THE EFFECTS OF REFERENTIAL QUESTIONS ON ESL CLASSROOM DISCOURSE

Cynthia A. Brock
ABSTRACT

Long and Sato (1983), in an examination of ESL teachers' questions in the classroom, found that teachers ask significantly more display questions, which request information already known by the questioner, than referential questions, which request information not known by the questioner. The main purpose of the present study was to test for some possible effects on adult ESL classroom discourse of higher frequencies of referential questions.

Four experienced ESL teachers and 24 non-native speakers (NNSs) enrolled in the University of Hawaii's English Language Institute participated in this experimental study. Two of the teachers were provided with training in referential questions; the other two teachers formed a control group not provided with training. All four teachers taught the same reading and vocabulary lesson to one group each of six NNSs.

It was hypothesized, on the basis of first language classroom research on attempts to increase the cognitive level of teachers' questions, that teachers could, with training, increase their use of referential questions. It was further hypothesized, on the basis of first language classroom research on the effects on student responses of higher cognitive
questions, that NNSs' responses to referential questions would be longer and more syntactically complex than their responses to display questions. The above hypotheses were supported by the results obtained.

Mot supported by the data were hypotheses predicting greater numbers of confirmation checks and clarification requests accompanying greater numbers of questions.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>CHAPTER I. REVIEW OF THE LITERATURE</td>
<td>1</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>18</td>
</tr>
<tr>
<td>RESEARCH QUESTIONS</td>
<td>19</td>
</tr>
<tr>
<td>CHAPTER II. METHOD</td>
<td>23</td>
</tr>
<tr>
<td>SUBJECTS</td>
<td>23</td>
</tr>
<tr>
<td>DESIGN</td>
<td>23</td>
</tr>
<tr>
<td>PROCEDURES</td>
<td>24</td>
</tr>
<tr>
<td>ANALYSIS</td>
<td>26</td>
</tr>
<tr>
<td>CHAPTER III. RESULTS</td>
<td>34</td>
</tr>
<tr>
<td>CHAPTER IV. DISCUSSION</td>
<td>53</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>66</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>76</td>
</tr>
<tr>
<td>NOTE</td>
<td>79</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>80</td>
</tr>
</tbody>
</table>
CHAPTER I

REVIEW OF THE LITERATURE

An abundance of questions is a hallmark of second language (SL) learners' exposure to the target language. In informal conversation between native speakers (NSs) and beginning-level non-native speakers (NNSs) questions are the form most frequently used by NSs to initiate topics and, as a consequence of the frequent shifts in topic, the dominant form used to address NNSs (Long 1981).

NSs' preference for questions in topic initiation may be due to the obligation to respond they generate, the assistance they provide to the NNS in the form of partially or fully preformulated responses and the salience added by the linguistic features, such as rising intonation, which distinguish them (Long 1981). Given that in many Third World societies "conversation...is the context known to be capable of producing fluent sequential bilinguals," (Long 1982:215), then questions may be a crucial input feature fostering development of SL abilities.

However, in spite of the prominence and possible importance of questions in language learners' input outside of the ESL classroom, they have been the subject of little research within the classroom. This
is despite the growing interest in classroom processes (Long 1980) and despite the fact that they seem to be just as pervasive a feature of ESL classroom discourse as they are of informal NS-NNS conversational discourse. White and Lightbown (1984), for example, counted 427 questions asked by an ESL teacher in a single 50-minute class. In fact, it appears that Long and Sato's (1983) study of the forms and functions of ESL teachers' questions in the classroom is the only one of its kind to date.

Analyzing the classroom speech of six teachers, as well as the speech of 36 NSs in informal conversations with NNSs, they found significant differences in the relative proportions of two types of questions, display and referential, asked in the two settings. Display, or known information, 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. The ESL teachers asked significantly more display than referential questions in the classroom, while the NSs in the informal conversational setting asked a majority of referential and virtually no display questions.

Since questions constrain what can appropriately be said in response (Stubbs 1983; Keenan, Schieffelin,
and Platt 1978), it may be the case that these two types of questions, display and referential, shape the language of responses to them in different ways. However, there do not appear so far to have been any investigations of this issue. The purposes of the present study were, accordingly, as follows: first, to investigate whether ESL teachers who received training in referential questions would ask greater numbers of them than teachers who did not receive training; and, second, if greater numbers of referential questions were asked by teachers receiving such training, to test for some possible effects on several features of ESL classroom discourse. With respect to the students' language, these features included the length and syntactic complexity of responses to the two types of questions and the number of connectives used in learner speech. With respect to the teachers' language, these features included the number of clarification requests and confirmation checks made.

In contrast to the general 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. While question types do not appear to be defined in these studies in terms of the distinction between display and referential, data are provided in them...
concerning at least four major issues of relevance to the present study: the intellectual level of teachers' questions; the relationship between student achievement and the use of questions at higher intellectual levels; the degree to which teachers can be trained to change the types of questions they ask; and the relationship between the types of questions teachers ask and certain features of their students' responses.

The intellectual or cognitive levels of questions are defined, in most of the studies reviewed here, according to two classification systems. One, developed by Bloom and his colleagues (1956), assigns questions to one of six levels. At the lowest level of this hierarchy are those questions calling for the recognition or recall of factual information, while at level two are questions dealing with comprehension and calling for explanation, interpretation or extrapolation. At the third, fourth and fifth levels are questions asking for the application of factual knowledge, for the analysis of relationships between elements, and for generalizing or synthesizing, respectively. Finally, questions at the highest cognitive level call for evaluation or judgment.

A second frequently-used system for classifying teachers' questions was developed by Gallagher and Aschner (1963). There are four cognitive categories in
the Gallagher and Aschner system, which is based on Guilford's (1956) theoretical structure of thinking. At the lowest level of this hierarchy are "cognitive-memory" questions, which call for the recognition or recall of factual information within a pre-determined and tightly structured framework, while "divergent" questions allow the respondent "to generate independently his own data within a data-poor situation or to take a new direction or perspective on a given topic" (Aschner and Gallagher 1963:187). At the highest level are evaluative questions, which call for expressions of judgment.

Regardless of the classifications system used, however, research results indicate that teachers tend to ask questions at low cognitive levels, primarily at the level of factual recall or recognition. This is true both in elementary schools (Guszak 1967; Willson 1973) and in secondary schools (Davis and Tinsley 1967; Gallagher 1965), and it is probably reasonable to assume that these questions at low cognitive levels, asking for factual recall or recognition, are display questions.

Despite this apparent tendency among teachers to question at low cognitive levels, there is evidence that they can, with training, increase the frequency in their classroom speech of questions at higher cognitive
levels (Gall 1970; Rogers and Davis 1970; Galassi, Gall, Dunning and Banks 1974; Chewprecha, Gardner and Sapianchai 1980; Arnold, Atwood and Rogers 1974). At least one study, however, suggests that the effects of training in questioning techniques may be short-lived (Crawley and Krockover 1979), and two others indicate that not all methods of training are equally effective. Swift and Gooding (1983) found that teachers receiving printed instructional material on questioning techniques once a week for eight weeks did not increase their use of higher level questions, and Galassi and his co-workers (1974) found that written transcripts of classroom dialogs were a more effective training tool than videotapes of the same dialogs.

Unfortunately, the effects on student achievement of an increase of questions at higher cognitive levels are unclear. Winne (1979), in a critical review of 86 studies addressing this question, found that, of the 53 studies he considered valid, 64% showed no effect on student achievement of a predominance of higher cognitive questions or of fact questions. Twenty-six percent showed that a predominance of fact questions was more beneficial to student achievement than a predominance of higher cognitive questions, and only ten percent showed the reverse.

In comparison with the substantial, albeit
inconclusive, body of research on the relationship between the cognitive level of questions and student achievement, there is very little research of possible correspondences between the level of the teacher's question and features of the student's response. However, the research that has been done has suggested, by and large, that the level of the question does have an effect on the student's response.

