INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Davis, Brian Kenneth

A STUDY OF THE EFFECTIVENESS OF TRAINING FOR FOREIGN TEACHING ASSISTANTS

The Ohio State University

Ph.D. 1984

University Microfilms International

Copyright 1984 by Davis, Brian Kenneth All Rights Reserved
A STUDY OF THE EFFECTIVENESS OF TRAINING FOR
FOREIGN TEACHING ASSISTANTS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

* * * *

The Ohio State University
1984

Reading Committee:
Frederick R. Cyphert
James W. Altschuld
John P. Humeke

Approved by
Frederick R. Cyphert
Adviser
Educational Theory and Practice
For
Lillie

Multum in parvo
ACKNOWLEDGEMENTS

This study would not have been possible without the assistance of those listed below. For their various contributions I am most grateful.

Jim Altschuld
Yun-Shiow Chen
Yun Chan Chi
Wen-Chi Chou
Ted Cyphert
Joe Damico
Bill Holschuh
Jason Hsu
Phil Buneke
Tian-Tzer Jeng
Leonard Jossem
Song Hyo Kim
Jules Lapidus
Lillie Lum
Gerry Martin
Paul Pelloquin
Debbie Phillips
Jagdish Rustagi
Roger Sell
Rick Sentieri
Kyung Yul Shin
Nien-Ling Tang
Chao-Ping Ting
Doug Wolfe
Wylbur
Sungchil Yeo
VITA

October, 1942............ Born: Leicester, United Kingdom

1965........ Dip. Ed., University of Newcastle-upon Tyne

1974...... B.Ed., University of Victoria, British Columbia

1978...... M.A., University of Victoria, British Columbia

1978-84...... Graduate Associate, The Ohio State University

PUBLICATIONS

ERIC Document Reproduction Service No. ED 222 447.

ERIC Document Reproduction Service No. ED 222 446.

Effective Lecturing: A Question of Clarity? Proceedings of 
The American Society for Engineering Education, The Ohio 
State University, Columbus, Ohio, April 13-14, 1984. 

FIELDS OF STUDY

Major Field: Teacher Education

Moral Education: Professor Kevin A. Ryan

Inservice Education: Professor Frederick W. Cyphert

Human Relations: Professor Susan J. Sears
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>The Role of the TA</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>TA Training</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Research Questions</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Definition of Terms</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Delimitations</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Limitations</td>
<td>21</td>
</tr>
<tr>
<td>II.</td>
<td>REVIEW OF THE LITERATURE</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Purpose</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Intervention Strategies</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Outcome Measures</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Literature Related to Training</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Program Content</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Foreign Teaching Assistant</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>Conclusion to the Review of the Literature</td>
<td>126</td>
</tr>
<tr>
<td>III</td>
<td>RESEARCH PROCEDURES</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Overview</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Sample</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Design</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>Instrumentation</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>Data Analysis Procedures</td>
<td>155</td>
</tr>
</tbody>
</table>
IV RESULTS AND CONCLUSIONS

Introduction...........................................173
Phase I: The Observation Instrument.................173
Phase II: Training Program Data......................175
Phase III: Principal Analyses.......................176

V DISCUSSION, SUMMARY, AND RECOMMENDATIONS
FOR FURTHER RESEARCH

Introduction...........................................201
Major Research Question............................201
The Observation Instrument........................205
TA Assimilated Knowledge and
Undergraduate Achievement.........................205
TA Use of Behavior and
Undergraduate Achievement.........................206
TA English Ability, Behavior
and Undergraduate Achievement.....................212
Different Measures of TA
English Ability.......................................215
Summary of Discussion of Results...................216
Recommendations for Further Research.............219
Questions Raised and Recommendations.............221
Concluding Remarks...................................224

