit differences are inscrutable and unalterable. Gumperz himself states, “With teachers whose teaching is responsive to student cues, who are tolerant of different communication styles and who succeed in setting up . . . rhythmic teacher-student exchanges, dialect speakers do as well as others in achieving control of expository English prose” (p. 144).

In the face of insights about the powerful determining effects of abstract ethnic discourse rules, a type of common sense should prevail about the pedagogical implications of evolving linguistic theory. The theory has developed in a certain manner—toward deeper, more powerful, less manageable aspects of language, more intrinsically psychological or, as in Gumperz’s case, social. This evolution makes the teaching situation seem more and more predetermined and unalterable. But the phenomena which these new insights attempt to explain have occurred all along, unseen or misnamed, and are not new to the classroom.

Moreover, linguistic theory is most suited to discover structural differences and, therefore, potential conflict. Except for the study of variability in work such as Labov’s (1972), linguistics is not
predisposed to observe successful efforts toward improved human contact and change—toward overcoming cultural and linguistic boundaries. Linguistics is not well suited to observe the dissolution of the differences it observes.

The existence of studies such as Gumperz’s is evidence that the more subtle and abstract features of language are knowable. Field linguistics has given us the method for decoding those aspects of language in which meaning is related directly to the structure of speech sounds. Strategies for attaining insight into more tacit structures like those described by Gumperz need to be made explicit—and not only for researchers but for teachers working in mixed classrooms.

I would suggest that a possible source of such a strategy may be the attempts by psychoanalysts to interpret and use their own seemingly unmotivated emotional responses to their patients. The differences between linguistics and psychoanalysis, and between the conduct of psychotherapy and teaching, are obvious, but these disciplines share the task of identifying and then analyzing types of communication which are not overtly structured in speech.

For example, Langs’s (1978) *The Listening Process* is a set of training seminars in which training analysts are grounded in techniques for monitoring their patients’ least explicit communications; it may therefore be well suited to helping teachers infer subliminal discourse strategies at play in their classrooms.

What is most striking in Langs’s (1978) approach is its recognition of a paradox: Although the trainees respond to their patients frequently and incorrectly during the clinical hour, they tend to be oblivious of the effects of their interventions on their patients. Langs’s method of “listening” emphasizes that analysts should intervene much less, that they should “contain” disconcerting or hostile projections from patients, “digest” them, and interpret them for diagnosis and for the conduct of the therapy. Langs encourages his trainees to develop “silent hypotheses” in place of interventions and to trust the patient to arrive at the intended intervention in due course.

In terms of psychotherapeutic method, Langs’s strategies complement techniques which Curran (1976) adapted from counseling to second language teaching. Curran uses these techniques to open natural, humane communicative channels between teacher and student. As such, Curran combats the tendency of second language learning to find formalistic, artificial channels shaped by rigid conventions of academic study or second language learning in general. Curran’s strategy assumes that teacher and student can free themselves sufficiently from their cultural
discourse strategies long enough to establish a more humane channel.

Langs, then, presents strategies which teachers might use to teach on the trans-cultural field of discourse which Gumperz helps us to understand. What’s more, Langs’s strategies can help teachers to tailor their actions from day to day to the extent of their evolving understanding of this field. If a therapist can maintain communicative fields for deeply disturbed patients, a teacher can maintain similar fields for those students whose discourse strategies the teacher does not yet understand. With a field of communication established, teachers can continue to analyze those strategies by listening closely to the students, attending to their own response, forming silent hypotheses, and seeing them invalidated or validated in the next classroom exchange.

REFERENCES


NEAL BRUSS

*University of Massachusetts/Boston*

*The Input Hypothesis: Issues and Implications*


There is a Monty Python routine in which a radio interviewer tries to get Miss Ann Elk, a dinosaur expert, to explain her new theory about the brontosaurus. After a great deal of hemming and hawing, false starts, and general time wasting, we are finally told this: Brontosaurus were very thin in the front, much, much thicker in the middle, and then very thin again at the end. Most of us would agree that, as a theory, this is rather unsatisfactory (indeed, the interviewer shoots Miss Elk before she can tell us her second theory). But then it was not meant to be taken seriously as a theory.

Reading *The Input Hypothesis*, which evidently is meant to be taken seriously, brings Miss Elk to mind. *The Input Hypothesis* is
the latest in a series of books and articles in which Krashen pretty
much repeats what he has said in all the other books and articles;
that is, he offers “what I call, perhaps audaciously, a theory of
second-language acquisition” (p. vii). (There are perhaps more
fitting words than “audaciously”; and in fact Krashen usually drops
the article and talks simply of second language acquisition theory, a
location that makes the complex error of suggesting that his theory
is a theory, that a second language acquisition theory exists, and that
his theory is it.) As just about everyone knows by now, Krashen’s
theory is comprised of five hypotheses. This book, however,
focuses on what is probably the most important of the five,
presumably in an attempt to explain and defend it in greater depth
than heretofore.
The book has three chapters: one describing the hypothesis and
offering putative evidence in its support; one dealing with various
problems with, and challenges to, the hypothesis; and one on
implications for the language teacher. I have criticized the
shortcomings of Krashen’s theory in some detail elsewhere (Gregg,
1984), and I wish to repeat myself here as little as possible. Rather,
I would like to concentrate on the Input Hypothesis as part of an
ostensible acquisition theory, specifically, to show how it reflects an
ignorance of the nature and goals of linguistic theory and of
language acquisition theory.
The Input Hypothesis states that “humans acquire language in
only one way—by understanding messages, or by receiving
‘comprehensible input’” (p. 2). There is more than a touch of Miss
Elk here. On the one hand, no one has suggested that comprehen-
sion is not necessary for acquisition. On the other hand, and more
important, Krashen does not claim that comprehensible input
causes acquisition, which claim is necessary if he wishes to rescue
his hypothesis from vacuousness. And if he did claim that
comprehension causes acquisition, he would then, of course, be
obligated to try to show how. After all, it does not amount to a
scientific hypothesis to assert that apples fall from trees because of
gravity or that birds fly south by instinct. According to Krashen,
there is no fundamental difference between first and second
language acquisition; that is, input is a sufficient condition for
acquisition. But in first language acquisition theory, that very fact is
the problem to be explained; it is not the explanation. And an
acquisition theory is supposed to explain acquisition.
The vacuity of the hypothesis makes Krashen’s “evidence” largely
irrelevant; the Input Hypothesis is consistent with just about any
evidence possible. Caretaker speech (CS), for instance, is
(allegedly) simplified input, hence comprehensible, hence evidence
for the hypothesis. But Ochs (1982) and others point out that there are cultures without CS. No problem! This just proves that they are providing comprehensible input in some other way. Ochs’s Samoan caretakers, for example, “provide repeated exposure to language they expect the child eventually to understand” (p. 7). The same, of course, could be said of audiolingual syllabuses.

Again, after saying that the CS is simplified, Krashen then blithely welcomes the findings of Gleitman, Newport, and Gleitman (1984) that children require a rich input—that is, input that is not simplified—and goes on to say, “Rich input provides the acquirer with a better sample to work with, more opportunities to hear structures he is ready to acquire” (p. 27). But Krashen’s Natural Order Hypothesis claims that “we acquire the rules of language in a particular order” (p. 1). Thus, the richer the input, the less likely the occurrence of any given “structure” in that input, and thus the longer the child will have to wait for it. In other words, despite Krashen’s denials, his theory predicts that fine-tuning of input would facilitate acquisition. Of course, there is in fact no fine-tuning for children (although there possibly could be for adults), so the point is moot; I raise it only to show how Krashen wants to have it both ways. The fact remains that CS has no bearing as evidence on the Input Hypothesis.

Of course, it is possible that Krashen has a narrower idea of comprehensible input in mind. The phrase “by understanding messages” suggests that input that is not in the form of a message (however defined) will not qualify as input. Thus, studying a conjugation chart presumably will not help. This seems like a plausible interpretation, especially since Krashen claims that “learning” cannot become “acquisition.” But on the other hand, he talks (pp. 46-47) of acquiring deviant forms, for example, in EFL situations where there is little comprehensible input. Japan is a case in point: Instruction typically is entirely in Japanese, so there are not many messages in English. And yet forms are acquired, which sounds rather like learning becoming acquisition. Or else, if it is not, then comprehensible input is not confined to true “messages” narrowly defined.

The Input Hypothesis claims that we move along the Natural Order by comprehending input containing the “next” rule. It is odd that in a book devoted to the Input Hypothesis, Krashen does not bother to describe this process, instead simply referring us to “Newmark’s ‘Ignorance Hypothesis’ and Current Second Language Acquisition Theory” (Krashen, 1983). There we are given an example of a “rule” that is at “i + 1,” next in line to be acquired: the past tense of sweep. This is an interesting example, for it suggests
how feeble a grasp Krashen has on the meaning of rule. If swept is a rule to be acquired and it is to be acquired in the Natural Order, one wonders where it might be in relation to, say, broke or drove. What kind of order could this be? What could possibly be natural about it? And even if we take less embarrassing examples of rules—such as the good old morphemes that gave us the hypothesis in the first place—the order is a phenomenon to be explained, not just appealed to.

Neither in the present book nor in the 1983 article—nor anywhere else, for that matter—is “i + 1” coherently defined. If “i” is supposed to refer to the learner’s competence at a given time, then it cannot be a rule. Therefore, it would be a category error to talk about comparing “i” with “i + 1” (e.g., Krashen, 1983, p. 140). (This has nothing to do with whether or not “i + 1” is operationalizable; see p. 68.) This vagueness in the use of words like competence and rule is not simply intolerably sloppy; it also reflects a profound misunderstanding of linguistic theory and its connection with second language acquisition theory.

For instance, Krashen attacks the “strong interface position” on the question of whether learning becomes acquisition—that is, the position that acquisition is always preceded by learning (a straw man, incidentally). If this position were correct, “language teaching [would be] truly ‘applied linguistics,’ completely dependent on research in formal linguistics: linguists discover a rule, . . . teachers teach it, and students learn it” (p. 39). But no linguist in the world is looking for new English irregular verbs or grammatical morphemes, and no teacher in the world is trying to explain subjacency or the move-alpha rule to language learners.

Krashen is confusing two different kinds of rules, and this confusion renders illegitimate his use of Chomsky to endorse his learning/acquisition distinction. For instance, Krashen quotes Chomsky as follows (the bracketed phrase is Krashen’s): “there is little doubt that [rules learned from a book] could not be consciously applied, in real time, to ‘guide’ performance” (p. 25, quoting Chomsky, 1975, p. 249). But this quotation is not faithful to the original: “[rules learned from a book]” refers to the Specified Subject Condition, a putative part of Universal Grammar (UG)—that is, a rule not learned from any book.

Actually, for what it is worth, Chomsky is not only not endorsing Krashen’s position, he is implicitly supporting a weak interface position. For instance, he characterizes a good traditional or pedagogic grammar as being in effect “a structured and organized version of the data presented to a child learning a language” (Chomsky, 1985, p. 15). In fact, in the very note that Krashen quotes
from, Chomsky (1975) states that “people learn language from pedagogic grammars by the use of their unconscious universal grammar” (p. 249).