Some researchers have analyzed student responses to determinewhether they are at the same level of the intellectual hierarchy as the teacher's question. In an early descriptive study of junior high classroom interaction, Gallagher and Aschner (1963) found that an increase in divergent questions by teachers was associated with an increase in the amount of divergent production by students. Similarly, Willson (1973) found, in an experimental study of second, fifth and sixth grade social studies discussions, that an increase in the mean cognitive level of questions asked by the teachers was accompanied by an increase in the mean level of the 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 and Rosseau 1980), Arnold, Atwood and Rogers (1974) did find a significant one-to-one correspondence between the
question level and the level of student response.

A few researchers have examined the effects of different question types on other aspects of student responses, such as their length and syntactic complexity. Although the results of these studies are not strictly comparable since no two investigations seem to employ the same system of analysis, the general pattern appears to be that responses to lower order questions, those calling for the recognition or recall of factual information, are shorter and less syntactically complex than responses to higher order questions calling for analysis, interpretation, prediction or the expression of subjective knowledge.

Smith (1978), for example, conducted two separate research studies yielding similar results. In the first, she compared the responses to factual questions with those to interpretive questions of 30 second graders and 30 fourth graders chosen at random from the second- and fourth-grade population of an elementary school. Interpretive questions "involved... analysis, reconstruction or inference of relationships" (Smith 1978:898). Second and fourth graders were chosen because they are presumed to be at two distinct Piagetian stages of cognitive development, the pre-operational period and the concrete operations period, respectively.
Each child was interviewed individually and asked factual and interpretive questions about stories and pictures presented during the interview, and both the second and fourth graders' responses to the interpretive questions were significantly longer than their responses to the factual questions. Furthermore, while the fourth graders' responses were longer than the second graders' responses to interpretive questions—presumably reflecting the difference in their cognitive development—there was no difference in their answers to factual questions.

In the second study, Smith analyzed both elementary and secondary students' responses to teacher questions at two levels: the "narrow" and the "broad". Narrow questions require recall or recognition of information. The broad questions were focusing questions requiring the students to develop a particular idea or answer by leading them toward it through clues as to what the answer was or as to the appropriate method of obtaining the answer... The broad questions allowed for several acceptable answers. (Smith 1978:898)

The twenty teachers participating in this descriptive study were also participating in a graduate level
teacher education project aimed at improving their questioning techniques, and as part of this project they designed broad as well as narrow questions. Tape recordings were made of the teachers asking the two types of questions when they returned to their classrooms, and transcripts of the tapes were analyzed by Smith. Unfortunately, she apparently conducted the analysis alone, without a second rater to help establish the validity of her categories (Frick and Semmel 1978). However, results of this analysis showed a significant difference in student response to the two types of questions. Answers to broad questions were significantly longer than answers to narrow questions.

Cole and Williams (1973) also carried out a descriptive investigation of the relationship between the cognitive levels of eight elementary school teachers' questions and the length of their students' responses. In addition, possible relationships were sought between the cognitive level of the teachers' questions and the cognitive level as well as the syntax of the students' responses.

Slightly modifying the Gallagher and Aschner (1963) cognitive hierarchy by combining the two categories at the top, Cole and Williams classified the teachers' questions and students' responses according
to three levels: cognitive-memory; convergent; and divergent and evaluative. Student responses were further categorized, with respect to length, as consisting of either one to three words, four to nine words, or ten or more words. Three syntactic categories were used for student responses: a word or phrase; a simple sentence; or a compound sentence or multiple sentences. An interrater reliability figure, presumably representing simple percentage agreement, of 0.92 as reported for Cole and Williams' use of these categories to code the questions and responses, and the results of their analysis indicate a strong positive relationship between the cognitive level of the teacher's questions and cognitive level, length and syntactic complexity of the pupil's response.

The results of at least one study, while consistent with the general pattern reported above, also include a finding which runs counter to what intuition might lead one to expect. Dillon (1981), in a descriptive study, classified teacher questions in a number of different ways, including a "fact" versus "opinion" dichotomy which is analogous to the lower and upper levels of the other cognitive-level systems. In addition, he classified questions as appearing either as part of a series of questions or in isolation, i.e., not preceded or followed by another question. He also
classified questions with respect to their respondent. That is, a question might be directed to or answered by either an individual student or several students.

Finally, he classified questions with respect to their structure:

The syntactic structure of [a question] indicated the minimum amount of response adequate on grammatical grounds. A closed [question] was so structured that a single word or phrase was sufficient in response. An open [question] required at minimum several phrases or a sentence. (Dillon 1981:2-3)

With two exceptions, there was no significant difference between types in any of the ways Dillon classified the teachers' questions. There was a difference between fact and opinion questions, with student response to opinion questions lasting significantly longer than the response to fact questions, and there was a difference in the length of response to open as opposed to closed questions. However, contrary to what one might expect, the mean length of response to closed questions was significantly longer than the mean length of student response to open questions.

These results may be due, though, to Dillon's
definition of student response as "the duration of student talk following upon one teacher utterance and terminating at the next" (Dillon 1981:2). Perhaps if he had examined individual students' responses, the patterns he found might have been different, given that a number of students might have produced short responses one after another in a series.

In addition to those studies focusing exclusively on the effects on classroom discourse of different question types, there is some overlap between research in this area and research on another classroom process variable, teacher wait-time. In an early study of this variable, which refers to the length of time a teacher pauses after asking a question or after a student has given a response, Rowe (1974) found that after teachers were trained to increase their wait-times to a criterion level of three seconds, there were a number of changes in their students' classroom language. Among these changes was a significant increase in the average number of words in their responses to questions. Although she does not define the terms or provide any relevant figures, Rowe also reports an increase in the amount of "speculative thinking" exhibited in their responses and in the amount of "evidence followed by or preceded by inference statements."
In addition to the changes in the students' language, Rowe found alterations in certain characteristics of the teachers' classroom language as well, including the kinds of questions they asked. Using Parsons' (1971) categories, which are defined neither in Rowe's paper nor in the original Parsons source, she classified teachers' questions as rhetorical, informational, leading or probing. Although, again, Rowe does not report relevant statistics, such as the mean proportion of informational questions asked before and after the wait-time training, she does offer the proportions of these types of questions asked by what she terms the "typical" teacher in the study. This profile of the typical questioning pattern suggests that the teachers asked predominantly informational questions before the training whereas afterwards the proportion of leading and probing questions increased substantially.

Similarly, Swift and Gooding (1983) found that teachers receiving successful training in increasing wait-times also asked significantly greater numbers of divergent and evaluative questions without having had any training in questioning techniques. Their students' responses were longer and the percentage of student talk was greater than in the classrooms of comparison group teachers not receiving the training.
Because alterations in question types is associated with longer student responses, and because it seems that teachers who increase their wait-times also tend to alter their questioning patterns, it is difficult to say to what extent the increase in questions at higher cognitive levels contributed to the lengthier student responses reported by Rowe and by Swift and Gooding.

Commenting on the fact that teachers' questioning patterns seem to alter as they increase their wait-times, Rowe remarks that

This is not to suggest, however, that giving explicit training in how to ask questions is not desirable. But it might not be totally surprising to find that people who receive such training are inadvertently slowed down, i.e. have longer wait-time. (Rowe 1974:221)

However, Tobin and Capie (1982), in the report of their study investigating the relationships between student achievement and a number of classroom processes, provide data which suggest that this is not necessarily the case. They randomly assigned middle-school science teachers and their intact classes to four different treatment groups: one receiving feedback designed to increase wait-time, a second
receiving feedback designed to increase the cognitive level as well as the clarity and relevance of teacher questions, a third receiving both kinds of feedback, and a fourth receiving a placebo treatment.

Although the results of a discriminant analysis showed that the treatments were effective for each group, the teachers receiving only the feedback on wait-time achieved a mean score on the cognitive level of their questions only slightly below that of the teachers receiving only the feedback on questioning techniques. The teachers receiving only the feedback on questioning techniques, on the other hand, achieved a mean wait-time only slightly longer than that of the control group.