APPENDICES

A. Report of Observed Teacher Behaviors.........228

B. Report of Observed Teacher Behaviors:
Behavioral Definitions for Items..................233

C. Report of Observed Teacher Behaviors........239

D. Report of Observed Teacher Behaviors:
Observers' Instrument..............................244

BIBLIOGRAPHY..........................................248
LIST OF TABLES

Table                                           Page

1. Change in TA Behavior(s) After Training as Measured by Non-Student Ratings............. 60
2. Change in TA Behavior as Measured by Student Ratings............................................. 63
3. College Level Studies Addressing Specific Constructs............................................. 108
4. Phases of the Study............................................. 133
5. Population From Which Sample was Drawn for Each Phase of the Study......................... 134
6. Factors Involved in the Design of the Study in Terms of TA Equivalent Pairs With Undergraduate Achievement Data............................................. 178
7. Phases of the Study, Purpose, and Analyses..... 160
8. Chi-Square and T-test Analyses for TA Use of Behaviors............................................. 166
9. T-test for Undergraduate Achievement and TA Use of Behaviors............................................. 167
10. ANOVA for Use of Behaviors and the English Ability of Trained and Untrained TAs........ 168
11. ANOVA for Undergraduate Achievement and the English Ability of Trained and Untrained TAs... 169
12. Undergraduate Achievement Scores and TA Assimilated Knowledge on Training Program Examination............................................. 180
13. Behaviors Used Significantly More Often by Treatment TAs (n=8) Than Control TAs (n=8).... 184
14. Grand Means Per Observation of the Ten Behaviors Measured Using a Frequency Scale..........................................................185

15. High and Low Use of Behaviors by TAs and Undergraduate Achievement.................................188

16. Undergraduate Achievement and TA Spoken English Level by All TAs, Treatment TAs, and Control...193
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Representation of the Relationship Between Use of Behavior 2 and Spoken English Level for Trained and Untrained TAs</td>
<td>191</td>
</tr>
<tr>
<td>2. Representation of the Relationship Between Use of Behavior 7 and Spoken English Level for Trained and Untrained TAs</td>
<td>192</td>
</tr>
<tr>
<td>3. Representation of the Relationship Between Use of Behavior 17 and Spoken English Level for Trained and Untrained TAs</td>
<td>193</td>
</tr>
</tbody>
</table>
Chapter I

INTRODUCTION

Many studies concerned with the preparation of Teaching Assistants (TAs) reference complaints or problems related to TAs' teaching ability (Alciatore & Eckert, 1972; Armenti & Wheeler, 1978; Bailey, 1977; Barrus, Armstrong, Renfrew, & Garrard, 1974; Brinton & Gaskill, 1979; Bray & Howard, 1980; Buckenmyer, 1972; Carroll, 1977, 1980; Costin, 1968 (a), (b); Damarin & West, 1979; Deboer, 1979; Franck & DeSousa, 1982; Golman, 1975; Haber, 1973; McGahie & Mathis, 1977; Monson, 1970; Schade & Bartholomey, 1980; Schulz, 1980; Staton-Spicer & Nyquist, 1979). It is not uncommon for such studies to quote from local or college newspapers.

A survey of about 100 students in various colleges conducted last quarter showed the two major problems were difficulty in understanding foreign TAs and the lack of TAs preparation for teaching a class. Because of these problems and the conditions of the state budget, which will make students pay higher tuition, Martin said that students deserve a better education (The Lantern, 1982(a), page 1).

There are a number of important reasons for addressing concerns about TA teaching effectiveness. Complaints from undergraduates about inadequate instruction by TAs is prob-
ably the most pressing argument for improving TA instruction especially when complaints are considered along with student and institutional financial concerns. This is not to imply that the only reasons for the improvement of teaching are political and financial. The improvement of instruction needs no justification simply on these grounds.

All beginning and intermediate language courses at the University of Wisconsin, Madison were reportedly taught by TAs and this figure rises to over ninety percent at the University of Michigan, Davis (Franck & Samaniego, 1981). While such figures are not currently available at The Ohio State University, a conservative estimate of the investment by that institution may be made. There are a total of 1,974 TAs currently employed of which nearly one quarter are foreign TAs. Over ninety percent of the foreign TAs are from countries where English is not the first language. Their percentage fluctuates from department to department. For Fall Quarter of 1983 in the College of Mathematical and Physical Sciences the percentage of foreign TAs was approximately thirty-three. Eighty percent of these were from Asian countries (Office of the Vice President for Personnel Service, O.S.U., 1984). In the departments involved in this study, Mathematics and Statistics, the percentage of foreign TAs exceed sixty.
The total number of TAs employed (1,974) permits a conservative estimate of dollar investment to be close to twenty million dollars per year with the inclusion of tuition waivers in the calculation. It might appear that value for money is not only the right of the undergraduate but also the employing institution.

The adequacy of preparation of TAs may impact upon the institution in ways other than that immediately associated with finances. TAs are responsible for a considerable part of the instruction received by undergraduates. They usually teach undergraduate courses which are traditionally the undergraduates' first contact with the department. Thus the TAs often represent the discipline and must do so well if the undergraduates are to be attracted to majoring in the department.