No one is denying the essential role of UG (except the mythical strong interface), but after all one does not study a pedagogic grammar unconsciously; that is, it sounds as if Chomsky does believe that learning can become acquisition. And if Chomsky’s cognize/know distinction is really the same as Krashen’s acquisition/learning distinction, as Krashen seems to claim (p. 24), Krashen can take little comfort from Chomsky’s (1980) claim that “the irregularities [of language] are learned” (p. 238). (Chomsky’s “irregularities” are Krashen’s “rules”: the various specific details, such as -ing, that distinguish one language from another.) Mind you, Chomsky may be absolutely wrong on this point; he is no expert. My point is simply that Krashen seems not to understand the Chomsky research program and hence its relevance, if any, to his theory.

Having brought the strong interface position to its knees, Krashen goes on to consider a weaker position, namely that learning can become acquisition in at least some cases. Krashen rejects this position (in favor of a no-interface position) on the grounds that it violates Occam’s Razor (“Entities are not multiplied beyond necessity”). Here we have another misunderstanding. Occam’s Razor is a principle of theory construction that bars the use of unnecessary constructs. (For instance, Krashen’s Output Filter, introduced here for the first time [pp. 44-46], violates Occam’s Razor, since it deals with a performance phenomenon, while Krashen’s theory is a competence theory.)

The claim that conscious knowledge of a rule may in certain cases be acquired is not a theoretical construct but a statement of fact, so far as anyone can tell; certainly no one has ever shown a shred of evidence against it. The fact that such acquisition is not necessary is neither here nor there. Or would be neither here nor there, except that Krashen’s Acquisition/Learning and Monitor Hypotheses claim that learning cannot become acquisition, In fact it is not the weak interface position but rather these two hypotheses—specifically, the Monitor construct—that must pass the test of Occam’s Razor.

It is worth noting that Krashen’s allegiance to Occam’s Razor does not prevent him from accepting a weak interaction position on the question of whether face-to-face interaction is necessary for acquisition (pp. 33-34). In other words, Krashen allows that while in theory one could learn a language without interaction, interaction can be helpful. This closely resembles the claim that while in theory one could acquire a language without any conscious learning,
learning can be helpful. The only difference is that the interaction question does not threaten the meretricious elegance of Krashen's theory.

Constructing a cogent theory of language acquisition is very, very difficult, which is one reason why no one has ever done it. Krashen has not even tried, as we can see more easily if we look at attempts being made in first language acquisition theory (e.g., Pinker, 1984; Wexler & Culicover, 1980). Thus, it is disturbing to see how well-received the theory seems to be.

Krashen himself indirectly suggests a possible reason when he says (pp. 58-59) that teachers in elementary and adult education are more taken with his theory than are teachers at the university level. To this, I would add that American teachers seem to be more receptive than Europeans. I can think of a couple of possible explanations. For one thing, university teachers are better educated in the relevant areas and also have a good deal more leisure to study Krashen’s writings critically. For another, Europeans are less infected by the anti-intellectualism that afflicts American elementary and secondary education (see, e.g., Hofstadter, 1962; for a reflex of this attitude in TESL, see Moskowitz, 1978). For the fundamental message of Krashen’s theory is that you do not have to know very much to be a good language teacher.

Krashen’s ideas have been around for almost a decade now; one of the noteworthy things about them is how little they have changed. Krashen has had plenty of opportunity to try to rescue them from their incoherence and shape them into something like a real theory, but what we have before us is almost identical to what we were offered years ago, with all its insufficiencies and contradictions intact. The Input Hypothesis offered Krashen a chance to give us a cogent elaboration of the linchpin of his theory; unfortunately, he has muffed the chance.

REFERENCES


REVIEWS


KEVIN R. GREGG
Matsuyama University, Japan

Support system for learning to write in English

Guided Composition Second Edition
Florence Baskoff, American Language Institute, New York University
271 pages • paper • Instructor’s Annotated Edition • 1984

Baskoff’s text for ESL/EFL students concentrates on writing paragraphs and short compositions. Each chapter contains a model composition that focuses on a rhetorical form, an organizational principle, and a grammatical topic. Exercises isolate specific grammatical and lexical problems for the non-native speaker and provide extensive practice in sentence writing. Composition assignments are based on the form and content of the model composition.

The Second Edition includes many new model compositions and dictations, an expanded handbook of English grammar and rhetoric, and a new Instructor’s Annotated Edition.

For adoption consideration, request an examination copy from your regional Houghton Mifflin office.

13400 Midway Rd., Dallas, TX 75244-5165
1900 So. Batavia Ave., Geneva, IL 60134
988 Lenex Dr., Lawrenceville, NJ 08648
777 California Ave., Palo Alto, CA 94304
Another Look at Passage Correction Tests

TERENCE ODLIN
Ohio State University

A passage correction test (PC) is a measure used to determine the ability to identify and correct errors that have been systematically inserted into a prose passage. Earlier studies of such tests have made use of two basic formats: the insertion of irrelevant words as distracters (e.g., Bowen, 1978; Davies, 1975) and the substitution of grammatically unacceptable forms for acceptable ones (e.g., Arthur, 1980). Performance on PCs of the former type has been shown to correlate significantly with other measures of EFL ability (Mullen, 1979), and the format used in Arthur’s study produced somewhat similar results. Other research (Abraham, 1983; Kaplan & Shand, 1984) suggests the value of editing tasks similar to PCs for the study of cognitive style and affective variables in second language learning.

PCs that are constructed through the substitution of ungrammatical forms for grammatical ones have been investigated less; however, they may be more valuable for the study of how some aspects of language develop, for example, how much the ability to edit one’s writing grows along with other abilities. While PCs constructed through the insertion of irrelevant words are a promising tool for some types of research, they do not so closely resemble the everyday task of editing one’s writing, since they only require individuals to detect forms randomly dispersed among acceptable structures.

The purpose of the study reported here was to examine in detail some characteristics of PCs as measures of editing skills. Specifically, its aims were (a) to show that success on a PC constructed by substituting unacceptable forms correlates significantly with a rather different type of language test, Harris and Palmer’s (1970) Comprehensive English Language Test (CELT); (b) to show that sensitivity to different types of errors correlates significantly with the CELT; and (c) to show that correlations among different types of PC items can distinguish patterns of developing awareness of forms and meanings in EFL.
THE STUDY

Subjects

The subjects, who were volunteers, included 25 individuals in the Intensive English Program (IEP) and 20 native speakers of English enrolled in an undergraduate composition course at the University of Texas in Austin. With the exception of 2 Japanese and 1 Arabic speaker, all of the IEP subjects were native speakers of Spanish.

The IEP subjects, whose ages ranged from late teens to early 30s, had all taken the CELT and been placed into one of six levels of ability. While the number of subjects is not large in relation to the number of levels in the IEP, the subjects were quite evenly distributed: 8 in Level 0 or 1, 9 in Level 2 or 3, and 8 in Level 4 or 5.

Format and Items

Since the PC designed for the study (see Figures 1 and 2) was intended to establish levels of ability ranging from minimal to advanced, the vocabulary in the passage was generally restricted to what would appear in textbooks for a beginning ESL/EFL course. Most of the words used in the test appear in the first two books of the revised Lado English Series (Lado, 1978). After the passage had been written, 28 errors, 7 of four different types, were substituted for the original words.

Figure 3 lists the four error types: lexical (LEX), grammatical (GRM), polarity (POL), and distributional (DIS). Grammatical errors include several types of common morphosyntactic and syntactic errors made by EFL students (irregular forms, agreement, verbal inflection, etc.).

Polarity (Clark, 1973) involves the dimensional information governing the antonymic relationship between two spatial terms. For example, both up and down involve vertical polarity; for purposes of classification, up can be considered the positive member of the polarity relationship and down the negative member. Similarly, above, over, and high can be construed as spatial terms with positive vertical polarity; and below, under, and low as spatial terms with negative vertical polarity.

Distributional errors are those involving an anomalous part-of-speech classification (Odlin, 1983; Odlin & Natalicio, 1982). For example, “airplane . . . flying up the house,” the example in Figure 3, originally read “airplane . . . flying over the house.” The POL and DIS errors were included to examine in more detail the semantic and syntactic features of dimensional terms previously studied by Odlin and Natalicio (1982).1

1 Without the picture accompanying the test, the LEX and POL errors in Figure 3 may not seem to be anomalous at all; it is at least plausible for water to come from a chimney (on a rainy day) or for a cat to walk above a flag (on a ledge over a flagpole).

Thus, while there are important differences among the item types, there is also a fundamental similarity between LEX and POL items on the one hand and between GRM and DIS items on the other hand: The former two item types require much more dependence on the information in the picture.
Instructions: Underline the errors in the following paragraph and correct them. Only individual words—not phrases or sentences—need revision. Two examples of correct revision are given in the first sentence. Use the picture to help you revise.

A Robertson family is at church. Two girls are playing in the living room which is above the basement. The children are stand next to the fireplace on which there are two candles above a picture of a woman. The picture is over the fireplace and have a dark frame. A fire is burning in the fireplace, and water is coming out the chimney. A airplane is flying up the house. It has a star on its tail. On the ground down is a high mountain with snow at the top. A man standing on the floor of the house is looking at the mountain. An American flag is flying from the balcony. Above the flag a cat is walking under the stairs while a dog is walking over. Outside a house on the other side stands a bush with a boy, Tom Robertson, on a down branch. Tom can look up and see her sisters playing in the living room. They are playing next to a table low a bookcase with two shelf. The boy cannot see the man and the woman, Mr. and Mrs. Robertson, in the room up the living room because he is not low enough. Mrs. Robertson is sitting in a chair, and she and her husband are talking with each other. A girl is looking out of a door in another room above. She can see the dog and cat above outside, but she cannot see the two children who are in the living room below the stairs leading to the basement. In the basement are a number of brooms, four of them over three other.

Note: LEX = lexical; GRM = grammatical; POL = polarity; DIS = distribution
Administration and Scoring

To forestall any ceiling effects with the EFL students, the PC was first given to the native speakers to determine a suitable length of time for any learner whose ability was close to that of a native speaker to complete the test. Since no native speaker needed more than 15 minutes to complete the test, a 15-minute time limit was imposed on the IEP subjects. It was useful in a few cases to provide an explanation in Spanish for some of the subjects in the lowest IEP levels (which consisted only of Spanish speakers), but most subjects had no difficulty in following test instructions. Subjects were not told how many errors were in the passage.

The performance of the native speakers revealed that two items, 17 and 28, were invalid. In the case of Item 17, the picture accompanying the test passage (see Figure 2) was ambiguous: The boy could be looking up or down. Thus, while some native speakers changed _up_ to _down_ (as was expected), many did not. Item 28 was deemed invalid because a considerable number of native speakers failed to change _other_ to _others_. Therefore, only 26 items were scored. The rare instances in which native speakers failed to note or correct other errors in the PC were interpreted as nothing more than oversights.

Of the several scoring systems applied to the data, three correlated highly with each other (from .95 to .98, \( p < .01 \) in all cases) and also with the CELT (from .83 to .87, \( p < .01 \) in all cases). All subsequent results are reported in terms of a system that counted not only successful revisions but also identifications of errors where no correction was attempted. That system had the highest correlations with the other systems and with the CELT, and the Kuder-Richardson reliability coefficient KR-20 calculated according to that system was .90.

RESULTS AND DISCUSSION

In addition to correlating highly with the CELT(.87), there is evidence that the PC can distinguish native from nonnative linguistic abilities as well as patterns of developing awareness of forms and meanings in EFL. The difference between native and nonnative abilities is
quite clear. Although the 8 most advanced students (those in Levels 4 and 5) had much higher scores than those in the less advanced levels, their scores were far below those of the native speakers: The 8 EFL students made (on average) only 13.8 identifications or corrections, whereas the native speakers made 23.1.