It does not seem to be the case, then, that a change in teachers' questioning patterns is necessarily accompanied by an increase in wait-time. However, the available evidence suggests that an increase in wait-time is accompanied by an increase in the cognitive level of questioning. Apparently there are no studies comparing the relative effects on the length and syntactic complexity of student responses of alterations in these two variables.

Most of the classification systems used in the studies referred to above describe the same intellectual continuum with the points along this
continuum simply named and defined in different ways. It seems reasonable that questions at the higher levels of the continuum, such as those calling for judgment, are likely to be referential questions, and questions calling for factual recall are likely to be display questions. However, the explicit distinction between display and referential questions seems not to have figured prominently in first-language classroom research, even though, as Mehan observes, "the use of known information questions has consequences for the knowledge that children display in the classroom" (Mehan 1979:291). 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" (Mehan 1979:294).

That the use of known information, or display, questions in the classroom generates discourse which is fundamentally different from everyday, or normal conversational, discourse is an important consideration for language teachers since everyday discourse is likely to be the target discourse for many second language learners. 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 conversational discourse learners are likely to experience outside of the classroom. Long and Sato (1983) reported that 76% of the questions in the 36 NS-NNS informal conversations they analyzed were referential, while only two questions out of 1,567 were display. As Long observes,

   if, as seems a reasonable assumption, practice at various kinds of speaking, like most tasks, is conducive of success, then the amount as well as type of practice [classroom second language learners] are allowed to engage in is important, as is...the study of how and why they obtain those opportunities in classrooms, or fail to do so. (Long 1979:3)

PURPOSE

It was assumed, on the basis of the Long and Sato (1983) findings, that teachers not trained in questioning techniques would ask a preponderance of display questions. It was hypothesized, however, on the basis of the studies of attempts to increase the
cognitive level of questions in the first-language classroom referred to above, that teachers could increase their use of referential questions. The main purpose of this study was to test for some possible effects on adult ESL classroom discourse of higher frequencies of referential questions.

RESEARCH QUESTIONS

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

(1) Teachers receiving a training session in the formation and use of referential questions will ask more referential questions in the classroom than teachers not receiving the training.

If the number of referential questions asked by teachers could be increased, it was expected that this increase would have a number of effects on classroom discourse. First, with respect simply to the volume of learner speech in the classroom as well as the syntactic richness of that speech, it was hypothesized that

(2) NNSs' responses to display questions are
shorter than their responses to referential questions.

(3) **NNSs'** responses to display questions are less complex syntactically than their responses to referential questions.

Since display questions are attempts to elicit from students information already possessed by the teacher, the student's response is matched up with what the teacher considers to be the correct answer and evaluated accordingly. There is no need for the teacher to check his or her understanding of the content of a student's response. The significant paucity of confirmation checks in ESL classroom speech (as compared to NS-NNS conversation) may be a reflection of the predominance in teacher questioning of display questions (Long 1983).

Referential questions, on the other hand, because they request information not already known by the teacher, may lead to the necessity for confirmation checks and clarification requests. A teacher may need to verify that *s/he* has accurately understood the information contained in a student's response. Since relative increases in confirmation checks and clarification requests in NNS-NS conversation (as compared to NS-NS conversation) form part of the body
of conversational adjustments posited to be necessary for second language acquisition (SLA) (Long 1983), it would be useful to know which classroom processes foster their occurrence. Accordingly, it was hypothesized that

(4) A greater number of referential questions is accompanied by a greater number of confirmation checks by the teacher.

(5) Confirmation checks by the teacher occur more frequently following referential questions than following display questions.

(6) A greater number of referential questions is accompanied by a greater number of clarification requests by the teacher.

(7) Clarification requests by the teacher occur more frequently following referential questions than following display questions.

As Mehan 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. (Mehan 1979:291)

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 of establishing their linear coherence as well.

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 which are necessary to form linearly coherent sequences (van Dijk, 1977a). Since these "connections between propositions are typically expressed by natural connectives such as and, because, yet, so, etc." (van Dijk 1977b:5), it is hypothesized that

(8) A greater number of referential questions will be accompanied by a greater number of connectives in learner speech.
CHAPTER II

METHOD

SUBJECTS

The subjects (Ss) for this study were 24 NNSs enrolled in classes in the University of Hawaii's English Language Institute (ELI). One of the Ss was from Afghanistan; the other 23 were from East Asian countries: Korea, China, Taiwan, Cambodia, Vietnam, and Japan. Sixteen of the Ss were enrolled in the ELI's most advanced courses where students' TOEFL scores typically average above 490; eight Ss were enrolled in a course at the next level below the most advanced where students' TOEFL scores typically average between 470 and 520.

DESIGN

Four groups of six Ss each were formed using a randomized block design to control for the difference in proficiency among Ss: the 16 more-advanced learners were randomly assigned as a block to the four groups, and the eight less-advanced learners were randomly assigned as a separate block to these same four groups.
Four ESL teachers, two women and two men, all with at least five years of ESL teaching experience and all enrolled in the MA program at the University of Hawaii, were assigned to a treatment and a control group. A randomized block was again used to control for gender. The two women were randomly assigned to the treatment and control groups as a block, and the two men were randomly assigned as a separate block to these two groups.

There were two teachers (one man and one woman) in the control group, then, and two in the treatment group. Each teacher was randomly assigned one of the groups of six 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. The passage, presented in its entirety in Appendix A, 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 two teachers in the control group. They were given the passage and the list of vocabulary items and instructed to allow students the first twenty minutes of the period for reading. The second twenty minutes were 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 two teachers in the treatment group were given the same reading passage and 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, while it contained the same items as that of the control group, also included
a sample referential question for each item. (Appendix B contains this list.) 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.

All four lessons were conducted at the same time on the same day, and in order to ensure that the lessons were of equal duration, the researcher provided a signal to the teachers to begin and to end the discussion portion of the lesson. The lessons were audiotaped, and the tapes transcribed for analysis.

ANALYSIS

Long and Sato's (1983) adaptation of Kearsley's (1976) taxonomy was used to categorize question types. Display or "evaluative" questions, according to Kearsley, "are asked to establish the addressee's knowledge of the answer" (Kearsley 1976:366). An example of this type of question from the corpus of the present study is: "What does temperament mean?" Referential questions are "intended to provide contextual information about situations, events.
actions, purposes, relationships, or properties" (Kearsley 1976:367). For the purpose of this study, it must be emphasized that referential questions request information not possessed by the questioner. An example from the corpus of this type of question is: "Do any of you have Filipino friends?"

To test Hypothesis (1), 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.

To test Hypothesis (2), mean lengths of response (in words) were calculated for Ss' responses to display questions and to referential questions. 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. The only exception to this was a contribution by the teacher characteristic of the back channel which occurred internal to the boundaries of a S's communication unit (defined below) and did not disrupt that communication unit (c-unit). If, however, such a contribution occurred at the boundary of a c-unit, the S's response was considered to have ended. An example from the corpus may clarify the distinction:'
Are you from the Philippines?

No I not. But other Asian

Yeah.

tend to uh to be with a patient during the

time they sick.

OK yeah.

To be with them.

In line 3 of the example the teacher's message occurs inside the boundary of the S's current c-unit without disrupting its continuation by the S in line 4. The teacher's expression of agreement in line 6, on the other hand, occurs after the end of the S's current c-unit and therefore marks the end of the S's response.

Words, for the purpose of this study, did not include pause fillers such as "uh." However, minimal expressions of assent such as "uh-huh" and "mhm" were counted as single words.

Repetitions of words were not counted in the analysis, and contractions were counted as single words. For example, in "They- they don't complain" "they" was counted only once, and "don't" was counted as one word. Proper nouns such as "Hong Kong" or "New York" were also counted as single words. If a S repeated an entire c-unit, as in "ill person, ill person," that c-unit was counted only once.

Finally, apparently semantically empty phrases and
words were left out of the analysis: "you know"; "(or) (something) like that"; and "well" when it was used not as an adverb but as a signal of the onset of a turn.