In recent years, a significant source of complaint by students taking introductory physics courses has been the quality of instruction in the laboratories. A large number of such complaints involve the language barrier between the TA and the student (in many cases the majority of TA's working in introductory physics courses are foreign born). In addition, students often report apathy, unpreparedness, and occasionally incompetence among TA's who otherwise spoke and understood English perfectly. It was recognized that widespread reports of unrewarding experiences in undergraduate labs might be dissuading potential students from taking what otherwise might be considered good courses (Armenti & Wheeler, 1978, page 122).
Courses taught by TAs are often service courses for other departments, beginning calculus for instance. Unless these courses are taught effectively requirements for progress in those departments may not be met (Staton-Spicer & Nyquist, 1979).

Additionally, if college and university enrollments decline as predicted by 1986 (Frankel, 1978), competition among departments will increase and effective teaching of undergraduate courses will become critical in the competition for these undergraduate students (Armenti & Wheeler, 1978). The employment of TAs primarily allows the provision of instruction in lower-division fundamental courses that maximize undergraduate student enrollment. At the same time it provides support for doctoral students working as TAs. Thus, it would appear that the teaching assistantship is presently designed for bringing more undergraduates into the department. The TA is repaying the institution for financial support, at the same time enabling the institution to increase student enrollment without hiring more faculty. Tave (1978) actually claims that graduates have been admitted to TA positions simply to fill the positions even though not suited to graduate studies. It is clear that the TA position is not designed for the purpose of providing graduate students the opportunity to develop skill in teaching nor as preparation for future employment (Sayer, 1978).
Finally, the problems resulting from inadequate instruction by TAs do not reduce only to short term concerns such as enrollment problems, student complaints, and cost-effectiveness. A long term consideration that has impact throughout the educational system has been addressed by some writers (Sayer, 1978; Staton-Spicer & Nyquist, 1979). Many graduates of higher education seek academic careers in higher education. In a sense then, graduate education can be seen as a training ground for future college and university teachers. It has been concluded that graduates most likely to succeed in the competitive market place will possess demonstrated teaching ability as well as proven research ability (Staton-Spicer & Nyquist, 1979).

No major university involved in large-scale teaching, research and community service activities could support these endeavors without employing large numbers of graduate teaching associates and assistants. Most of these graduate apprentices will someday staff the classrooms and laboratories, the work benches and podiums and the auditoriums in which they now assist senior faculty members. These graduate teaching associates, in addition to their teaching and research functions, are also serious graduate students working towards higher degrees in their fields of specialization (Haenicke, 1983).

It has been argued elsewhere that all graduate education should include a teacher training component (Bailey, 1977). At least it would seem appropriate to provide such training for TAs especially if the experience of teaching at the college level influences career choices in this direction.
Unlike the graduates in the "flush" academic marketplace of the early 1960s, the current graduates have found that a doctorate no longer guarantees a teaching position within higher education. Enrollments have stabilized and expansion has slowed; our graduates have flooded higher education; there are no longer sufficient numbers of jobs to accommodate this demand. As a direct result of this situation, an increasing percentage of recent Ph.D graduates has turned to the secondary schools and junior colleges to seek employment (Sayer, 1978, page 213).

The Role of the TA

Each year thousands of students enter the universities of the United States not only to take courses as students but to work in those institutions as teaching assistants (TAs). White (1977, 1978) found from enrollment data gathered from member institutions of the Council of Graduate Schools in the United States that there were 109,048 graduate teaching assistants (GTAs) in 285 institutions and that across disciplines there was a general trend for such appointments to increase. In large colleges and universities it is not unusual for an undergraduate student at some time to be taught by a TA.

Such teaching assistants are usually students working on a masters or doctoral degree in the discipline in which they teach. Assistantships are a major form of financial support for graduate students and are usually awarded on
the basis of academic potential. A teaching position may be given as a result of lack of funding for a research position or simply because of a need for someone to assume responsibility for laboratory and recitation classes. The recruitment of TAs has even been extended in recent years to include undergraduates (Arbes & Kitchener, 1974; Boeding & Vattano, 1976; Carroll, 1977; Crawford, 1983; Kanaga, 1979; Borbaan & Hillis, 1976).

The responsibility of a TA can vary considerably (Nowlis, Clark, & Rock, 1968; Rose, 1972). At one end of a continuum there may be no contact with undergraduate students resulting from a role involving clerical work; grading material and preparing material for the instructor's use. Responsibility may involve the laboratory setting where it can vary from maintenance of equipment, through grading reports, to supervising experiments and giving instruction. Many courses in which a member of the faculty lectures may have in addition a number of recitation groups during which the assistant answers students' questions, reviews lecture material and may instruct the group. The responsibility may include setting and grading quizzes and the provision of individual tutorial sessions for which an office area is provided. At the extreme end of the continuum the assistant may have more or less complete responsibility for all aspects of the teaching (Staton-Spicer and Nyquist, 1979).