Individuals' success on the four item types (Figure 3) suggests that while an awareness of each type of error is a predictor of general EFL ability, the awareness of certain semantic errors is an especially revealing predictor. The value of the four item types as predictors can be seen in Table 1, a summary of a stepwise multiple regression performed on the EFL data (with IEP CELT scores as the dependent variable). The results show that item types treated as independent variables have significant Pearson product-moment correlations with the CELT (Simple \( r \) column) and that combinations of independent variables have even higher correlations (Multiple \( r \) column).

### Table 1

<table>
<thead>
<tr>
<th>Error type</th>
<th>Simple ( r )</th>
<th>( R^2 )</th>
<th>Change in ( R^2 )</th>
<th>Multiple ( r )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>POL</td>
<td>.811</td>
<td>.658</td>
<td>.658</td>
<td>.811</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>GRM</td>
<td>.669</td>
<td>.716</td>
<td>.058</td>
<td>.846</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>LEX</td>
<td>.588</td>
<td>.758</td>
<td>.042</td>
<td>.871</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>DIS</td>
<td>.643</td>
<td>.758</td>
<td>.000</td>
<td>.871</td>
<td>&lt; .01</td>
</tr>
</tbody>
</table>

Note: POL = polarity; GRM = grammatical; LEX = lexical; DIS = distribution.

Figures in Table 2 indicate that distinct types of metalinguistics awareness are put to use on the test. For example, the nonsignificant correlation between LEX and GRM suggests that a good ability to correct vocabulary errors does not necessarily imply much of an ability to correct grammatical errors. In contrast, success on POL items, which have a prominent semantic role in the passage, correlates significantly with success on every other type. Moreover, the high (.811) correlation of POL scores with the CELT (see Table 1) suggests that sensitivity to errors involving forms with a prominent semantic role in a piece of discourse can predict more general language abilities. Such results are consonant with recent research by Kaplan and Shand (1984), Odlin (in press), and others on relationships between metalinguistics awareness and communicative competence.

The relationships among the item types suggest that PCs are a useful tool for second language acquisition research. Although they resemble other tests in a variety of ways, PCs offer a unique combination of advantages. Like grammaticality judgment tests, PCs require individuals to use a simple metalinguistics procedure: judging the appropriateness of
linguistic forms. Unlike grammaticality judgment tests, however, PCs provide subjects with a coherent discourse. Since PCs do not depend on one’s ability to use the information in an isolated sentence to imagine a universe of discourse, they have a distinct advantage over grammaticality judgment tests (see Chaudron, 1983).

Like dictations, PCs require a holistic processing strategy, since subjects must consider every word, every phrase, and every sentence as a unit within some larger unit (see Oller, 1973). Unlike dictations, they can be scored with systems that are relatively easy to devise, and once a system is established, scoring can be done rather quickly (from 5 to 10 minutes for a test with two dozen or so items).

<table>
<thead>
<tr>
<th>Error type</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. LEX</td>
<td>.31</td>
<td>.50*</td>
<td>.24</td>
</tr>
<tr>
<td>2. GRM</td>
<td>—</td>
<td>.59**</td>
<td>64**</td>
</tr>
<tr>
<td>3. POL</td>
<td>—</td>
<td>—</td>
<td>.74**</td>
</tr>
<tr>
<td>4. DIS</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: LEX = lexical; GRM = grammatical; POL = polarity; DIS = distribution. * p < .05. ** p < .01.

Like cloze tests, PCs use an ordinary prose passage that can be easily transformed to test for specific structures (see Oller & Inal, 1971). Unlike cloze tests, they measure a subject’s sense of the normality of a phrase: Any subject who consistently ignores something anomalous clearly has a certain tolerance that a cloze test cannot indicate, since a blank space in a cloze test draws a subject’s attention to the phrase in which the blank is inserted.

The disadvantages of the PC format are relatively minor: Scoring takes a little more time than it does on grammaticality judgment tests and cloze tests (though not nearly as long as on dictations). Also, scoring a PC may not be as straightforward as scoring a grammaticality judgment test, since a variety of responses may sometimes be acceptable. (In that difficulty, PCs resemble cloze tests and dictations.)

Recent research on error detection suggests that systematic attention to certain types of errors can help students in writing classes learn to edit their papers for such errors (Lalande, 1982). Additional evidence suggests that passage correction is a useful activity to develop such systematic attention. Some PC formats have been found to be valuable as classroom or writing-lab exercises to promote consistent monitoring of errors such as sentence fragments and violations of number agreement and of tense consistency (Odlin, 1985). Indeed, while PCs show promise as measures
of metalinguistics awareness and other language abilities, their greatest
value may be for the teaching of editing and related writing skills.

REFERENCES

Abraham, R. (1983). Relationships between use of the strategy of monitoring and

Arthur, B. (1980). Gauging the boundaries of second language competence: A

Bowen, J. D. (1978). The identification of irrelevant lexical distractions: An editing
task. *TESL Reporter*, 12 (1), 1-3, 14-16.


Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65-

Davies, A. (1975). Two tests of speed reading. In R. Jones & B. Spolsky (Eds.),
*Testing language proficiency* (pp. 119-127). Arlington, VA: Center for Applied
Linguistics.

McGraw-Hill.

Eckman, L. Bell, & D. Nelson (Eds.), *Universals of second language acquisition*


Language Journal*, 66, 140-149.

Schachter (Eds.), *On TESOL ’79* (pp. 187-192). Washington, DC: TESOL.

doctoral dissertation, University of Texas at Austin.

presented at the 19th Annual TESOL Convention, New York.

Review of Applied Linguistics*.


Oller, J. W., Jr. (1973). Discrete-point tests versus tests of integrative skills. In J. W.
Oller, Jr., & J. Richards (Eds.), *Focus on the learner: Pragmatic perspectives for

Quarterly*, 5, 315-326.

Author’s address: Department of English, The Ohio State University, 164 West
17th Avenue, Columbus, OH 43210
Previous research on anxiety and foreign language learning (see Scovel, 1978, for a full review of the literature) has focused primarily on the effects of anxiety on overall proficiency in a second language, which is typically measured by discrete-skills tasks or end-of-course grades. However, such measures of proficiency are likely to obscure some of the more subtle effects of anxiety on second language performance. For example, anxiety might affect the content and elaboration of second language speech as well as overall fluency and grammaticality.

Indeed, research on the effects of writing apprehension has found that native-speaking students with higher levels of writing anxiety write shorter compositions, use less intense words, and qualify their writing less (Daly, 1977; Daly & Miller, 1975). If nonanxious second language students are more apt to attempt ambitious topics which require more complicated explication than their level of proficiency permits, they may actually appear to be less proficient than students whose anxiety restricts them to safer topics. Yet the nonanxious students may be the ones communicating at the higher level.

Source of variation in the content of second language performance is, however, a relatively unexplored topic. For instance, Kleinmann (1977) found that the grammatical structures used by ESL learners varied with their level of facilitating anxiety; the informational content of their language was not examined, however.

This study (Steinberg, 1982) explored the effect of induced anxiety on the content of oral descriptions, in a second language, of stimulus pictures. It was hypothesized that subjects undergoing an anxiety treatment and those undergoing a nonanxiety treatment would be differentiated by the proportion of interpretive to denotative content in their descriptions, with the anxiety group responding less interpretively. Since the study dealt with environmentally manipulated anxiety, it addressed an area readily susceptible to the intervention of the classroom teacher, that is, the atmosphere provided for student communication.

METHOD

Subjects

Twenty Spanish-speaking young adults enrolled in an intensive ESL program at the University of Texas at Austin volunteered to serve as research subjects. All were students at the low-intermediate level who
agreed to participate in a study of “the way people speak in another language.” They were not informed that the study would focus on their responses to anxiety.

To control for proficiency biases, evaluations of each subject’s current oral ability in English were obtained from the subject’s classroom teacher. On the basis of these evaluations, subjects were placed in either a high- or low-proficiency cell and were then randomly assigned by cell to the two treatment conditions.

**Procedure**

Subjects were interviewed individually by the same researcher in an empty classroom; all were informed of the presence of an audio recorder. The task consisted of describing in English three pictures (Numbers 2, 8BM, and 5) from Murry’s (1935-1943) Thematic Apperception Test (TAT). The subjects were asked to address three specific areas in their descriptions: (a) the elements in the picture, (b) the actual events depicted, and (c) what the subjects imagined to be happening in the picture. Thus, subjects were to respond with both objective information and their subjective interpretations.

The TAT pictures were chosen because their ambiguity is well suited for the elicitation of interpretive as well as denotative material; in addition, their availability permits replication by other researchers. As a control for possible vocabulary problems, words basic to each picture were provided on a piece of paper and their referents indicated in the picture. Subjects could also request additional vocabulary from the researcher. All interviews were audio-recorded.

A Spanish language version of the Anxiety scale of Zuckerman and Lubin’s (1960) Multiple Affect Adjective Checklist (MAACL) was administered as a check on the effectiveness of the experimental conditions. Upon completion of the experimental task, subjects were given the MAACL and instructed to check off all adjectives which described how they felt at that moment.

**Treatments**

**Anxiety condition.** To foster a stressful environment, the experimenter pointed out the presence of audio as well as video recorders, trained a video camera on the subject, and conspicuously played with the controls during the interview. The subject was brusquely shown to a seat at a narrow lecture desk, several feet distant from the experimenter, who maintained a cold and official posture toward subjects in the anxiety group. Task instructions were stress-loaded by emphasizing that the interview was an indicator of basic English skills and that good performance was crucial to the success of the experiment. However, in accordance with human subject guidelines, all subjects were also informed...
that the experiment was in no way connected to their academic institution and that the results would be confidential.

**Nonanxiety condition.** The subjects receiving this treatment sat in a comfortable armchair and were not subjected to the presence of a video camera. The warm, personal manner of the experimenter toward these subjects was also designed to reduce stress. She greeted them at the door, exchanged a few pleasantries before beginning the task, and maintained a smiling and supportive presence throughout the interview. Finally, the task instructions to the subjects in the nonanxiety condition emphasized that while it was hoped that the subjects would perform to the best of their ability, the experience was supposed to be interesting and enjoyable for them and they were not to worry about being evaluated.

**Analysis**

The audio-recorded interviews were evaluated by three native raters, all experienced ESL teachers. After a brief training session, the raters were instructed to determine the proportionate amounts of denotative and interpretive information provided in each interview and to indicate their judgments along a scale (see Figure 1). Denotative responses were those referring to actions and elements clearly shown in the TAT pictures; interpretive responses were those containing projective references to events not specifically depicted in the instrument.

**FIGURE 1**

Rater Instructions

*Please rate each picture description according to the amount of denotative or interpretive material it contains.*

4. Performance is heavily loaded with personal interpretation of picture, going beyond the elements actually present.

3. Performance contains a significant, but not striking, amount of interpretation. The amounts of denotative and interpretive material are approximately equal.

2. Most information is denotative, with a few interpretive elaborations.

1. Communication is almost entirely denotative; almost no interpretation is provided.

<table>
<thead>
<tr>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>2.5</th>
<th>3</th>
<th>3.5</th>
<th>4</th>
</tr>
</thead>
</table>

The subjects received a score for their description of each picture; these scores were added to yield one score per subject per rater. Finally, the scores were converted to z-scores and summed across raters to yield a
Student Response Style score. This procedure controls for differential use of the scales by the raters. Thus, each subject ended up with a single composite score reflecting performance on the three communicative tasks as judged by the three raters.