Hypothesis (3) was tested by measuring the mean number of sentence-nodes (s-nodes) per communication unit. The communication unit (c-unit) was identified by Loban (1966) and defined as

either independent grammatical predications
or answers to questions which lack only the repetition of the question elements to satisfy the criterion of independent predication. Given this definition, the single word 'yes' can be admitted as a whole unit of communication when it is an answer to a question. (Loban 1966:7)

In adapting the c-unit as a measure of NNS speech, a segment 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.

A further adaptation was the expansion of the notion of the repetition of question elements to permit analysis of questions shaped by their place in the discourse context, as in the following example from the corpus:
1 T: For example, was your grandfather or is your
grandmother very influential in family
decisions?
2 S: My grandmother.
3 T: Oh your grandmother. More so than your
grandfather?
4 S: Uhhuh.

The teacher's question in lines 5 and 6 lacks only repetition of elements from the previous question to qualify as an independent predication. Similarly, the S's response in line 7 lacks only the repetition of elements both present and ellipted in the teacher's question to qualify. Communication unit in this study was therefore defined so as to include segments which lacked only repetition of question elements, either present or ellipted, to satisfy the criterion of independent grammatical predication.

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." (Freed 1978:43) Infinitives and gerunds, then, as well as tensed verbs, were taken to signal an underlying s-node. Modals, such as "can" and "have to," on the other hand, were not considered to be signals of underlying s-nodes.

The mean number of s-nodes per communication unit
in learner responses to referential questions was compared with the mean number of s-nodes per communication unit in learner responses to display questions.

To test Hypotheses (4) through (7), the transcripts were coded for confirmation checks and clarification requests following the definitions in Long and Sato (1983). That is, confirmation checks are either Yes/No or uninverted questions spoken with rising intonation that involve exact or semantic, complete or partial repetition of the previous speaker's questions, and serve either to elicit confirmation that their user had heard and/or understood the previous speaker's previous utterance correctly or to dispel that belief. (Long and Sato 1983:275)

Clarification requests, on the other hand, are any expressions by a NS designed to elicit clarification of the interlocutor's preceding utterance... They require that the interlocutor either furnish new information or recode information previously given. (Long and Sato 1983:276)

To test Hypotheses (4) and (6), totals of each of
these two types of repair were calculated for the treatment and control groups. That is, for Hypothesis (4), the total number of confirmation checks asked by the two teachers in the treatment group was tallied and compared with the total number asked by the two teachers in the control group. For Hypothesis (6) the total number of clarification requests asked by the treatment group teachers was compared with the control group teachers' total.

To test Hypothesis (5), the total number of confirmation checks made by all four teachers in their turns immediately following learner responses to referential questions was compared to the total number of confirmation checks in turns immediately following learner responses to display questions. Similarly, to test Hypothesis (7), the total number of clarification requests made by the teachers in their turns coming immediately after learner responses to referential questions was compared to the total number made in turns immediately following learner responses to display questions.

To test Hypothesis (8), the total number of connectives 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 the class of connectives was
defined as those appearing in the extensive list compiled by Celce-Murcia and Larsen-Freeman (1983:371-377), and only those initiating a clause were counted. No connective was included in the tally if the clause it appeared in was interrupted, either by self-interruption or interruption by another speaker. A random sample from the corpus containing 119 questions was coded by a second rater for display and referential questions, confirmation checks, clarification requests and comprehension checks. The simple percentage nominal agreement for these five categories was 0.91. However, using Cohen's kappa, which corrects for asymmetrical marginals and for chance agreement (Cohen, 1960; Frick and Semmel 1978), the nominal agreement obtained was 0.76.

In all hypothesis testing the acceptable level of probability was set at 0.05.
CHAPTER III

RESULTS

For the purposes of reporting the results, the teachers in the control group will be referred to as Teacher 1 (T1) and Teacher 2 (T2), while the two teachers in the treatment group will be referred to as Teacher 3 (T3) and Teacher 4 (T4).

Hypothesis 1: The treatment group teachers receiving a training session in the formation and use of referential questions will ask more referential questions in the classroom than the control group teachers not receiving the training. The two 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 (Table 1). Table 2 presents a comparison of the number of referential questions asked by teachers in the two groups. 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 teachers who received the training in the formation of referential questions asked significantly more of them than the teachers who did not receive the training ($\chi^2 = 93.58$, df = 1, $p < 0.001$), using the correction for continuity for a one-way $\chi^2$ with one degree of freedom.

Table 1

Frequency of Referential and Display Questions in the Control and Treatment Groups

<table>
<thead>
<tr>
<th></th>
<th>Number of referential questions asked</th>
<th>Number of display questions asked</th>
<th>Total number of questions asked</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T1 and T2)</td>
<td>24</td>
<td>117</td>
<td>141</td>
</tr>
<tr>
<td><strong>Treatment group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T3 and T4)</td>
<td>173</td>
<td>21</td>
<td>194</td>
</tr>
</tbody>
</table>

$\chi^2 = 172.52$, df = 1, $p < 0.001$
### Table 2
**Referential Questions Asked by Teachers in Treatment and Control Groups**

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted number of referential questions</th>
<th>Adjusted number of referential questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Group (T1 and T2)</strong></td>
<td>24</td>
<td>33.12</td>
</tr>
<tr>
<td><strong>Treatment Group (T3 and T4)</strong></td>
<td>173</td>
<td>173</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 93.58, \, df=1, \, p < 0.001 \]

### Table 3
**Frequencies of Display and Referential Questions for Individual Teachers**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Treatment group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td><strong>Display</strong></td>
<td>59</td>
<td>58</td>
</tr>
<tr>
<td><strong>Referential</strong></td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>73</td>
<td>68</td>
</tr>
</tbody>
</table>
Table 3 shows the frequencies of the two types of questions for each of the four teachers. An examination of this table suggests that individual teachers' performances in each of the two groups reflected, in a roughly equivalent manner, the patterns shown in Table 1 of the balance between the two question types for their respective groups.

Hypothesis 2: Responses to display questions are shorter than responses to referential questions. Table 4 presents the mean lengths of responses to the two question types by all learners in all four classes. 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 ($t=3.92$, $df=221$, $p < 0.0005$).

Tables 5 and 6 show the mean lengths of responses to referential and display questions by learners in each of the four classes. An examination of these tables suggests variability in these measures within the treatment and control groups.

Hypothesis 3: Responses to display questions are less complex syntactically than responses to referential questions. Table 7 presents the results for responses to the two question types by all learners in all four groups. The mean number of s-nodes per
Table 4
Mean Length (in Words) of Learner Response

<table>
<thead>
<tr>
<th>Learner response to</th>
<th>Mean</th>
<th>Number of responses</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>referential questions</td>
<td>10.00</td>
<td>121</td>
<td>15.17</td>
</tr>
<tr>
<td>display questions</td>
<td>4.23</td>
<td>102</td>
<td>5.03</td>
</tr>
</tbody>
</table>

\[ F = 9.10, \ p < 0.05^a \]

\[ t = 3.92, \ df = 221, \ p < 0.0005 \] (one-tailed test)

\(^a\) The F-test tests for homogeneity of variances. Since the assumption of equal variances was untenable, the approximate method for use with unequal variances suggested by Cochran and Cox (1950) was used. This method adjusts the value of \( t \) required for significance.
### Table 5

Mean Length (in Words) of Learner Response to Referential Questions (by Class)

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>Number of responses</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1's class</td>
<td>4.45</td>
<td>11</td>
<td>5.62</td>
</tr>
<tr>
<td>T2's class</td>
<td>6.40</td>
<td>5</td>
<td>6.19</td>
</tr>
<tr>
<td>T3's class</td>
<td>13.65</td>
<td>55</td>
<td>20.73</td>
</tr>
<tr>
<td>T4's class</td>
<td>7.56</td>
<td>50</td>
<td>7.16</td>
</tr>
</tbody>
</table>

### Table 6

Mean Length (in Words) of Learner Response to Display Questions (by Class)

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>Number of responses</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1's class</td>
<td>2.61</td>
<td>52</td>
<td>3.55</td>
</tr>
<tr>
<td>T2's class</td>
<td>6.23</td>
<td>34</td>
<td>6.41</td>
</tr>
<tr>
<td>T3's class</td>
<td>6.00</td>
<td>5</td>
<td>4.42</td>
</tr>
<tr>
<td>T4's class</td>
<td>4.91</td>
<td>11</td>
<td>4.30</td>
</tr>
</tbody>
</table>
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 \( t=4.50, df=277, p < 0.0005 \).