TA Training
Most TAs have little or no formal teaching experience and face their initial teaching experience with trepidation (Krockover, 1980). Chase (1970) reported that TAs want better preparation, supervision, and evaluation, but usually do not know where to direct or how to express their requests to invite faculty to become more involved in this part of their graduate training. Most new TAs have had little if any prior contact with undergraduates in the instructional setting, have not been trained in teaching techniques, are unprepared in test preparation or administration, and have no experience in evaluating anyone using letter grades (Krockover, 1980).

A strong background in the subject matter that they will teach must be a necessary but not sufficient condition for effective teaching. Allen states that few have received adequate—if any—preparation on methods of teaching. They are often left to themselves to discover the best procedures for teaching and at best get an awkward start in their teaching, with slow improvement and inadequately developed competence (Allen, 1976).

In virtually no other job would a person be expected to perform well without training. Teachers of elementary and secondary students must complete a thorough sequence of courses in educational methods and theory before they are licensed to teach. Yet GTAs seem to be assigned to classes with no experience and little training (Yoder & Hugenberg, 1980, page 14).
The training of TAs for their teaching role has received increased attention in the last few years (Sharp, 1981). The provision of such programs of preparation is, however, not necessarily widespread. Jossen (1982) surveyed 102 physics departments listed in the 1980-81 edition of the AIP publication "Graduate Programs in Physics, Astronomy and Related Fields" as employing ten or more graduate teaching assistants. Of the 91 responding 60% had no formal teaching preparation and only 17% reported programs designed to improve teaching which lasted a week or more. Such programs were described as being voluntary and not well attended.

Ervin & Huyskens (1982), in a review of the literature relating to the preparation of TAs in foreign language departments, found a "substantial improvement" comparatively in the decade since 1970. The figure for departments requiring a methods course for TAs had risen from twenty-five percent in a 1970 to thirty-eight percent in a 1979 survey. These modest figures certainly reflect an improvement on those of over 15 years ago when it was reported that less than ten percent of the departments in colleges and universities providing programs for TAs gave systematic instruction in how to plan a course, methods of classroom teaching, or evaluation of student performance (Costin, 1968).
While acknowledging that there is not unanimous support for the training of TAs (Bailey, 1977; Yoder & Hugenberg, 1980) the implementation of such programs provides a possible solution to the problem of dissatisfaction with the instruction provided undergraduate students. Such programs in their varied forms have received much attention in the literature over the past decade. Carroll (1980) in a review of the empirical research on TA training programs reports forty-eight descriptive accounts of various programs "none (of which) reported actual data or statistics on the effects of training" (page 168). In summarizing the results of twenty-four studies which did assess effects he points out that only eleven of these included a control group in their design. Carroll (1980) recommends that "substantially more effort ought to be devoted to assessing the effects of TA training programs rather than simply describing innovative ways of conducting such programs," and that studies "of matched or equivalent groups of participants and non-participants would be most useful" (page 176).

Levinson-Rose and Manges (1981) in a similar review of research focussed on intervention strategies in training programs suggested that:

A well-defined field of inquiry should draw upon coherent theory, subscribe to high standards of research, and build upon previous research in a systematic way. By these criteria, research on
improving college teaching is not a well-defined field. For most studies, the basis in theory is strained and for some it is nonexistent. Work on major conceptual issues remains, for example, before programs can be validated, we must clarify the nature of "instruction" and the meaning of "improvement" (Levinson-Rose & Menges, 1981, page 418).

The intervention strategy selected for this present study is one based upon the concept of modelling. It incorporates microteaching techniques and the use of protocol videotapes of TAs teaching (for details, see Review of the Literature).

There are numerous descriptions of innovative TA training programs and instructional interventions in the literature but only few have investigated the effectiveness of either the programs or various program components (Carroll, 1978; Sharp, 1981). In many cases published descriptions of training programs fail to provide evaluation data (Allen, 1976). Carroll (1977), noting the lack of empirical research on the effects of TA training programs, identified only four studies that: included student variables among criteria of effectiveness; reported reliable data to assess change in teaching behaviors; utilized an experimental design. All four focusses on a few highly specific teaching competencies. In a review of empirical research on TA training programs Carroll (1980) noted the limited extent
of such research, at which point six studies were identified meeting the three criteria specified above.