RESULTS

Table 1 displays the means and standard deviations of the unstandardized (raw) and standardized Student Response Style scores. A $t$ test of significance was applied to the group means; the resulting $t$ value was significant: $t (18) = -2.02, p < .03$ (one-tailed test). Thus, the hypothesis that anxiety-group members would respond less interpretively than their nonanxiety-group counterparts was supported.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Unstandardized scores</th>
<th>Standardized scores</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>2.38 0.58</td>
<td>-0.64 1.60</td>
<td>-2.02*</td>
</tr>
<tr>
<td>Nonanxiety</td>
<td>2.86 0.48</td>
<td>0.64 1.30</td>
<td></td>
</tr>
</tbody>
</table>

Note: Higher numbers indicate a more interpretive response style; unstandardized scores are divided by nine (3 raters x 3 tasks) to convert to scale units.

*$p<.05.$

A manipulation check on the treatment effects was computed as a design control. The Pearson product-moment correlation coefficient between the Spanish version of the MAACL and the experimental conditions was $r = .51, p < .01$, indicating that the anxiety treatment was moderately successful. Sixty percent of the subjects in the anxiety condition reported being anxious, whereas only 10% reported being anxious in the nonanxiety condition.

Further analysis of the group assignment/MAACL relationship showed that the correlation figure may underestimate to some extent the true treatment effects. As discussed above, each cell was balanced for the proficiency level of the students. While the Student Response Style scores of the high-proficiency subjects in the anxiety condition reflect the same treatment effects as found for the anxiety group as a whole—$t (9) = -1.94$; 

1 Interrater reliability was assessed using the Pearson product-moment correlation coefficients. All three correlations ($r = .62, .69, .69$) were significant ($p < .01$), indicating a moderate degree of interrater agreement. The composite scores used in the study increase measurement stability by reducing the effects of the biases of individual raters.
p < .05 (one-tailed test)—the high-proficiency subjects had a mean MAACL Anxiety score of −6.6, while the mean for the low-proficiency group was 9. (A negative score indicates a lack of anxiety.)

Thus, the scores of the high-proficiency students, who did not perceive themselves as anxious as a result of the stress condition, attenuated the correlation between treatment assignment and MAACL scores. As the MAACL is a self-report measure, it is likely that the anxiety condition also affected the high-proficiency subjects without their being consciously aware of it.

DISCUSSION

While other studies have examined the influence of anxiety on overall proficiency in a second language, this study examined the effect of environmentally induced anxiety on a more subtle aspect of second language performance: the degree of subjectivity, of personal input, in the second language message. It was found that subjects undergoing an experimental treatment aimed at making them feel anxious and “on the spot” described visual stimuli less interpretively than did subjects in a relaxed, comfortable environment.

To what extent can the results of this study be generalized to the second language classroom? While the anxiety condition was somewhat artificial, the situation probably seemed quite credible to the many students who feel the constant pressure of evaluation in the second language classroom. Further research in second language classrooms is necessary to determine the relationship between the content of second language speech and anxiety in natural settings.

This study has important implications for teachers who believe that language teaching and learning should be based on genuine communication in the target language. Realistic communication is both subjective and objective, requiring the speaker to discuss personal reactions to and interpretations of facts, as well as the facts themselves. The results of this study suggest that students may be less likely to attempt these kinds of messages in a stressful, nonsupportive environment.

REFERENCES


BRIEF REPORTS AND SUMMARIES
The Influence of Background Knowledge on Memory for Reading Passages by Native and Nonnative Readers

HELEN ARON
Union County College

Schema theory research has provided evidence of the importance of background knowledge in reading comprehension. Specifically, content schemata are previously established patterns of background knowledge existing in the mind of a reader and are used to create meaning from text. During the reading process, selected “new” information from the text is related to “old” information acquired from the reader’s previous world knowledge (Kintsch & van Dijk, 1978).

“Through membership in a culture, an individual has privileged information which is represented in a rich system of schemata” (Steffensen & Colker, 1982, p. 2). However, when the cultural backgrounds of the author and reader of a text differ, the reader may inappropriately instantiate schemata (Adams & Bruce, 1982). The schemata needed for reading comprehension in a second language (L2) are often nonexistent or contain information inaccurate for the L2 setting. That is, there is a mismatch between the background knowledge presupposed by the text and the background knowledge possessed by the reader (Carrell & Eisterhold, 1983).

Since the mid-1970s, a number of empirical studies on cross-cultural comprehension have been based on schema theory (e.g., Connor, 1984; Johnson, 1982; Lipson, 1983; Steffensen & Colker, 1982; Steffensen, Joag-Dev, & Anderson, 1979). In general, these studies have found that subjects read passages with native themes more rapidly than passages with nonnative themes. Subjects recall a greater amount of information from native reading and listening passages, produce more culturally appropriate elaborations of the native passages, and generate more culturally biased distortions of the foreign passages. When portions of a foreign passage are familiar, there is significantly greater recall of the familiar portions than of the unfamiliar parts.

The study reported here was designed to investigate whether the potential mismatch in background knowledge between text and reader might affect the placement of ESL students into remedial reading classes.
In New Jersey, all entering freshmen at 2-year and 4-year state colleges must take a battery of basic skills tests in reading, writing, and mathematics. Any student scoring below a predetermined minimum grade is required to enroll in a remedial course. This study attempted to assess whether second language speakers might be receiving low scores on the reading test (and assignment to a remedial reading class) because they did not possess the background schemata expected of examinees.

Previous cross-cultural studies of the effect of background knowledge on memory or on comprehension have used narrative material composed specifically for use in those investigations. In this study, the stimulus material was expository prose and was taken from a standardized reading test in the target language.

METHOD AND PROCEDURES

Subjects

The sample consisted of 62 subjects, all of whom were first-semester freshmen at a New Jersey community college. Thirty-one of the subjects were native speakers of English, were born in the United States, and were considered to be on grade level, that is, enrolled in English 101 (Freshman Composition). The other 31 subjects were not native speakers of English, were not born in the United States, and were enrolled in English 111, a 3-credit freshman composition course taken in lieu of English 101 by ESL students. Students enrolled in English 111 have either completed or been exempted from the ESL program; they are presumed to be comparable in English language proficiency to the English 101 students. The 17 female and 14 male nonnatives represented 17 different countries of national origin and had lived in the United States for a period of time ranging from less than a year to 14 years. The average length of residence was 4.6 years.

Materials and Administration

All subjects were asked to read 2 passages selected from the total of 12 passages on the Reading Comprehension subtest of the New Jersey College Basic Skills Placement Test, 1982-3 Academic Year, Form 3EJP7 (1982). Both passages are at college reading level, according to the Flesch readability formula. The first passage, chosen because its theme is universal, rather than specific to the United States, is as follows:

Some anthropologists suggest that when human beings became toolmakers, they also began to develop language. Lacking speed or strength, claws or fangs, they survived only because their hands and brains gave them unique abilities. Language enabled them to warn others of approaching danger, to direct others to a food source and to instruct their companions in the use of tools. With language, they could signal to others even in the dark, at distances, and when their hands were busy. These abilities considerably improved human beings’ chances of survival.

1 From the New Jersey College Basic Skills Placement Test, 1982-3 Academic Year, Form 3EJP7 (pp. 3 and 8, respectively), 1982, Trenton, NJ: New Jersey Basic Skills Council. Copyright 1982 by the State of New Jersey. Reprinted by permission.
The second passage, which was chosen because it is bound to U.S. culture, is as follows:

If, in the early 1860s, the remaining free tribes believed that the White people's Civil War would significantly delay the pioneers' invasion of Indian territory, they were soon disillusioned. During the next thirty years, Cochise, Geronimo, Sitting Bull, and many other tribal leaders would have to fight against western expansion in every way they could. Their feats would be recorded by historians biased against the Indian cause. Even so, their names would become as well known as those of the people who opposed them. Most of them, young and old, would be driven into the ground long before the symbolic end of Indian freedom came at Wounded Knee in December 1890. Still, a century later, in an age without heroes, with their roles in history reexamined, they would come to be considered among the most heroic of all Americans.

This passage contains four proper names (Cochise, Geronimo, Sitting Bull, Wounded Knee) that an American student would have encountered in history classes but that would probably be unfamiliar to nonnatives. Although nonnatives could understand words and phrases like civil war, white people, pioneers, Indian tribes, Indian territory, Indian freedom, and westward expansion as vocabulary items at the level of literal comprehension, they would be unable to comprehend the passage at the interpretive level without knowledge of specifics about American culture, that is, the historic problem of Indians versus whites.

All subjects were tested individually in an untimed condition. The two passages were read one at a time, and the order of presentation of passages was alternated. After indicating to the examiner that they were ready, the subjects gave an oral recall of everything in the passage that they could remember. This protocol was tape-recorded. When a subject finished retelling the content of a passage, the examiner prompted once by asking, “Is there anything else you can remember?” The same procedure was followed for the second reading passage.

Scoring

The tape recording of each subject’s recall of the reading passages was transcribed. Grammar errors were transcribed verbatim; however, references to the pragmatic conditions of the task (e.g., “That’s all I can remember about what I read”) were omitted.

Next, all the protocols were scored holistically by two teams of raters. Holistic scoring was selected as the method for assessment because the group would be judged against itself rather than against a norming population. General impression marking, a procedure developed by the Educational Testing Service (Cooper, 1977; Odell, 1981), was used. Consistent with the normal procedure for scoring the writing portion of the New Jersey College Basic Skills Placement Test, each protocol was rated by each of two judges on a 6-point scale.

The raters for the subjects’ protocols were four full-time faculty members from the English department at a New Jersey community
college. Each was familiar with general impression marking and had been previously trained in this procedure by a staff member at the Educational Testing Service. For the purposes of interrater reliability, the Educational Testing Service considers that raters are in agreement if the difference between the scores they assign is 2 points or less. Using this standard, the raters for the protocols in this study had 100% agreement.

FINDINGS

Two t tests were performed to calculate the significance of the difference between the means of the protocol scores (see Tables 1 and 2). The first, which compared the scores of native and nonnative speakers for recall of the passage with a universal theme, indicated no significant difference ($t = .99$, n.s.). However, the second t test, which compared recall by native and nonnative speakers of the passage with a U.S. culture-bound theme, indicated a significant difference in group performance ($t= 2.25$, $p < .05$).

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison of Native and Nonnative Speakers’ Recall Scores for Passage With a Universal Theme</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Native</td>
</tr>
<tr>
<td>Nonnative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison of Native and Nonnative Speakers’ Recall Scores for Passage With a U.S. Culture-Bound Theme</td>
</tr>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Native</td>
</tr>
<tr>
<td>Nonnative</td>
</tr>
</tbody>
</table>

* $p < .05$.

Thus, while native and nonnative subjects appeared to bring similar previous knowledge to the passage with a universal theme, they seemed to bring differing degrees of pertinent, previously acquired knowledge to the passage with a culture-bound theme.
These findings, which can be added to a growing body of cross-cultural schema theory research, are of course limited by the content of the two reading passages. While the passages appear to be typical of selections found on standardized reading tests for college freshmen, different results might have been obtained with other passages. Nonetheless, English placement and proficiency tests containing passages that require U.S. culture-bound background knowledge may well discriminate against ESL students.