<table>
<thead>
<tr>
<th>Learner response to referential questions</th>
<th>Mean c-units</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.19</td>
<td>177</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learner response to display questions</th>
<th>Mean c-units</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.56</td>
<td>102</td>
</tr>
</tbody>
</table>

\( F=1.83, p < 0.05^a \)

\( t=4.50, df=277, p < 0.0005 \)

\( \text{(one-tailed test)} \)

\( ^a \) The F-test tests for homogeneity of variances. Since the assumption of equal variances was untenable, the approximate method for use with unequal variances suggested by Cochran and Cox (1950) was used. This method adjusts the value of \( t \) required for significance.
Tables 8 and 9 show the mean number of s-nodes per c-unit in responses to display and referential questions by learners in each of the four groups. An examination of these tables suggests variability in these measures within the treatment and control groups.

Table 8

Mean Number of S-nodes per C-unit in Learner Responses to Referential Questions (by Class)

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>Number of C-units</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1's class</td>
<td>.23</td>
<td>53</td>
<td>.54</td>
</tr>
<tr>
<td>T2's class</td>
<td>.97</td>
<td>33</td>
<td>1.16</td>
</tr>
<tr>
<td>T3's class</td>
<td>1.00</td>
<td>5</td>
<td>.71</td>
</tr>
<tr>
<td>T4's class</td>
<td>.73</td>
<td>11</td>
<td>.65</td>
</tr>
</tbody>
</table>
Table 9
Mean Number of S-nodes per C-unit in Learner Responses to Display Questions (by Class)

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>Number of c-units</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1's class</td>
<td>0.23</td>
<td>53</td>
<td>0.54</td>
</tr>
<tr>
<td>T2's class</td>
<td>0.97</td>
<td>33</td>
<td>1.16</td>
</tr>
<tr>
<td>T3's class</td>
<td>1.00</td>
<td>5</td>
<td>0.71</td>
</tr>
<tr>
<td>T4's class</td>
<td>0.73</td>
<td>11</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Hypothesis (4): A greater number of referential questions is accompanied by a greater number of confirmation checks. The total number of confirmation checks made by the two control group teachers was 13; the total for the treatment group teachers was 21 (Table 10). Since the control group teachers took a total of only 141 turns, while the treatment group teachers took 244, or about 1.73 times as many 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. Since this was a slightly greater number of confirmation checks than the number made by the
treatment group teachers, Hypothesis (4) was not supported by the data. The difference between the two groups was not statistically significant ($\chi^2 = 0.005$, $df=1$, n.s.), using the correction for continuity for a one-way $\chi^2$ with one degree of freedom.

Table 10

Total Confirmation Checks in All Teacher Turns in Control and Treatment Groups

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted number of confirmation checks</th>
<th>Adjusted number of confirmation checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>13</td>
<td>22.49</td>
</tr>
<tr>
<td>$\text{(T1 and T2)}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment group</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>$\text{(T2 and T3)}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2 = 0.005$, $df=1$, n.s.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11 shows the frequencies of confirmation checks made by each of the four teachers in their turns immediately following learner responses to display and referential questions as well as the number made in their other turns. Only the two treatment group teachers made confirmation checks in turns other than those turns immediately following learner responses to display and to referential questions.
Table 11
Raw Frequencies of Confirmation Checks

<table>
<thead>
<tr>
<th>Following learner responses to display questions</th>
<th>Following learner responses to referential questions</th>
<th>In other turns</th>
<th>Total during lesson</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>T2</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T3</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>T4</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>11</td>
<td>9</td>
</tr>
</tbody>
</table>

Hypothesis (5): Confirmation checks occur more frequently following referential questions than following display questions. There were a total of 14 confirmation checks made by the teachers in their turns immediately following learner responses to display questions and 11 made in their turns immediately following responses to referential questions (Table 12). Since there were only 102 responses to display questions, but 121 to referential questions, or about 1.19 as many responses, 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. Hypothesis (5), therefore, was not supported by the data. The difference between the two groups was not statistically significant (\( \chi^2 = 0.78, \text{df}=1, \text{n.s.} \)), using the correction for continuity for a one-way \( \chi^2 \) with one degree of freedom.

Table 12

<table>
<thead>
<tr>
<th>Confirmation Checks Following Display and Referential Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted frequency</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Confirmation checks following learner responses to display questions</td>
</tr>
<tr>
<td>Confirmation checks following learner responses to referential questions</td>
</tr>
</tbody>
</table>

Hypothesis (6): A greater number of referential questions is accompanied by a greater number of clarification requests. The total number of clarification requests made by the two control group teachers was five (Table 13). and the total number of
clarification requests made by the two treatment group teachers was also five. 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. Since the adjusted frequency of 8.65 clarification requests made by the control group teachers was greater than the frequency for the treatment group teachers, Hypothesis (6) was not supported by the data. The difference between the two groups was not statistically significant ($\chi^2 = 0.52$, df=1, n.s.), using the correction for continuity for a one-way $\chi^2$ with one degree of freedom.

Table 13

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted frequency of clarification requests</th>
<th>Adjusted frequency of clarification requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>5</td>
<td>8.65</td>
</tr>
<tr>
<td>(T1 and T2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment group</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>(T3 and T4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\chi^2 = .52$, df=1, n.s.
Table 14 shows the frequencies of clarification requests made by the teachers in their turns immediately following display and referential questions, as well as the numbers made in their other turns. T3 did not make any clarification requests during the lesson, while frequencies for the other three teachers were low to a roughly equivalent extent.

Table 14
Raw Frequencies of Clarification Requests

<table>
<thead>
<tr>
<th></th>
<th>Following learner responses to display questions</th>
<th>Following learner responses to referential questions</th>
<th>In other turns</th>
<th>Total during lesson</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>T2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>T3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T4</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

Hypothesis (7): Clarification requests by the teacher occur more frequently following referential questions than following display questions.

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 perform a statistical analysis of the results.

Hypothesis (8): A greater number of referential questions is accompanied by a greater number of connectives in learner speech. Learners in the classes characterized by higher frequencies of referential questions (T3's and T4's classes) used a total of 71 connectives in all their turns during the lessons. Learners in the control group classes without the increase in referential questions used 11 (Table 15). However, the total number of turns taken by learners in the control group classes was 155, while the total number taken in the treatment group was 242, or about 1.56 as many turns. In fact, this difference was significant ($\chi^2 = 18.62, df=1, p < 0.001$), using the correction for continuity for a one-way $\chi^2$ with one degree of freedom. To adjust for the unequal number of turns taken by learners in the two groups, the number of connectives in the control group was weighted by a factor of 1.56. This adjustment yielded a total number of 16.5 connectives in learner speech in the control group. As predicted, the learners in the treatment group used a significantly greater number of
connectives ($\chi^2 = 32.72, \text{df}=1, p < 0.001$), using the correction for continuity for a one-way $\chi^2$ with one degree of freedom.

Table 15

Frequencies of Connectives in Learner Speech in Control Group and Treatment Group Classes

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted number of connectives used by learners</th>
<th>Adjusted number of connectives used by learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group classes (T1 and T2)</td>
<td>11</td>
<td>16.5</td>
</tr>
<tr>
<td>Treatment group classes (T3 and T4)</td>
<td>71</td>
<td>71</td>
</tr>
</tbody>
</table>

$\chi^2 = 18.60, \text{df}=1, p < 0.001$

Table 16 presents the distribution of connectives by class. While learners did not use any connectives during T1's lesson, learners in T3's class, where the frequency of connectives was highest, used almost twice as many as learners in the other treatment group.
teacher's class, where the second highest frequency occurred.