Today the only evidence available which suggests that the program is fruitful is testimony from participants. In order to demonstrate that such preparation actually results in improved teaching, it will be necessary to define criteria of teaching effectiveness and secure a body of empirical data for program assessment (McGahie & Mathis, 1977).

This present study focuses upon a teaching program whose goal was the acquisition by the participants of certain lower-inference teacher behaviors which relate to and facilitate the clear communication of the subject matter of instruction to undergraduate students. The identification of such behaviors and their relationship to student outcomes is of great importance.

Clearly, designing the content of the training programs to be evaluated is a key issue for future research. It is hoped that the design of programs for research will be an iterative process in which the results of previous research are incorporated cumulatively in the design of forthcoming programs (Carroll, 1980, page 177).

Specific teaching skills that are essential to college level instruction should be identified and taught in these programs. Secondly, criteria should be developed for measuring the successful achievement of the objectives of such programs (Smith, 1974, page 200).
The training program that was used as the treatment effect in this study was developed after an intensive review of the literature relating to: program content; intervention strategies; outcome measures. Particular emphasis was placed on the preparation of foreign TAs. The intervention strategy selected was a modelling approach using locally-produced protocol videotapes and microteaching experiences. The content selected was certain lower-inference teacher behaviors drawn from a number of studies and the outcome measure chosen was that of the achievement of the undergraduate students of the TAs. The findings of Chapter II, the Review of the Literature, formed the bases for these selections.

Statement of the Problem

Criticism concerning the quality of instruction received by undergraduate students in lecture and laboratory courses taught by teaching assistants is common on college campuses (Schade and Bartholomew, 1980). Concern for the adequacy of the preparation of TAs and the improvement of their teaching effectiveness has emerged on many university campuses across the United States (Bray & Howard, 1980; Carroll, 1980; Damarin & West, 1979; Deboer, 1979; Krockover, 1980; Staton-Spicer & Nyquist, 1979).
The Ohio State University is no exception to the phenomenon of concern about TA preparation and instruction. The training and quality of TAs has been questioned by students, faculty, departments, and by TAs (The Lantern, 1982 (d); The Lantern, 1982 (e); OSU on Campus, 1982; The Graduate Voice, 1983).

At The Ohio State University undergraduate concerns usually focus on two aspects of instruction by TAs (The Lantern, 1982 (a)). One concern is the inability of TAs in general to provide adequate instruction. The second concern is the difficulty that TAs, and foreign TAs in particular, experience in communicating clearly content matter to their students (Dege, 1979; Hood & Reardon-Anderson, 1979; The Lantern, 1982 (e)).

It is a fact however, that many graduate teaching associates are new to the podium. Some have just recently completed their baccalaureate degrees before being thrust into their new role as instructor. They are ill-prepared and inadequately supervised as they wrestle with problems of curricular content, organization, presentation and evaluation. They have difficulty responding to legitimate questions from students. They have had no experience with classroom teaching and procedure. And, many are uncertain about the most effective way of presenting the basics of their discipline. In addition some GTAs from non-English speaking countries have difficulty with the language and with the customs of our students (Haenicke, 1983).

The three purposes of this study were:
(1) To investigate if novice foreign TAs, trained to use certain lower-inference teacher behaviors, will be significantly better teachers as measured by the achievement of their students when compared to an equivalent group of novice foreign TAs not trained in this manner.

(2) To investigate the relationship between the achievement of the students of the TAs and the spoken and written English ability of the TAs of those students.

(3) To investigate the relationship between the achievement of the students of the TAs and the ability of the TAs to both assimilate and demonstrate certain lower-inference teacher behaviors.

Research Questions

The study sought to provide answers to the following specific research questions related to the training program designed as the treatment:

1. Major Research Question

Will the training program designed as the treatment for novice teaching assistants in the Departments of Statistics and Mathematics produce treatment TAs whose teaching is statistically significantly better than that of control TAs as measured by the achievement of the TAs' undergraduate students?
2. Can the lower-inference teacher behaviors selected for study be identified with consistency in the classroom by undergraduate students?

3. Do the lower-inference teacher behaviors selected for study represent reliable measures of the variable clarity?

4. Will the participant TAs in the training program assimilate the knowledge concerning the lower-inference teacher behaviors?

5. Will there be a statistically significant difference in the achievement of the undergraduate students of the treatment group TAs who do and who do not assimilate the knowledge concerning the lower-inference teacher behaviors?

6. Will there be a statistically significant difference in the demonstration of the lower-inference teacher behaviors by the treatment and control group TAs over a six-week observation period in the classroom?