Mandatory testing in reading, which is currently required at the elementary, secondary, and/or college levels in 29 states and is under consideration in 2 more (Gambrell, 1985), may be unfair to second language speakers. Their subsequent placement may be partially based on how closely their background knowledge matches that presupposed by the test rather than based on an assessment of their second language skills.

REFERENCES
New Jersey college basic skills placement test, 1982-3 academic year (Form 3EJP7). (1982). Trenton, NJ: Board of Higher Education.
Author’s Address; Department of English, Union County College, Cranford, NJ 07016
The phrase “process, not product” has now risen to the heights, or perhaps sunk to the depths, of the great buzzwords of TESOL’S past: It is the “communicative competence” of the mid-1980s. Though initially offering fresh insight into an important area of teaching, it has now been miscast as a complete theory of writing; its adherents spread its gospel with religious zeal, and those who fail to pay homage to it in any paper or discussion on writing are ostracized; indeed, it has reached the point where, to quote Tallulah Bankhead, “there is less here than meets the eye . . .”

The uncritical acceptance of this approach is attested to by the fact that discussions of its shortcomings are almost nowhere to be found. Nevertheless, before anyone fully embraces this approach, the following points should be considered:

1. The process-oriented approach fails to prepare students for at least one essential type of academic writing.

At the “Key Questions About Writing” roundtable discussion which took place at TESOL ‘85, speaker after speaker stressed the need for students to produce multiple drafts of papers in order to allow the process of evaluation and revision to go forward. At one point, someone in the audience asked the panel to relate this dictum to essay examination writing. One of the members of the panel, having no ready answer, simply dismissed examination writing by claiming that it was not “real” writing. Real men don’t eat quiche, and real foreign students . . .

Against this cavalier view of what is real and what is not stands our students’ indisputable need to write adequate in-class examination essays several times per semester. Whether an approach which constantly emphasizes multiple revisions will
eventually lead to a fluency that facilitates fast essay writing is an unanswered question. It is claimed, of course, that emphasis on process leads to a better product, that process and product do not stand in opposition to each other. But is there only one process and only one product? In Raimes’s (1985) study of the writing process, students were asked to "Tell [Maria Chen] about something unexpected that happened to you" (p. 236). Does such a study tell us about the process involved in producing a lab report or an annotated bibliography?

I would claim that there are as many different writing processes as there are academic writing tasks and that anyone who claims to understand the former had better have a specific taxonomy of tasks in mind. In other words, even if it can be said of using multiple drafts that "it works," the question is, For what? For writing essay examination answers under pressure of time? No one knows.

2. The inductive orientation of the process-centered approach is suited only to some writers and some academic tasks.

Some writers do better by producing an outline first—the "radical outliners," as one dissenting member of the roundtable panel called them (see Reid, 1985). In addition, many of the academic writing assignments I examined at Western Illinois University (Horowitz, 1985) left no choice but to write in a "top-down" way because they required students to follow a tightly structured, question-by-question or point-by-point outline.

Teaching students to write and revise according to the demands of an audience is useless unless those demands are realistic simulations of academic demands. Going far beyond the usual demands of coherence, relevance, and so on, most academic writing tasks, at least at the university where I work, require students to present data, usually obtained through written sources, according to a fairly explicit set of instructions. Does an inductive approach prepare students for these kinds of tasks? No one knows.

3. In claiming that certain internal states of mind are superior to others for successful writing, the process-oriented approach ignores the fact that radically different orientations to a situation can be equally successful.

A basic dogma of process-oriented teaching is that good writing is "involved" writing, that students write best when they care about their subject. It is assumed that students who choose their own topics and answer the questions they are truly curious about will be more highly motivated, better writers.

Although it would be difficult to disprove these assumptions, two
facts diminish their importance. First, students rarely have a free choice of topics in their university writing assignments. Teaching students to write intelligently on topics they do not care about seems to be a more useful goal than having them pick topics which interest them.

Second, while it may be true that the “good” writers who have been studied exhibited certain characteristics, it seems both futile and culturally insensitive to try to make over “bad” writers in their image. Many of our students, for better or for worse, have been highly conditioned by the demands of their native education systems to see THE TEST or THE PAPER or, most of all, THE GRADE as the be-all and end-all of the educational process. This may offend some teachers’ humanistic sensibilities and may, according to certain Western psychological theories, prevent these students from reaching their full human potential, but who are we to try to change the value structures of our students? And again, are typical American students “thirsty for knowledge”? Are we? Unless a strategy is truly nonadaptive to a situation (that is, will result in failure), we, as teachers, would be better advised to tap into the motivation behind it than to try to restructure our students’ thought patterns.

4. The process-oriented approach gives students a false impression of how university writing will be evaluated.

One of the panelists in New York said that process-oriented teachers take a humanistic approach, responding to the student rather than to the student’s writing. Another panelist rejoined that while teachers using this approach may respond to the student, examination readers will surely respond only to the writing itself. The “gentle” approach of process-oriented classrooms may foster a false impression of the realities of academia, where our students’ product-oriented attitudes may in fact be more adaptive.

In sum, the process-oriented approach is a collection of teaching techniques which have certain merits in certain situations. Multiple drafts? Of course. Too many of our students believe that once it is down on the page, their job is finished (they are only partly right). Group work? Certainly. Our students surely can teach each other as much as or more than we can teach them. Get it down on the page and then organize it? This will help some of our students prepare for some academic tasks. Choose topics of personal interest? This has always been an effective technique at the lower levels. Gentle peer evaluation? Since we are teaching a developmental skill, we certainly must walk the line between discouraging our students with low grades and giving them a false impression of their abilities.
Yet, despite the undeniable merits of these techniques as techniques, teachers should be extremely cautious about embracing an overall approach which, in its attempt to develop their students’ writing skills, creates a classroom situation that bears little resemblance to the situations in which those skills will eventually be exercised. “Resisting the bandwagon” (Leki, 1985, p. 6) is really quite easy when one realistically assesses what is there and what is not there in the process-oriented approach.

REFERENCES

Toward a Methodology of ESL Program Evaluation

ALAN BERETTA
University of Edinburgh

In contrast to the prominence it has had in education journals, the question of evaluation methodology and purpose has received relatively little attention in language teaching journals. However, two recent contributors to the TESOL Quarterly have heightened awareness of the importance of program evaluation (especially method evaluation) and set out some of the major problems involved. Richards (1984) points out that many currently popular methods have yet to be evaluated in any rigorous manner, and Long (1984) emphasizes the need to incorporate systematic observation as a means of explaining results.

This discussion builds on these initiatives (though occasionally taking issue with them in the process). The arguments here are
offered in the hope that they will give rise to further discussion and refinement. The starting point for this discussion is that evaluation is first and foremost applied inquiry. The choice of methodology, to some extent, flows naturally from this fundamental perspective. I would suggest that (a) we conduct our investigations in the field rather than in artificially controlled “laboratory” settings, (b) we consider the effect of total programs rather than isolated components of them, (c) the duration of the studies be long-term rather than short-term, and (d) randomization is not always practicable or crucial. I will elaborate on the kinds of questions that attempts at external validation might address and propose a viable agenda for cumulative investigation, affirming throughout that relevance is the principal consideration in all ESL program evaluation undertakings.

EVALUATION AS APPLIED RESEARCH

Second language acquisition researchers can perhaps afford to invest decades in trying to arrive at a coherent theory of language learning, but in the meantime teachers must be given some form of sustenance (Lightbown, 1985). They might obtain usable information from evaluators if evaluators were oriented toward providing such a service. Unfortunately, evaluators have often behaved more like “basic” researchers than applied. They have attempted to capture small pieces of ground of which they can be certain and have appealed to an indeterminate number of follow-up experiments to replicate their findings or capture other small pieces of ground until eventually a theory is formed. These other experiments are usually never mounted, but even if they were, they would require an inordinate length of time to carry out, and teachers would as a result be left high and dry. An example is in order.

Freedman (1976) conducted a study into methods in which all students were randomly assigned to treatments (to control for initial learner differences) and teachers were dispensed with and replaced by pretaped lessons (to control for differences in teachers). The duration of the experiment was restricted to one lesson (because the longer a treatment, the greater the likelihood that uncontrolled variables will intervene), and the scope of the inquiry was confined to the effects of implicit and explicit modes of teaching the French subjunctive (because complete methods are continually shifting mosaics, and by isolating one element, control is possible). By putting the study in quarantine, Freedman argued that she had achieved an internally valid experiment. She recommended further specific comparisons to contribute to an overall theory (Freedman,
1976, p. 25). In the event, other components were never tested and reported (at least not in any form accessible to this author).

Having been cleansed of all typical classroom influences, such as the presence of teachers, Freedman’s study can have only extremely remote implications for practice. As an example of laboratory method comparison, although exceptionally austere, it is not an isolated case (see also Levin, 1972; Seliger, 1975; Von Elek & Oskarsson, 1973; Wagner & Tilney, 1983). These are all instances of well-conducted research, but they are not evaluation. Evaluation, as has often been averred (Suchman, 1972; Talmage, 1982), is applied research and must confront the real world. Our primary goal is to provide feedback to teachers in the short run. If that is accepted in ESL, as it seems to be in education (Cronbach, 1980), it will give us a criterion against which to judge the acceptability of our methods of inquiry.

At this stage, it should be stressed that I am not advocating that we relinquish all control and concern ourselves only with relevance. I am stating a priority. What is relevant must take precedence over what can be tightly controlled, but it is at the same time quite clear that we do not assist anyone if our findings are completely undependable. Nevertheless, we might start from considerations of what will be usable and only then proceed to judge what controls are possible and desirable.

EXTERNAL VALIDITY

Another way of saying that a study is usable is to say that it has a high degree of external validity. External validity is the level of representativeness of an investigation (Campbell & Stanley, 1963). That is to say, we are concerned with the extent to which results obtained in one setting, population, and time may be generalized to another (Glass, 1982). Even the slightest contemplation of this matter will lead us to realize that external validity cannot be established with any certainty, induction being largely immune to logic (Popper, 1972). However, because it may inform teacher practice, it is of the essence in program evaluation.

When all is said and done, probably every one of us would wish to be able to generalize our findings, so it is perhaps a little surprising that Long (1984) should list and describe six threats to internal validity but fail even to allude to external validity. To complement Long’s necessary and worthwhile elaboration of the potential impediments to reliability, I would like now to look at some of the ways in which we might increase generalizability.
Setting

The more the settings of our evaluations resemble regular classrooms, the greater the degree of ecological representativeness and the more confident we can be in extrapolating to other settings (Bracht & Glass, 1968). By contrast, a study which takes place in contrived conditions has no credible relationship with what might happen in real classrooms. What happens in a stripped-down environment may not parallel what occurs in the field. This can be illustrated by the example (cited by Good & Power, 1976, p. 47) of Kounin’s (1970) finding that teachers’ techniques of restraining individual students had a ripple effect on classmates in a laboratory setting but made no impression at all in naturalistic conditions. (A number of further examples of the disparity between laboratory and field have been brought together by Beretta, 1985.) Laboratory studies, then, remain in limbo as far as teacher practice is concerned, since the results cannot be applied beyond the confines of the experiment.