Table 16
Raw Frequencies of Connectives in Learner Speech

<table>
<thead>
<tr>
<th></th>
<th>Number of connectives used by learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1's class</td>
<td>0</td>
</tr>
<tr>
<td>T2's class</td>
<td>11</td>
</tr>
<tr>
<td>T3's class</td>
<td>46</td>
</tr>
<tr>
<td>T4's class</td>
<td>25</td>
</tr>
</tbody>
</table>

Although in the absence of videotapes of the lessons it was impossible to determine with complete certainty when a speaker took a series of successive turns in exchanges with the teacher, it was possible to tell with a reasonably high degree of certainty when such a series occurred. An examination of the transcripts gave the impression that there was a tendency in the treatment group classes for learners to have lengthy exchanges with the teacher. A count of the number of instances in all classes of learners taking more than two turns in succession showed that
this impression was correct (Table 17) and statistically significant (Table 18). The raw frequency of series of successive turns by learners in the control group was again weighted by a factor of 1.56 to adjust for the difference in the number of turns taken by learners in the treatment and control groups.

Table 17

<table>
<thead>
<tr>
<th></th>
<th>Number of instances of more than two successive turns taken by same NNS speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1's class</td>
<td>3</td>
</tr>
<tr>
<td>T2's class</td>
<td>6</td>
</tr>
<tr>
<td>T3's class</td>
<td>15</td>
</tr>
<tr>
<td>T4's class</td>
<td>23</td>
</tr>
</tbody>
</table>

51
Table 18

Frequency of Series of Successive Turns by Same NNS Speaker (By Group)

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Treatment group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted number of instances of more than two successive turns by the same NNS speaker</td>
<td>9</td>
<td>38</td>
</tr>
<tr>
<td>Adjusted number of instances of more than two successive turns by the same NNS speaker</td>
<td>14.04</td>
<td>38</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 10.12, \; df=1, \; p < 0.005 \]
The results reported above provide support for four of the eight hypothesized effects on ESL classroom discourse of an increase in the number of referential questions asked by the teacher.

As predicted by Hypothesis (1), the two teachers receiving the training session were able to increase the number of referential questions they used in the classroom. That this result was obtained after a single training session only twenty minutes long is not surprising in light of the fact that posing referential questions is not a new behavior for teachers. Indeed, referential questions are a common discourse phenomenon beyond the walls of the classroom, and the teachers' familiarity with them may have contributed to the ease with which their use in the classroom was increased.

The differences in the language produced by learners in response to the two question types were pronounced. First, with respect to quantity, learners' responses to referential questions were, on average, more than twice as long as their responses to display questions (Hypothesis 2). Second, responses to referential questions displayed more than twice the syntactic complexity of responses to display questions.
Finally, in the two classes in which referential questions were prevalent, learners used a far greater number of connectives to make explicit the links between the propositions they expressed (Hypothesis 8).

This last finding describes one of the overall differences characterizing the two classes in which the number of referential questions asked was greater since all turns taken by learners were analyzed for the number of connectives they contained. Another overall difference was in the significantly greater number of turns taken by learners in the two classes with the higher numbers of referential questions.

With respect to the expected alterations in the interaction between the teachers and the learners, the results were not as predicted. In fact, perhaps because of the generally high level of proficiency of the learners involved, the number of clarification requests made by the teachers was so small that statistical analysis of the difference in the numbers of those following responses to display questions and those following responses to referential questions was precluded (Hypothesis 7).

Cell frequencies of total clarification requests in the treatment and control groups were just high enough to perform a statistical analysis, but the
number of clarification requests made by the teachers asking predominantly display questions was slightly, although not significantly, higher than the number made by the teachers asking predominantly referential questions. This finding was counter to the prediction made in Hypothesis (6). Again, it seems likely that the generally high level of proficiency of the learners involved resulted in far fewer instances of unintelligible speech necessitating clarification than might have been the case with students of lower proficiency. With learners of lower proficiency the predicted patterns might have occurred.

While there were sufficient numbers of confirmation checks to permit statistical analyses, the results were not as predicted. The increase in referential questions in T3's and T4's classes was not accompanied by a greater total number of confirmation checks made (Hypothesis 4). The number of confirmation checks made by the two teachers asking predominantly display questions was greater, by an amount that was not statistically significant, than the number made by the teachers asking an increased number of referential questions. Furthermore, confirmation checks occurred immediately following learner responses to display questions slightly (but not significantly) more frequently than they occurred following learner
responses to referential questions (Hypothesis 5).

As with the number of clarification requests made, these findings are probably due to the learners' generally high level of proficiency. An example from the corpus serves to illustrate this point.\(^1\)

1 T: . Are there different ways of showing
deferece or respect?
2 S: An arrangement marriage.
3 T: How?
4 S: Arrangement marriage.
5 T: Arrangement.
6 S: Marriage.
7 T: Marriage?
8 S: Arrangement marriage maybe- kind of-
9 T: OK. An arranged marriage. Husband and wife.
10 S: Mhm.
11 T: Uh-huh.
12 S: But depend if the parents decide could marry.
13 T: OK.
14 S: And you did.
15 T: Yeah it depends on what they said.
16 S: Yeah ( ) if the father said ( 
17 ) then they're out of luck. ((laughter))
18 Right now it's different but.
19 T: Now it's changing?
20 S: It's changing but.

56
The teacher seems to use the confirmation check in line 20 more as a means of prompting the learner to continue speaking than of checking his understanding of the meaning of what the learner has said. The confirmation check in line 8, on the other hand, seems to be part of an attempt by the teacher to understand the learner's statements in lines 3 and 5. If the proficiency level had been lower, the necessity might have arisen more frequently for the kind of meaning negotiation of which the confirmation check in line 8 appears to be part. If, in addition, it is true (as it seems impressionistically to be) that confirmation checks were often used as prompts by teachers even when they understood what a learner had said, then, given the learners' high level of proficiency, it is not surprising that in this study the expected pattern of confirmation checks did not occur.

The example above also serves to illustrate a particular problem with this study brought about by the use of the communication unit as a unit of analysis. Although the response in line 3 above is contextually appropriate in the sense Widdowson (1978) uses the term, because it lacks more than repetition of elements in the teacher's question to be an independent
grammatical predication, it is not a c-unit as defined in this study. As a consequence, the response in line 3 above and the others in the corpus like it were left out of the analysis of syntactic complexity. Fortunately, responses that had to be excluded for this reason were few enough in number to avoid seriously reducing the amount of the corpus analyzed for the measure of syntactic complexity. For example, the results for the syntactic complexity analysis for T4's class, which had the greatest number of responses (seven to referential questions) excluded, were altered an insignificant \((t=0.71, \text{n.s.})\) amount when a mean was calculated including those responses as if they were c-units. However, it is conceivable that in other ESL classroom situations—with lower levels of proficiency or greater numbers of students attempting to speak—the problem might be more serious.

An interesting issue in the interpretation of the results reported for the measures of syntactic complexity and length of response is the degree to which these two variables are independent. It is possible that longer responses imply greater syntactic complexity, or vice versa. Although it is impossible to draw any conclusions on the basis of data drawn from only two teachers, it is interesting nevertheless to compare the results from T2's (control group) lesson.
### Table 19
Comparison of Mean Length of Response to T2's Display and to T4's Referential Questions

<table>
<thead>
<tr>
<th></th>
<th>Mean length of response (in words)</th>
<th>Number of responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses to T2's display questions</td>
<td>6.23</td>
<td>34</td>
</tr>
<tr>
<td>Responses to T4's referential questions</td>
<td>7.56</td>
<td>50</td>
</tr>
</tbody>
</table>

$t = 1.33$, $df = 82$, n.s.