7. Is there a significant relationship between the behaviors used by the TAs and the achievement scores of the undergraduate students of the TAs?

8. Is there a statistically significant relationship between the TAs' scores on tests of spoken and written English and their demonstrated performance of the lower-inference teacher behaviors?
9. Is there a statistically significant relationship between the TAs' scores on tests of written and spoken English and the achievement of their undergraduate students?

10. Is there a relationship between the TAs' scores on tests of written and spoken English and on the training program examination?

Definition of Terms

Teaching assistant (TA):
A graduate student working toward a master's or doctoral degree having responsibility for the provision of undergraduate instruction through the leading of a recitation class.

Teaching associate (TA):
This term is used synonymously with teaching assistant throughout.

Graduate teaching associate (GTA):
This term is used synonymously throughout with TA.

Foreign TA:
A TA whose first language is not English.

Novice teaching assistant:
A TA who has not previously taught in the English language in the United States and has had no formal teacher training.
Recitation class:
A class of undergraduate students led by a TA meeting for one hour twice a week for the purpose of discussing the material presented by faculty in the lectures and working the problems assigned by those faculty.

Modelling:
A process by which new responses may be learned or behaviors changed in an observer through observing the behavior of others without the observer actually enacting the behaviors or receiving reinforcement (Bandura, 1969).

Protocol videotapes:
These materials consist of videotapes of TAs teaching demonstration recitation classes designed to draw attention to the use of certain lower-inference teacher behaviors.

Microteaching:
A scaled-down lesson of five to ten minutes in length during which the TA assumes the role of a recitation leader and teaches subject content to other TAs playing the role of undergraduates while demonstrating the teacher behaviors identified in the training program.

Lower-inference behaviors:
"Those teacher behaviors whose occurrence in the classroom situation can be objectively observed and counted" (Hines, 1981, page 16). By contrast, a behavior requiring observ-
ers to make subjective judgements about its occurrence in the classroom situation would be designated a higher-inference behavior.

**Teacher performance:**
Those behaviors that the teacher, in this case the TA, uses on the job rather than those that he or she knows about (Medley, 1982). The knowledge of certain teacher behaviors may be necessary but not sufficient for the use of those behaviors in the classroom situation.

**Teacher effectiveness:**
The effect that a teacher's performance has on his or her pupils (Medley, 1982), measured in this case by the achievement on the final examination of the undergraduate students of the TAs.

**Teacher clarity:**
"In the general sense, teacher clarity refers to behaviors that teachers evidence that result in learners who have adequate interest and aptitude gaining knowledge or understanding of a topic or subject at hand" (Cruickshank, 1984).

**Delimitations**
The selection of subjects for this study was limited to new TAs in the departments of Mathematics and Statistics at The
Ohio State University. The majority of these TAs were non-native speakers of English. Attendance by the TAs at the summer training program was not mandatory but those who did attend were given a stipend for the quarter. Any generalization of the findings of this study to a different population of TAs in different departments may be in serious error and should be made with great caution.

These two departments were selected for the following reasons: (1) there is a high incidence of foreign TAs in them, (2) there is a commonality of subject matter, (3) there are usually many prospective novice TAs available during the Summer Quarter so that training may take place prior to the commencement of teaching in the following quarter, and (4) measures of student achievement can be reliably determined.

The subjects of the treatment and control groups involved in the study were assigned as paired leaders in elementary one-hundred level courses (Math. 151, Stat. 125, Stat. 133). These courses are for undergraduate students who will not necessarily be majoring in the subjects taken. Generalization from this undergraduate population to other populations should be made with caution.

Any broad generalizations across the university should be preceded by replication of findings with a number of
variables addressed. These should include a variety of departments outside of the College of Physical and Mathematical Sciences and should involve undergraduate students at a number of levels as the majority of the undergraduates involved in this study were in the freshman or sophomore year.

The study was designed to warrant the making of causal claims based on an experimental design. Any such findings must be subject to the constraints described below under Limitations. Principal analyses performed used t-tests and analyses of variance. Table 7 in Chapter III shows the analyses performed and the dependent variables for each phase of the study. The significance levels were established a priori at the 0.05 level.