Treatment

The question here is whether we should concern ourselves with whole methods or should segment methods and test elements separately. The difficulty of segmenting components is partly that they may not be readily identifiable, but more important, that when they are treated in isolation, they may not behave in the same way as they would in the company of other components. In other words, given the likelihood of interaction effects (Cronbach & Snow, 1977; Good & Power, 1976), we cannot assume that variables will exhibit the same effects individually as in combination.

Even when separate elements are not abstracted from particular methods, the same problem arises. For example, Von Elek and Oskarsson’s (1973) comparison of implicit and explicit teaching of grammatical rules may not have been consciously derived from audiolingual and cognitive code methods, but they acknowledged that “it is obvious that each of our experimental strategies has a great deal in common with current methods” (pp. 14-15). The point is that when these laboratory-tested elements are restored to an integrated program, whatever its label, the effects will be unpredictable, and external validity will therefore be considerably constrained.

Dealing with complete methods, on the other hand, should establish whether or not elements work in combination, and this provides more immediate grounds for generalization. Perhaps
certain elements are more important than others, and perhaps our explanatory tools (such as systematic observation) will indicate why a program works; in any case, however, the main objective of securing usable information is achieved.

Scriven (1977) makes this point with an analogy from the design of automobile engines. In a number of instances, an engine had been designed that was clearly superior to its competitors. Rather than waiting to find out which of perhaps 30 variables was principally responsible for increased performance, however, the manufacturer decided to go into production on the basis of only the comparative evaluation. Scriven observes “this is the way we have to work in any field where there are too many variables and too little time” (p. 357).

**Population**

How far is the population (learners and teachers) we have sampled from representative of the population to which we wish to generalize? With respect to ESL methods, we do not really have prespecified target populations. We appeal instead to all interested persons. The extent to which interested persons are likely to be able to interpret our findings may nevertheless depend partly on how broad a sample we use in our studies. Wagner and Tilney (1983) used 21 subjects: 9 advanced English language students, 3 English language instructors, and 9 graduate music students. In such a case, generalization would be restricted to a very limited population. By contrast, the 328 pupils in the four regular schools involved in the evaluation of the Bangalore Project (Beretta & Davies, 1985) may have many traits in common with the students of other South Indian ESL teachers. Obviously, however, we take whatever subjects we can get, and their representativeness is probably more often than not a matter of availability.

It might be argued that evaluations which are dependent on volunteer teachers to implement a particular method cannot be generalized to nonvolunteer teachers. This is not especially problematic, since the adoption of methods in our profession is largely governed by choice. Unlike bilingual education, for instance, ESL method is apolitical, so adoption is not typically imposed from without. Therefore, although volunteer teachers may have different characteristics from nonvolunteer teachers, a problem of generalizability hardly arises because adoption too is usually voluntary.
Duration

A great number of ESL method studies have been what Eisner (1984) would call “educational commando raids” (p. 451). Get in. Get the results. Get out. For example, Seliger (1975) took 65 minutes, Lim (1968) 1 hour 45 minutes, McKinnon (1965) 2 hours 15 minutes, and Freedman (1976) a single lesson.

Snow (1974) advised that “most generalizations about school learning need to be built on research using substantial samples of learning time” (p. 281). This seems sensible because learning takes place over an extended period of time (see P.D. Smith, 1970, p. 6), and it would be unfair to find for or against a method after examining the effect it produces only within an exceedingly brief time span. Information about learning based on one or two lessons may have its purposes, but they are not evaluation purposes. Such information says nothing about representative conditions (which include duration).

A natural limit on duration is set by the length of a course of study. Other equally practical limits, like the imminence of external examinations, may also inhibit duration. Nevertheless, methods should surely be given as prolonged a hearing as local pressures allow.

EXTERNAL VALIDITY AS CREDIBLE REASONING

In ESL, the adoption of innovative methods is not usually on a vast scale (though see D.A. Smith’s [1962] account of the retraining of 27,000 teachers in a structural method in South India). If our goal were large-scale implementation, we would have to sample, as far as possible, all the identifiable target populations, and we would be interested in grand averages. In ESL methodology, however, adoption typically remains a matter of individual or perhaps institutional choice. On reading evaluation reports, ESL teachers are not especially interested in national means but in whether a certain method is likely to work for them, in their specific circumstances, with their particular learners. This requires quite different emphases and a quite different role for external validity.

Clearly, if we are able to glean usable information from studies carried out in only a few schools, there can be no logical or even statistical grounds for generalizing to other circumstances. We can, however, collect appropriate data and prepare evaluation reports in such a way that there are good psychological bases for extrapolation (Mahoney, 1978).
In this sense, external validity embraces construct validity; it involves a construction of reality on the basis of what is known about the relevant situations (Cronbach, 1982). Teachers know about their own situations, and the evaluator’s task is to enable them to put an adequate construction on the situations under study.

The evaluator can fulfill this function by thoroughly explaining results. One way of doing this is to monitor behavior through systematic observation (Long, 1984), but a number of other data-collecting procedures can also aid explanation and promote credible reasoning. For example, historical narratives can illuminate in an unconstrained manner that is denied to fixed observation schedules, and retrospective accounts have the decided advantage of being able to focus on classrooms which were particularly successful or unsuccessful (Cronbach, 1982). Teachers’ stages of concern can be determined through questionnaires (Hall, George, & Rutherford, 1977), and the implementation of innovations can be explored through focused interviews (Hall & Loucks, 1977) or through checklists mapping configurations (Heck, Stiegelbauer, Hall, & Loucks, 1981). Information gathered from such multiple perspectives might complement systematic observation and provide a more comprehensive picture for practitioners. Thus, evaluation reports could be brought to life and inference from them rendered plausible.

CONTROL

A long-term study in a natural setting involving complete methods has all the benefits of providing usable information but lacks some of the control that is possible in a short-term, artificially conditioned inquiry into method components. There is thus a trade-off between relevance and proof. To an extent, however, the measurement of implementation replaces actual control. In other words, being able to explain results plausibly (through observation, interviews, narratives, and so on) substitutes for the relative certainty associated with the isolated variables of a laboratory investigation.

Because of the logistical problems involved, it is not often possible to achieve randomization in the field. (Students cannot be randomly assigned without causing considerable upheaval in school schedules, and normally we have too few schools at our disposal for the class to be the unit of analysis.) Sometimes with great cooperation, though, it proves possible (e.g., Green, 1975). A method is especially ripe for randomized field experiment when it has been thoroughly investigated and cumulative findings lead us to
expect a positive result. (Otherwise we may be in danger of stretching the researcher-practitioner relationship too far.) This conforms with the observe-correlate-experiment procedure by now familiar (and productive) in education research (Brophy & Good, 1984).

An insistence on true experiment (Long, 1984; Richards, 1984) is problematic. In Cronbach’s (1982) words, “The outmoded recommendation that the program evaluator prefer true experiments is hopelessly ambiguous” (p. 324). For Long (1984), true experiment involves the setting up of randomly assigned experimental and control groups so as to permit causal inference about the efficacy of a method (p. 410). This may be possible in a laboratory study, but Long clearly has in mind the examination of complete methods in a natural setting, as his example of a process evaluation indicates (pp. 415-416).

As I have argued above, this kind of evaluation would allow plausible explanation and inference but not the control necessary to sanction causal claims. (After all, systematic observation may miss much that is responsible for change.) Laboratory precepts cannot be shifted intact to field practice. If true experiment is to legitimize causal statements, then true experiment is beyond the evaluator’s reach. It would, perhaps, be reasonable and helpful to delete the word causality from our evaluation vocabulary.

A VIABLE AGENDA FOR EVALUATION STUDIES

Many academic investigations end with recommendations for further research, an implicit acknowledgment that stand-alone studies are a great deal weaker than cumulative findings. This is, of course, commendable. However, such recommendations can sometimes appear to be somewhat disingenuous if the suggested studies are obviously impracticable or too inconvenient ever to be carried out. There can be no doubt that evaluation would benefit from cumulative studies investigating the same or similar themes in a variety of settings. They would help to shorten still further the distance that generalizations must travel. But what kind of inquiries could really be expected?

We may immediately rule out multi-site ventures and an insistence on randomization: Present levels of funding preclude the former, and the latter depends heavily on cooperation and availability. Also, we can probably not expect duration to stretch over years because career structures, doctoral dissertation schedules, and pressures on schools seem likely to encourage shorter spells.
If we were to seek data from a number of sources on the implementation and effects of, for example, the Natural Approach (Krashen & Terrell, 1983), we might anticipate that the modal study would take place in one or two schools, involve intact classes, and last perhaps only a semester. While we might expect that some would be longer and that some would be randomized, the anticipated modal investigations would greatly increase our confidence in extrapolation. What is clear is that unless we construct a viable, cumulative agenda for ESL program evaluation and constantly refine our methods of inquiry, our potential for providing usable findings will be unduly limited.

CONCLUSION

Once the decision is made to undertake a field study of the effects of complete methods over time, we bid farewell to proof. It barely needs to be mentioned, since it has been said so often (e.g., Stern, 1983), that methods are not static, standardized treatments but instead, constantly varying, often overlapping, interacting sets of behaviors. Some may feel, then, that it seems too “unscientific” to attempt to measure such apparent chaos, keeping tabs on turbulence. However, I would argue that if we use all the means at our disposal of documenting what happens when innovations are implemented and if we use such controls as are feasible and desirable, we at least arrange our priorities to provide for plausible extrapolation. As primarily applied researchers, evaluators can have no business retreating to well-ordered universes of their own construction. Had we but world enough and time, we could join our second language acquisition research colleagues and devote our energies to theoretical advancement; but evaluation is nothing if it is not timely and relevant.

Although I have emphasized the differences between my perception of ESL program evaluation and Long’s, there is much common ground. Long (1984) makes a valuable contribution by insisting that we should measure the independent variable (i.e., the teaching process) if we are to begin to explain our results. This is his main point, one with which I am in total agreement. Where we part company is in the relative importance we would attach to internal and external validity.

Discussions about methodology so easily become polarized. In educational evaluation, we have witnessed the quantitative versus qualitative skirmishes, which seem finally to have reached an accommodation (Cook & Reichardt, 1979). In ESL, we may draw on this experience, modifying perspectives to cater to our own

152
specific demands. Most of us would probably agree that what is required is a judicious balance between internal and external validity, between reliability and usability, and between certainty and relevance. By airing our differences, we may arrive at the balance that best answers our needs.

REFERENCES


Now in two volumes

SECOND LANGUAGE ACQUISITION IN CHILDHOOD

BARRY McLAUGHLIN
University of California, Santa Cruz

"... a masterful overview of the variegated research on the process of second language acquisition. It is up-to-date, treats the topic thoroughly [484 references], and evaluates the voluminous [and heterogeneous] research at a well-tuned level of exposition..."

—American Scientist

"...thorough, balanced, and lucid...there is currently no other treatment of these issues that can claim to provide the breadth and penetration of McLaughlin's book. It would make an excellent text for graduate, and possibly undergraduate, courses."

—Contemporary Psychology

These works challenge traditional views of second language learning in preschool and school-age children and feature discussions of: the critical period, a comparison of second language acquisition in children and adults, simultaneous bilingualism, international perspectives on bilingual education, instructional models and individual difference variables in the classroom. Together, they offer comprehensive, up-to-date coverage of research in an increasingly important area.