### Table 20
Comparison of the Mean Syntactic Complexity of Responses to T2's Display and to T4's Referential Questions

<table>
<thead>
<tr>
<th></th>
<th>Mean number of S-nodes</th>
<th>Number of per c-unit c-units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses to T2's display questions</td>
<td>.97</td>
<td>33</td>
</tr>
<tr>
<td>Responses to T4's referential questions</td>
<td>1.08</td>
<td>59</td>
</tr>
</tbody>
</table>
and T4's (treatment group) lesson. Table 19 shows a comparison of the mean lengths of response to referential questions in T4's class and to display questions in T2's class. The difference between the two is not significant (t=1.33, df=82, n.s.).

A similar comparison of the results of the analysis of syntactic complexity (Table 20) suggests that where the length of response to the two question types is held relatively constant, there may not be a significant difference in syntactic complexity. It may be that because responses to referential questions tend to be longer, they also tend to be more syntactically complex. It is possible that there is no effect of referential questions separate from that brought about by the increase in the length of responses to them.

Because the focus of this study was on the effects of display and referential questions on classroom discourse, teacher wait-times were not calculated. However, this is a limiting factor in the interpretation of the results reported insofar as it is not possible to determine to what extent an increase in teacher wait-time may or may not have been correlated with the longer and more complex student responses to referential questions. It may be the case that teachers tend to wait longer after asking referential
questions and that this added pause is a factor in the longer and more complex responses observed. On the other hand, it may be the case that referential questions elicit longer, more complex responses despite the fact that the wait-time after them is roughly the same as the wait-time after display questions. Since wait-times were not calculated for the purposes of this study, though, it is impossible to address this question.

A second possible limitation of the study is the fact that the Cohen's kappa coefficient for interrater reliability was only 0.76. Frick and Semmel note that the more uneven the distribution of category frequencies is, the greater the effect chance correction has on simple percentage agreement, which in the present study was 0.91. Since the distribution of category frequencies was uneven throughout the corpus, including the sample coded for the purposes of the measure of interrater reliability, this may have contributed to the magnitude of the effect of chance correction. However, Frick and Semmel assume that simple percentage agreement of 0.85 to 0.90 or greater corresponds to "nearly perfect agreement" (1978:175) and report that if simple percentage agreement is 0.90, then Cohen's kappa "from the experience of the present authors, typically falls around 0.80 or greater" (1978:175).
Perhaps the most serious limitation of this study, however, was the very small number of teachers involved. Thus, while the results reported support the hypotheses stated above, verification with larger number of teachers is necessary to establish confidence in these findings. In addition, it will be necessary to investigate what effects group size and proficiency level might have on the results obtained. Finally, it will be necessary to determine to what extent the effects of training persist and produce permanent alterations in teachers' questioning patterns.

With verification, however, these results have clear implications for the ESL classroom. With what seems to be a relatively easy alteration for a teacher to make, he or she can effect substantial changes in the amount and kind of practice ESL students obtain in the classroom. In responding to referential rather than to display questions, learners appear to speak at greater length using more syntactically complex speech. It also appears that learners in classes where greater numbers of referential questions are posed also tend to take more speaking turns.

That referential questions may produce an increase in the amount of speaking learners do in the classroom is relevant to at least one current view of second language acquisition (SLA). In a paper presenting the
results of a study of the acquisition of French by Canadian children in elementary school immersion classrooms, Swain (1983) 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 extra-linguistic information, "it is possible to comprehend input--to get the message" (Swain 1983:19) 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" (1983:20).

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

In addition to speaking more in classes characterized by higher frequencies of referential questions, it seems that learners also use a far greater number of connectives. Since connectives are
those global elements (Burt and Kiparsky 1975) 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 at least, mistakes in the use of connectives linking clauses within sentences were more likely to cause breakdowns in communication than mistakes in the use of articles. If as seems likely, global elements like connectives are equally important to oral communication, then it may be useful to know that posing referential questions provides increased practice in their use.

Finally, responding to referential questions is practice in language use, in Widdowson's (1978) sense. As Widdowson observes,

A knowledge of use must of necessity include a knowledge of usage but the reverse is not the case: it is possible for someone to have learned a large number of sentence patterns and a large number of words which can fit into them without knowing how they are actually put to communicative use. (Widdowson 1978:18-19)

Since providing an interlocutor with information he or she does not already have is fundamentally communicative language use, referential questions can
serve to initiate practice in communicative language use in the ESL classroom.
Who in any major city of the United States has not met at least one person with a Chinese-Malayan face, Spanish surname, and American nickname? This person, a Filipino, could easily be mistaken for a Japanese, Chinese, Korean, or Vietnamese national. In the decade ending in 1974, a total of 120,269 Filipinos migrated to the United States, the annual numbers rising from 3,130 in 1965 to 32,857 in 1974; while the annual numbers of temporary workers, trainees, and exchange visitors increased from 2,691 to 32,203.

Before World War II, only a handful of Filipinos came to the United States, mostly laborers brought over for the big plantations in Hawaii and the farms of California and the Pacific Northwest. Recently, more professionals have migrated to America, usually bringing with them their immediate families. Modern U.S. technology, seen as a means of improving professional competence and enhancing socioeconomic status, is one of the reasons for this migration.

Essentially, the Filipino is an Asian with a strong Malayan base, derived from the highly complex
culture of the Indonesian Madjapanhit Empire.

In the sixteenth century, the West started influencing the Philippines. The 300 years or so of Spanish hegemony began when Ferdinand Magellan first claimed the Philippine Islands for Spain in 1521. Then, as an outcome of the Spanish-American War, the Philippines was ceded to America for $20 million, a sale consummated without the knowledge or consent of the Filipino people.

Another country that has exerted tremendous influence on Filipino thought and behavior is China, which traded with the Filipinos from the Tang dynasty in the seventh century.

Filipino Culture

Although the Filipino culture has arisen from such a mixed heritage, there are some basic traits that most Filipinos will manifest. Individualization is necessary, for of course, no one person will demonstrate all the elements of his culture.

The Philippines consists of roughly 7,000 islands, big and small. Filipinos display characteristics that differ, depending on the region they hail from. Due to the Chinese influence on family solidarity, people from the same region tend to socialize only among themselves. This clannishness is evident among the
many organizations in Filipino communities, which to some observers might appear to be rivalry. Thus, lack of spontaneous cooperation is not surprising, but it is both a weakness and a strength.

If at all possible, the nurse should identify her Filipino patients' regional idiosyncrasies. She can probably obtain help in this from a Filipino worker in the hospital.

The younger generation's values in some ways clash with the traditional. In the Philippines today, youths resemble their western counterparts, particularly in matters of dress, music and social values. In the U.S., one sees this orientation, whether they were brought up by parents who had migrated or they are third-generation American children. Despite this, one may still observe in them such traditional values as respect for elders, love of family, and preservation of self-esteem.

The strong feeling for family, a quality derived from the Chinese influence, is manifested by old-fashioned patterns imposed by the family patriarch or the equally authoritative matriarch. Respect and deference are always given to one's elders, whose words and decisions one dares not question. The young receive solicitous protection from their elders. In the absence of both parents, the eldest child has the
say and must be obeyed.

The implications for health care are important. Filipino patients always have their families hovering over them, perhaps to the irritation of the nursing staff. The sick Filipino child feels lost without his mother constantly at his bedside. When grandparents are ill, sons, daughters, and even grandchildren take turns keeping them company and supporting their husbands or wives.

A daughter newly delivered follows traditional customs related to activity, food, and hygiene, which may be contrary to what the doctor or nurse prescribes.

Single adults with no relatives in the vicinity have swarms of visitors. These are concerned friends, who recognize the loneliness that illness can bring. Nurses would do well to capitalize on this custom for whatever assistance it provides the patient.

Certainly, family-centered nursing is indicated.

Filipinos are deeply religious and God-fearing. They have a deep sense of destiny, a heritage from the Asian religions of pre-Spanish times, coupled with their firmly rooted faith in the God whom the Spanish missionaries brought to the islands, Spain's greatest contribution to the collective Filipino soul.