**Limitations**

The treatment group consisted of eight novice TAs who were available to receive training during the summer. An equivalent number were selected as a control group from those novice TAs who did not arrive until the start of the following quarter. The control group received no training prior to teaching in that quarter. In order to control for language ability differences, the TAs' level of spoken English was not assessed until one week prior to the TAs commencing teaching in the Fall Quarter of 1983.
Given the difficulty of obtaining a true control group, partial control may be achieved by selecting from those novice TAs who could not be present during the summer. The treatment and control groups therefore were not randomly assigned and the study is consequently subject to the limitations of a quasi-experimental design. In order to reduce these limitations, steps were taken to establish treatment and control group pairs that were as equivalent as possible.

As the presence or absence of novice TAs for training during the summer is usually determined by chance factors such as difficulty in relocating it may be argued that the two populations from which the groups of novice TAs were drawn, all TAs available in the summer and all TAs not available in the summer, are similar. In order to further increase the equivalence of groups matched pairs of treatment and control TAs were established. Matching subjects provides a means of controlling variance by building variables into the design. The variables used for matching should be substantially related to the dependent variable. Details are to be found under Design in Chapter III.

The matched pairs of TAs were assigned randomly to recitation classes within each department. It was not possible to assign all TAs to the same course in the Statistics Department, nor was it possible in the Mathematics Department
to always assign each member of the TA pair to recitation groups taught by the same faculty member. These limitations were imposed upon the study as a result of departmental exigency.

The examinations for undergraduate students in both departments are common for the course to which TAs are assigned. In the Statistics Department the examination takes the form of multiple choice questions. This reduces concerns over reliability of the grading of the examination but there can be no control by the researcher over the content validity of the examinations nor is it possible to control for spuriousness resulting from uncontrolled variables such as failure of some recitation TAs to cover all of the material examined. In general, the probability of such an event is not high given the structuring of the courses, but discrepancies could occur.

While undergraduates are not randomly assigned to recitation classes by the researcher they do register for classes on a chance basis. The composition of the students in the recitation classes may therefore be reasonably considered to be similar. In order to further enhance this condition the recitation classes upon which the study focussed for its data collection were those that took place at the same time and on the same day for each member of the paired TAs. Each class was fifty minutes in duration.
Undergraduate students have a tendency to "vote with their feet". That is, they demonstrate their level of satisfaction with a recitation leader's teaching by remaining or leaving for a different section. Undergraduates registered in one section but attending another could produce bias in the effectiveness measure based on undergraduate achievement. While it is impossible to prevent undergraduate movement, steps were taken to reduce the potential for biasing the results. TAs were required to monitor the class enrollment for students who were present but not registered in the TA's section. No such instances were detected.

Those new TAs who participated in the summer treatment also took content-related courses concurrently in their respective departments. The control group of new TAs did not take content-related courses prior to teaching in the Fall Quarter. Consequently, the presence or absence of this experience, both academic and cultural, may be considered an extraneous variable and a threat to internal validity. This was deemed unlikely to prove critical as the study focused on TA teaching behaviors rather than knowledge of subject matter. It was seen to be unlikely that the short time that treatment TAs spent in this country prior to teaching in the Fall Quarter would have significant impact upon their development when compared to the effect of the
training program. Nevertheless, in order to reduce this potential further, certain critical data such as spoken English ability were collected from both groups immediately prior to teaching commencing in the Fall Quarter. Details may be found under Design.

The data collected in this study are limited by the quality of measurement attained. The study concentrates primarily on certain lower-inference teacher behaviors that research supports as characteristic of teachers who communicate clearly content matter to their students. It must be conceded that there may be other important teaching behaviors that were not addressed in this study. The behaviors utilized were identified as lower-inference behaviors. The degree to which these behaviors are lower-inference and can therefore be observed relatively easily is a potential limitation. As described under Design, steps were taken to test the perceptions of undergraduate students concerning these behaviors. This provided a reliability test situation for the observation of behaviors by undergraduate students. This preceded behaviorally defining them, and the training of expert observers in the observation of them. The presence or absence of the behaviors observed is a function of the perceptions of the observers and the level of agreement of their observations.
Each TA pair involved in the study was randomly assigned to one of the observers who visited the classroom three times during a six-week period at two-weekly intervals to collect data. The period of time involved, six weeks during one quarter, may be considered a limitation. Ideally, each TA might have been visited a number of times by a number of observers over several quarters. Temporal and financial constraints prohibited this.
Chapter II
REVIEW OF THE LITERATURE

Purpose
The purpose of this review is to ascertain from the relevant literature information concerning four major areas in the preparation of TAs. These areas are: the intervention strategies commonly used in TA preparation; the outcome measures most frequently utilized in that preparation; the literature related to the content of TA training; the concerns specific to the preparation of foreign TAs. There will be particular focus on foreign TAs in a separate section as the training program forming the treatment in this research will be designed to prepare foreign TAs to teach in the classrooms of the United States. Prior to these areas being addressed in detail a general introduction to the state of research in TA preparation will be given.