Volume 1: Preschool Children
Contents: Bilingualism in Society, First-Language Acquisition, Language Learning in Childhood and Adulthood, Acquisition of Two Languages in Childhood, Successive Acquisition of Two Languages in Childhood.
0-89859-378-6 1984/280 pp./$24.95

Volume 2: School-Age Children
0-89859-565-7 1985/304 pp./$29.95

Now available as a set
0-89859-822-2 $42.50

Mention this ad and save 15% when accompanied by payment. Write:

LAWRENCE ERLBAUM ASSOCIATES, INC.
365 Broadway, Hillsdale, NJ 07642
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 3. testing and evaluation
   learning and teaching; issues in research and research methodology
2. curriculum design and development; 4. professional
   instructional methods, materials, and techniques 5. language planning
   professional standards

Because the Quarterly is committed to publishing manuscripts which contribute to bridging theory and practice in our profession, it particularly welcomes submissions which draw on relevant research in such areas as 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 which then address implications and applications of that 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
1. The TESOL Quarterly invites submissions in four categories:

   Full-length articles. Manuscripts should usually be no longer than 20 double-spaced pages. Submit three copies plus three copies of an informative abstract of not more than two hundred words to the Editor of the TESOL Quarterly:

   Stephen J. Gaies
   Department of English
   University of Northern Iowa
   Cedar Falls, IA 50614 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 non-print 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 5 double-spaced pages, although comparative reviews or
review articles may be somewhat longer. Submit two copies of reviews to
the Review Editor:

Vivian Zamel
English Department
University of Massachusetts/Boston
Harbor Campus
Boston, MA 02125

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 which 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 investiga-
tion. Although all reports and summaries submitted to this section will be
considered, preference will be given to manuscripts of 5 double-spaced
pages or fewer (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, Brief Reports and Summaries:

D. Scott Enright
Department of Early Childhood Education
University Plaza
Georgia State University
Atlanta, GA 30303

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. Contribu-
tions to The Forum should generally be no longer than 5 double-spaced
pages. Submit two copies to the Editor of the TESOL Quarterly at the
above address.

2. 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, N.W., Washington,
D.C. 20036. The Publication Manual is also available in many libraries and
bookstores.

3. 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.

4. 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 which they would like to have
included. Double spacing should be used throughout.
5. The *TESOL Quarterly* provides 25 free reprints of published full-length articles, and 10 reprints of published reviews.
6. Manuscripts submitted to the *TESOL Quarterly* cannot be returned to authors. Authors should be sure to keep a copy for themselves.
7. It is understood that manuscripts submitted to the *TESOL Quarterly* have not been previously published and are not under consideration for publication elsewhere.
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.
A scholarly tour-de-force...

A Comprehensive Grammar of the English Language

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik

Longman is honored to present a major contribution to the study of English grammar. The authors, widely acclaimed for their previous works, share the fruits of twenty years of research and collaboration. Their findings are gathered together in the most complete and authoritative reference work of its kind. A Comprehensive Grammar of the English Language spans such areas as English grammar, semantics, pragmatics and lexicology. It includes material on spelling, pronunciation, word formation, register, regional variation, non-standard speech and much more.

No teacher, no school, no library can afford to be without it.

Also of interest...

Stephen Krashen's
The Input Hypothesis: Issues and Implications
is an important addition to second-language learning theory. A must for everyone in the field.

95 Church Street,
White Plains, NY 10601

[Longman Logo]
PUBLICATIONS RECEIVED

Publishers are invited to send copies of their new materials to the TESOL Central Office, Georgetown University, Suite 205, 1118 22nd Street, N.W., Washington, D.C. 20037. Packages should be labeled REVIEW COPIES.

TESOL Quarterly readers are invited to contribute evaluative reviews, comparative reviews, and review articles for consideration for publication in the Review section of the Quarterly. These should be sent to the TESOL Quarterly Review Editor, Vivian Zamel, University of Massachusetts/Boston.

TESOL gratefully acknowledges receipt of the following publications.


PUBLICATIONS AVAILABLE
FROM THE TESOL CENTRAL OFFICE

Publications on a wide range of topics of relevance to TESOL professionals are available from the TESOL Central Office. Many of these publications are offered at reduced rates to TESOL members, and single copies of several of these are available at no charge. An annotated list of publications can be obtained by contacting TESOL, Georgetown University, Suite 205, 1118 22nd Street, N.W., Washington, D.C. 20037. 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.

REFERENCE GUIDES


REVISED

English Language and Orientation Programs in the United States. A directory of programs and courses in English as a second or foreign language at U.S. institutions of higher education and nongovernmental agencies, for students from abroad. Institute of International Education, 1984. $8.95

1985 TESOL Membership Directory. TESOL, 1985. 263 pp. $5.00 ($4.00)


Summary Listing of Professional Journals in ESL/EFL. TESOL Newsletter, 1985. 8 pp. $1.00

OTHER TESOL PUBLICATIONS

Teacher Preparation


Classroom Practices

Classroom Practices in Adult ESL. Donna Ilyin and Thomas Tragardh, editors. TESOL, 1978. 209 pp. $4.50 ($4.00)

Classroom Practices in ESL and Bilingual Education. Muriel Saville-Troike, editor. TESOL, 1973. 84 pp. $2.00 ($1.75)

On TESOL Series

On TESOL '84: A Brave New World for TESOL. Penny Larson, Elliot L. Judd, and Dorothy S. Messerschmitt, editors. Selected papers from the 18th Annual Convention of TESOL in Houston, March 1984. TESOL, 1985. 335 pp. $13.00 ($11.00)

On TESOL '82: Pacific Perspectives on Language Learning and Teaching. Mark A. Clarke and Jean Handscombe, editors. Selected papers from the 16th Annual Convention of TESOL in Honolulu, May 1982. TESOL, 1983. 329 pp. $11.50 ($10.00)


On TESOL '78: EFL Policies, Programs, Practices. Charles H. Blatchford and Jacquelyn Schachter, editors. Selected papers from the 12th Annual TESOL Convention in Mexico City. TESOL, 1978. 264 pp. $8.00 ($7.00)

On TESOL '76. John F. Fanselow and Ruth H. Crymes, editors. A sampling of the teaching done at the Tenth Annual Convention in New York City. TESOL, 1976. 276 pp. $6.50 ($5.50)

On TESOL '74. Ruth Crymes and William E. Norris, editors. Selected papers from the Eighth Annual Convention in Denver. TESOL, 1975, 240 pp. $6.50 ($5.50)

Bilingual Education


Testing


The Construct Validation of Tests of Communicative Competence. Adrian S. Palmer, Peter J.M. Groot, and George A. Trosper, editors. A collection of papers including proceedings of a colloquium at TESOL '79 in Boston. TESOL, 1981. 165 pp. $8.50 ($5.50)

Concepts in Language Testing: Some Recent Studies. Eugene J. Brière and Frances Butler Hinoftis, editors. Papers from the colloquium on "Research and testing methods in evaluating second language proficiency" at the 12th Annual TESOL Convention in Mexico City. TESOL, 1979. 100 pp. $5.00 ($4.00)

Writing

Writing and Composition: Supplement No. 1 to the TESOL Newsletter. Ann Raines, editor. Nine articles on ESL writing/composition, plus tips for writing teachers and a brief list of books and journals. TESOL, 1984. 16 pp. $1.00 (discount for quantity orders)
General

$2.00

$2.00

$6.50 ($5.50)

$5.00 ($4.00)

$4.00 ($3.50)

$1.00

Convention Programs. The following convention programs contain the entire program as well as the abstracts of papers presented at that convention.

$4.00 ($3.50)

15th Annual in Detroit, 1981.
449 pp.
$5.00 ($4.00)

16th Annual in Honolulu, 1982.
263 pp.
$4.00 ($3.50)

311 pp.
$7.00 ($6.50)

18th Annual in Houston, 1984.
442 pp.
$7.00 ($6.50)

228 pp.
$6.00 ($5.00)

Guidelines for the Certification and Preparation of Teachers of English to Speakers of Other Languages in the United States. 8-page pamphlet. TESOL, 1976. Single copies, no charge. Multiple copies, 50 cents each.


FROM OTHER PUBLISHERS


TEFL, Perceptions, and the Arab World. John W. Bagnole. Africa-Middle East Educational-Training Services, 1976. 50 pp. $2.00

Adapting and Writing Language Lessons. Earl W. Stevick. Foreign Service Institute, 1971. 431 pp. $10.00


The Administration of Intensive English Language Programs. Ralph Pat Barrett, editor. NAFSA, 1982. 109 pp. $2.50
BACK ISSUES OF THE TESOL QUARTERLY

Order by catalog listing TQ, followed by number of volume and issue, e.g., TQ 1:4 = TESOL Quarterly, Vol. 1 (1967), No. 4 (December).

1968 Vol. 2, nos. 1, 2, 3, 4
1969 Vol. 3, no. 2—$1.00 each number
1971 Vol. 5, no. 1
1973 Vol. 7, nos. 1, 2—$2.00 each number
1974 Vol. 8, nos. 1, 2, 3, 4
1975 Vol. 9, no. 4
1976 Vol. 10, nos. 1, 2, 3, 4
1977 Vol. 11, nos. 1, 2—$3.00 each number
1978 Vol. 12, nos. 1, 2, 3, 4
1979 Vol. 13, nos. 1, 2, 3, 4
1980 Vol. 14, nos. 1, 2, 3
1981 Vol. 15, nos. 1, 2, 3, 4—$4.00 each number
1982 Vol. 16, nos. 1, 3, 4
1983 Vol. 17, nos. 1, 2, 3, 4
1984 Vol. 18, nos. 1, 2, 3, 4—$5.00 each number
1985 Vol. 19, nos. 1, 2, 3, 4—$6.00 each number

Special price for all volumes
listed above (2–19) ——$125.00

Issues not listed above are no longer in print. They may be obtained from:
University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106

Back Issues of the TESOL Newsletter
A complete set of Vol. 19 (6 issues, 1985)——$8.00
Vol. 18 (6 issues, 1984)——$6.00
Vol. 17 (6 issues, 1983)——$5.00
Vol. 16 (5 issues, 1982)——$5.00
Vol. 15 (6 issues, 1981)———$4.00
Vol. 14 (3 issues, 1980)———$2.00
Vol. 13 (4 issues, 1979)———$2.00
Vol. 12 (5 issues, 1978)———$2.50
## TESOL ORDER FORM

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Title</th>
<th>Unit Price</th>
<th>Total Price</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please add postage and handling as follows:

<table>
<thead>
<tr>
<th>Total Price of books</th>
<th>Postage/Handling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1-$15</td>
<td>$1.50</td>
</tr>
<tr>
<td>$15-$25</td>
<td>$2.00</td>
</tr>
<tr>
<td>$25-$40</td>
<td>$3.00</td>
</tr>
<tr>
<td>$40-$70</td>
<td>$5.00</td>
</tr>
<tr>
<td>$70-$100</td>
<td>$7.50</td>
</tr>
<tr>
<td>over $100</td>
<td>$10.00</td>
</tr>
<tr>
<td>over $150</td>
<td>$15.00</td>
</tr>
</tbody>
</table>

- All orders under $40 must be pre-paid. Orders for which payment is not included will be returned.
- All orders within the USA are shipped library rate; international orders are shipped foreign book rate. Allow 4-5 wks. for USA delivery; 3-4 mo. outside the USA.
- Add $1 to postage rates for United Parcel Service (USA only).

Subtotal

Postage & Handling

Total

Ship to: ____________________________

□ Bill me (orders over $40)
- Please make checks payable to TESOL.
- When sending funds from outside the USA, please remit in the form of a US Postal Money Order, a check drawn on a USA bank or a foreign bank draft on a USA bank. Do not send checks drawn on foreign banks.
- Prices subject to change.