An expression often used, whose origin Filipinos themselves may not know, is Bahala na. It is a
corruption of *Bahala*, the name of the ancient god of the pagan Filipinos. In uttering *Bahala na*, Filipinos are saying that they are leaving things in the hands of God.

Another expression they may use is *talaga*, meaning "destined, inevitable." Ill Filipinos tend to attribute their condition to the will of God and to cope with their illness by praying and hoping that whatever God's will is, it is the best for them. Although both patients and families find it hard to accept a poor prognosis, they keep on hoping despite their resignation. This attitude explains why Filipino patients are uncomplaining and frequently suffer in silence.

The nurse can try to penetrate the facade of cheerfulness, silence or fatalism, including the family in her explorations. Filipino patients and their families appreciate openness on the part of their doctors and nurses, for their reliance on God and fate strengthens them and enables them to cope.

Temperament

Filipinos, being essentially Asian, have gentle, mild, and passive temperaments--the temper of people whose collective unconscious has been anchored in the idea of a harmonious balance between man and nature.
They generally are neither assertive nor aggressive. They appear guarded or reticent. This, unfortunately, can be misconstrued as an inferiority complex. Too often, they are labeled as passive-aggressive, as having a personality disorder with anger as its underlying cause. This anger produces anxiety, which is usually handled through covert and passive means. Behavioral manifestations of this disorder and behavior common to Filipinos are procrastination and stubbornness, associated with intense cravings for acceptance, generosity, and verbal demands for attention. Whether this behavior in Filipinos is a personality disorder is difficult to determine. I would venture to say that this attitude stems from centuries of Spanish repression, which inhibited a proud people. As a result of more than three centuries of Spanish hegemony, Filipinos learned the value of silence and prudence. To preserve security, they are submissive to authority and reluctant to express their opinions.

No wonder they regard doctors and nurses as authority figures and do not question whatever regimen is imposed on them. They easily win the labels "good patient," "cooperative," and "uncomplaining." Yet, is this how one should perceive such behavior? Shouldn't the nurse use assessment skills to determine what is
behind this behavior?

Interpersonal Relationships

Filipinos generally behave agreeably, even to the extent of personal inconvenience. This is called pakikisama, which means going along with others.

Related to pakikisama is hiya or "shame," which is Asian. The Chinese and the Japanese call it "face," and the Spanish orgullo, or "self-pride." Hiya is a painful emotion arising from a relationship with an authority or society that inhibits self-assertion in a situation perceived as dangerous to one's ego.

Another trait is amor propio, which is Spanish and means "self-esteem." When a Filipino's amor propio is wounded, he preserves his dignity through silence or aloofness, believing that to do otherwise would demonstrate a lack of self-pride. Filipinos believe that having accepted the doctors or nurses implies that they have confidence and, by the same token, they trust that it is mutual.

On the other hand, even when they mistrust the doctor or nurse, Filipino patients seldom tell them to their faces. They beat around the bush for fear of hurting the other person's feelings or, perhaps, have a family member or friend make known their feelings.

Commonly, a hesitant yes means no. This is an
effort to avoid a direct, blunt no. Still another custom is the use of a euphemism, such as kuwan, meaning "thing." It can replace any expression one does not ordinarily use in polite company, and its nonverbal counterpart is a nod, which means yes.

The Filipino language has hierarchical terms for yes and no. The term depends on whether a Filipino is speaking to a person of lower, equal, or superior status. When a Filipino does not know another's status, a silent nod avoids giving possible offense. Filipino patients often address a physician or nurse properly as "Dr. Jones" or "Ms. Smith." If they don't know their names, they nod.

A silent nod may also be a defense in a belittling situation. If an authority figure gives instructions in a belittling way, a Filipino patient may nod. If a nurse rattles off instructions and the patient nods, it could mean he understands, or it could mean that her instructions are inadequate but he wants to spare her feeling and preserve his amor propio.

Food Preferences

Eating habits vary with Filipinos' region of origin. Drawing origin from various cultures but displaying regional characteristics, Filipino food was prepared by the series of Malay migrations, spiced by
commercial relations with Chinese traders, stewed in 300 years of Spanish rule and hamburgered by American influence on the Philippine way of life.

The main influence in this potpourri is Spanish cookery. Certainly, to Europeans and South Americans, nothing can be very strange about Filipino dishes.

The French will recognize a lighter bouillabaisse in our sinigang, which is fish in a tart broth with tomatoes and tamarind (a fruit noted for its tartness, native to the Philippines). The Germans will find arroz con caldo like their Suppenham. Both are rice in chicken broth.

Despite regional differences, there are national dishes known all over the Philippines. Among these are adobo (pork, chicken, beef, or a combination simmered in a vinegar and garlic sauce) and dinuguan (pork flesh and innards spiced with whole peppers and stewed). Another is pancit (long, uncut rice noodles sauteed with meat and vegetables). Still another is lumpia (a roll of vegetables and meat in a paper-thin rice wrapper). Filipinos like lechon (whole pig roasted outdoors for long hours over charcoal). Some Americans, no doubt, have been treated to some of these foods by their Filipino friends.

The Filipino staple is rice, ordinarily boiled to fluffiness. It is eaten at every meal. For breakfast,
rice usually is fried with a touch of garlic and eaten with an egg, sausage, or fried fish. Rice is the main bulk of Filipino meals, a must on the dinner table.

Attractively served western dishes may not suit Filipino patients, even when they are familiar with American food which lacks spices. Perhaps the hospitalized Filipinos may eat the first few meals, then start craving rice and home-cooked dishes. A Filipino child might ignore other foods but pick at ice cream and desserts. Food from home can help, but the nurse needs to tell the family that they can bring food.

Filipinos use salt generously. On top of the already salted and spiced dishes, Filipinos pour a salty, brownish clear liquid called patis, a preparation of fish or shrimp extract. Some use soy sauce. Another sauce used on shrimp and other dishes, while cooking or at the table is bagoong, which is highly salted. In fact, a humble meal can be made with a cup of rice doused with bagoong, plus tomatoes or any leafy vegetable. For Filipino patients who must limit their salt intake, a careful examination of food habits is essential.
Appendix B

Vocabulary Items from the Reading Text and Sample Referential Questions

**competence**

What are you most competent in?

**enhancing**

What do you think the best way is for a person to enhance his or her position in society?

**hegemony**

What effects do you think three centuries of hegemony by a foreign power would have on a local culture?

**heritage**

What do you think some of the advantages are of having a mixed heritage?

**manifest**

What are some basic traits you think most Americans manifest?

**patriarch**

What are the characteristics of a patriarch in your culture?

**matriarch**

What are the characteristics of a matriarch in
your culture?

defere

How does one display deference in your culture?

prognosis

Do you think a doctor should tell a patient that his or her prognosis is poor?

penetrate

Do you think a nurse should try to penetrate a facade of cheerfulness in a patient?

What do you think are good ways to cope with stress?

harmonious

What does the idea of "a harmonious balance between man and nature" mean to you?

anxiety

How do you think anger produces anxiety?

procrastination

What is something you procrastinate about?

cravings

What foods do you get cravings for?
Do you think being in a foreign culture makes you more prudent?

Do Americans seem inhibited to you?

Why do you think we have euphemisms?

Are there ways of talking to people in your culture that reflect a hierarchy?

What way would you consider to be a belittling way to give instructions?

What is something you would feel reticent to do?
NOTE

Transcribing conventions used in this study were as follows: a question mark indicates rising intonation; a period indicates falling intonation; a hyphen indicates interruption, either self-interruption or interruption by another speaker; single parentheses enclosing blank spaces indicate the presence of words the transcriber could not reliably transcribe; double parentheses enclose the transcriber's comments providing extra-linguistic information.


van Eijk, Teun. 1977b. Semantic macro-structures and knowledge frames in discourse comprehension. In


Rogers, Virginia, and O. L. Davis. 1970. Varying the cognitive levels of classroom questions: an analysis of student teachers' questions and pupil achievement in elementary school social studies.