---

TESOL, Georgetown University, Suite 205, 1118 22nd Street, N.W., Washington, D.C. 20037, U.S.A.
Hopscotch

New

Regents’ new six-level ESL/EFL program for children aged 5-12 is:

- **ACTIVITY-BASED**
  HOPSCOTCH allows children to learn by doing—the way children learn best. And its abundance of "hands-on" activities, songs, chants, and games are all geared to their natural interests and abilities.

- **FLEXIBLE**
  HOPSCOTCH can be used with large or small classes. And it offers activities for pairs, small groups, or large groups so you can tailor-make lessons just for your class.

- **EASY TO USE**
  HOPSCOTCH’s clear organization insures that both students and teachers always know what to do. The step-by-step instructions in the Teacher's Editions reduce preparation time, too.

- **BROAD IN SCOPE**
  HOPSCOTCH presents students with interesting, familiar material in such content areas as math, science, and social studies to make teaching and learning English more rewarding than ever before.

**HOPSCOTCH** offers a full range of components: textbooks in full color, activity books, teacher's editions, and cassette programs with music and sound effects to make language practice come alive in the classroom.

To learn more about these and other quality ESL materials, call toll free or contact your nearest Regents representative.

1-800-822-8202
In AK, HI, and NY call 212-889-2780

REGENTS PUBLISHING COMPANY, INC.
TWO PARK AVE.
NEW YORK, NY 10016

The ESL/EFL Specialists
WHILE EVERYONE ELSE WAS PROMOTING YESTERDAY’S SOUND, SONY PERFECTED THE NEXT GENERATION

SONY’s NEW line of high quality full-featured duplicators will meet all of your audio cassette duplication needs; whether voice or music, monaural or stereo. SONY systems can reproduce from 1 to 43 copies in a single pass while being simple to use. Further, they include features such as short tape indicator, audio end detect, A or A+B selection, manual or fully automatic duplication, and superior frequency response.

And only SONY duplicating systems carry a full TWO YEAR WARRANTY against head wear. All in all, SONY high speed duplicating systems last longer, cost less and produce superior reproductions...THE NEXT GENERATION of duplicating systems has arrived.

SONY AV PRODUCTS

NATIONAL DISTRIBUTOR
Educational Electronics Corporation
P.O. Box 339 • Inglewood, CA 90306-0339 • 213-677-8167
Branching out...

NEW ESL TITLES FOR 1986 FROM HBJ

Grammar
HARBRACE ESL WORKBOOK
Sheila Y. Graham and Wynn J. Curtis
Paperbound

UNDERSTANDING AMERICAN SENTENCES
Lily A. Kapili and Ber H. Kapili
Paperbound

Reading & Writing
GATEWAY
A Cross-Curricular Reading & Writing Book
Len Fox
Paperbound

Writing
BRIDGES
From Sentence to Paragraph
Beth Snyder and Barbara Auerbach
Paperbound

Conversation
GETTING TOGETHER
Susan Stempleski, Abson Rice, and Julie Falsetti
Paperbound

Readers
WORKING WORLD
Language and Culture of the Job Market
Maria Maniscalco Baskin and Lois Wasserman Manton
Paperbound

NEW AMERICAN PROFILES
Lecette Rollet Kenan
Paperbound

PORTS OF ENTRY
Scientific Concerns
Abelle Mason
Paperbound

ENCOUNTERS
Third Edition
Paul Pimsleur, Beverly Pimsleur, and Donald Berger
Paperbound

THE AMERICAN SCENE
A Basic Reader for ESL
Mary Thurber
Paperbound

For further information, please write:
HARCOURT BRACE JOVANOVICH, PUBLISHERS
College Department
1250 Sixth Avenue, San Diego, CA 92101
The Sound Pronunciation Cards provide pictorial illustrations of proper placement of the articulators for correct pronunciation of 24 consonants, 12 vowels, and 4 diphthongs—the sound system of Standard American English!

Contents:
- 40 cards 3¾" × 5½" illustrating lip, teeth, tongue, and jaw positions
- 2 90-minute cassettes with model pronunciations of speech sounds & words

Cards are appropriate for:
- Teaching English as a second language
- Modifying social/regional and foreign dialects
- Improving articulation and pronunciation
- Instructing phonics-based reading
- Training the hearing impaired

Author: Bernadette Anderson, Ph.D., CCC-SP

SOUND PRONUNCIATION CARDS

ORDER HERE
☐ Complete set of cards, posters and cassettes $90.00
☐ Set of cards only $30.00
☐ 2 90-minute cassettes only $20.00
☐ 40 11" × 17" Sound Pronunciation Posters $45.00
☐ More information on the set

(Prices include postage and handling. Please send check or money order—no cash.)

Charge my ☐ Visa ☐ MasterCard

No. & exp. date:
Signature:

NAME:
INSTITUTION:
ADDRESS:
CITY/STATE/ZIP:

Send to: Communication Improvement Center, 30 W. Washington, Suite 1324, Chicago, IL 60602, (312) 842-0516
KING ABDULAZIZ UNIVERSITY
JEDDAH—SAUDI ARABIA

ENGLISH LANGUAGE CENTER

The ELC provides courses in scientific and technical English for approximately 1,100 male and 500 female students drawn from six colleges at King Abdulaziz University.

Courses range from intensive programs of 20 hours per week for foundation students to specialised reading skills courses of five hours per week. The Center has an academic staff of 60 well-qualified teachers recruited mainly from the U.K. and U.S.A. and is supported by fully equipped in-house media services -- a graphics studio, a reprographics and printing shop, TV studio, photographic studio and two language laboratories. A 20-position Apple computer laboratory has recently been established and now offers a variety of programs in computer-assisted language learning.

The ELC is currently recruiting well-qualified teachers with relevant experience for June 1986. Teaching couples are especially encouraged to apply.

QUALIFICATIONS: M. A. in TEFL/TESL or Applied Linguistics.

EXPERIENCE: Minimum two years’ experience teaching TEFL/TESL overseas.

STARTING SALARY: Attractive salaries depending on qualifications and experience.

BENEFITS:
1. Generous housing and furniture allowances;
2. End of contract gratuity;
3. Sixty days paid summer leave each year;
4. Baggage allowance;
5. Educational allowance for school-age children;
6. Transportation allowance;
7. Annual round-trip air fares.

CONTRACT: One-year, renewable.

APPLICATIONS: Write, including full resume, current telephone contact and photocopies of qualifications and letters of experience to:

SUPERVISOR GENERAL
ENGLISH LANGUAGE CENTER
KING ABDULAZIZ UNIVERSITY
P.O. BOX 1540,
JEDDAH-21441
SAUDI ARABIA.
1986 TESOL SUMMER INSTITUTE
A RAINBOW OF PERSPECTIVES IN TESOL

A program designed for beginning and experienced teachers, administrators, researchers and graduate students interested in broadening their knowledge and skills.

JULY 7 - AUGUST 15
UNIVERSITY OF HAWAII AT MANOA

Visiting professors and University of Hawaii faculty members will offer a full range of introductory and advanced courses in teaching and learning English as a second or foreign language, bilingual and deaf education.

Three-week and six-week credit classes, workshops, colloquia, lectures, demonstrations and panels will be offered.


For more information write:
Pamela Pine, Assistant Director, 1986 TESOL Summer Institute, Department of ESL, University of Hawaii at Manoa, 1890 East-West Road, Honolulu, Hawaii 96822 USA

TESOL
Teachers of English to Speakers of Other Languages
The TESOL Quarterly is available in Microform.

Please send additional information for The TESOL Quarterly.

Name ____________________________

Institution _______________________

Street ____________________________

City ______________________________

State ___________ Zip ______________

300 North Zeeb Road, Dept. P.R., Ann Arbor, Michigan 48106 U.S.A.

University Microfilms International
TESOL MEMBERSHIP APPLICATION

Please Do Not Use for Renewal Membership—For New Membership Only

MEMBERSHIP APPLICATION ONLY

Name (Print) ____________________________

Mailing Address ____________________________

City __________________ State ____________ Zip. __________

Position Title ____________________________

Institution ____________________________

Please make check in U.S. funds drawn on a U.S. bank payable to TESOL and mail to:

TESOL
Georgetown University
Suite 205
1118 22nd Street, N.W.
Washington, D.C. 20037, U.S.A.

MEMBERSHIP CATEGORIES
Membership includes subscription to TESOL QUARTERLY and TESOL NEWSLETTER.
Check appropriate box:

☐ REGULAR .......................................................... $ 40.00
☐ STUDENT (for those engaged in at least half-time study) .......... $ 20.00
☐ JOINT (two-member household) ................................ $ 60.00
☐ INSTITUTIONAL/LIBRARY ....................................... $ 75.00
☐ COMMERCIAL ..................................................... $200.00
☐ PARAPROFESSIONAL ☐ RETIRED ☐ UNEMPLOYED or ☐ VOLUNTEER $ 20.00

Additional mailing fee
☐ Foreign Surface Mail add $5.00
☐ Foreign Air Mail add $15.00

INTEREST SECTIONS
MARK with (x) the Interest Section in which you wish to be active and vote.
MARK with (—) not more than two additional Interest Sections in which you have a secondary interest.

☐ Teaching English Internationally
☐ English as a Foreign Language, for Foreign Students in English-speaking Countries
☐ English to Speakers of Other Languages, in Elementary Education
☐ English as a Second Language, in Secondary Schools
☐ English as a Second Language, in Higher Education
☐ English as a Second Language, in Bilingual Education
☐ English as a Second Language, in Adult Education
☐ Standard English as a Second Dialect
☐ Applied Linguistics (relevant linguistic studies and research)
☐ Research
☐ Refugee Concerns
☐ Teacher Education
☐ Computer-Assisted Language Learning
☐ Program Administration

176
ESL Software That *Talks* To Students

- Oral language instruction
- Supplementary to ongoing instruction
- Functional English language skills
- Cross referenced to other popular ESL materials

*Harmony™* is a 14-title microcomputer software series developed by the Houston Independent School District specifically for elementary-aged ESL students.

Available for use with the Apple II Plus and Apple II/e microcomputers, Harmony uses Jostens' Ufonic® voice synthesis system to provide students with natural human speech simultaneously with pedagogically-sound instructional content.

Harmony materials include Instructional Disks, Skillbuilder Disks, and Student Workbooks—all designed to provide students with meaningful, critical, and communicative English language skills that they can use immediately—at school, at home, and at play!

The Harmony Starter Package, that includes the Ufonic voice system hardware, is available immediately. For additional information or to arrange for a demonstration, please write or call our toll-free number today.

1-800-323-7577 1-800-942-6995
National  in Illinois

**JOSTENS LEARNING SYSTEMS, INC.**  800 East Business Center Drive  Mt. Prospect, Illinois 60056
"...the best program on the market..."
—John W. Oller, Jr.
University of New Mexico

EXPRESS ENGLISH:
Beginnings 1 and 2
Transitions

The beginning and intermediate levels of a bold, new communicative series

LINDA A. FERREIRA

NEWBURY HOUSE
PUBLISHERS, INC.
ROWLEY, MA 01969
Order Dept.: 1-800-343-1240
In Mass., call: (617) 948-2840
Main Office: (617) 948-2704
Telex: 178296 NHP UT

To order an examination copy of the Student Text, please call or write the Order Department