Summaries of findings will be presented throughout the review. The review will include recommendations which will focus specifically on what the literature seems to suggest for the design of a training program for TAs in general and foreign TAs in particular.

Introduction
The literature is replete with studies describing TA training. Two recent reviews have attempted to evaluate this abundance of data. Levinson-Bose and Menges (1981) found the quantity of relevant literature to be larger than anticipated but of lower than hoped for quality. Of 17 studies examined, involving TAs in workshops and seminars, confidence placed in the results was only considered high in one instance, a study by Bray and Howard (1980).

This conclusion was generally supported by a review by Carroll (1980) of empirical research on training programs for university TAs. He stated that studies reported consisted largely of descriptive data with a very limited amount of evaluative or empirical research on the effects of TA training. Carroll identified forty-eight descriptive accounts of various TA training programs "none (of which) reported actual data or statistics on the effects of training" (page 168). He also summarized the results of twenty-four studies which did assess TA training effects on student variables but points out that only eleven of these included a control group in their design. Carroll recommended that "substantially more effort ought to be devoted to assessing the effects of TA training programs than simply describing innovative ways of conducting such programs, and whenever possible that the design should include a control group" (page 176).
The literature review that follows examines sixty-four studies referring to TA preparation, of which thirty-five were empirically based using a pre-experimental, quasi-experimental, or true-experimental design as defined by Campbell and Stanley (1963). The remainder were descriptive. A large number of publications relating to but not of TA preparation are also referenced.

Both Carroll and Levinson-Rose and Menges felt that research on the improvement of college-level teaching was not building upon previous research in a systematic way. Nevertheless, they concluded that evidence existed suggesting that workshops and seminars can affect student ratings, teacher classroom behavior, and student learning. The reviews covered a variety of intervention strategies and outcome measures employed in the preparation of TAs. Their conclusions, along with those of the writer and others based on additional studies, are reviewed briefly below.

**Intervention Strategies**

The most commonly occurring intervention strategies appearing in the literature concerning TA preparation seem to be ones based on modelling approaches. These take the form of live demonstrations, the use of protocol materials, and the use of microteaching experiences. These three areas will be briefly reviewed and a summary of findings presented.
The empirical base for modelling as a means of producing learning lies in social learning theory. Although this approach has also been termed "imitation", "observational learning", and "vicarious learning" the basic contention is that new responses may be learned or behaviors changed through observing the behavior of others without the observer actually enacting the behaviors or receiving reinforcement (Bandura, 1969). That behavior may be modified by observing another engaging in the behavior that is desirable has been a conclusion reached by a number of research studies (e.g., Bandura & Walters, 1963; Krumboltz & Thoresen, 1969).

1. Live Demonstrations

One quasi-experimental study (Costin, 1968) cited demonstrations of discussion methods and of lectures by the faculty. Participants were able to evaluate these and rated the experience as especially helpful. The treatment group in this study made significant gains only in student ratings of student-teacher rapport.

Numerous descriptive studies (Armenti & Wheeler, 1978; Bailey, 1977; Brinton & Gaskell, 1979; Goepper & Knorre, 1980; Grasha, 1978; Monson, 1970; Russo, 1982, 1978) generally support the value of modelling. "Outstanding" or "distinguished" faculty (Rose, 1972) have been used to present different teaching styles (Monson) and to relate their own
teaching experiences to problems relevant to TAs (Rose). Such components of programs were generally found to be highly valued by participants. Instructors modelling "how to" and "how not to" were favorably received by program participants (Brinton and Gaskill) as was the modelling of particular instructional methods (Grasha).

Generally positive student ratings of TAs were found to result for those participating in a program in which instructors demonstrated each skill as it was introduced with the TAs playing the role of students (Goepper and Knorre). In another study (Szymanski) the "apparent continued progress of TAs indicated positive outcomes emerging as a result of the training program" during which TAs were appointed to a different member of faculty each semester. The faculty member's responsibility included modelling effective instructional techniques.

TAs showed "genuine and significant improvement" as a result of a program which included role playing an office hours session with the professor playing a confused student. Emphasis was placed upon TAs being clear as to what the student was really asking and the need for the TA to ask clarifying questions rather than launch immediately into a lecture (Armenti & Wheeler). Role playing also formed a part of two programs (Bailey; Brinton & Gaskill), both "well accepted" by TAs, the former designed for for-
eign TAs. In this program (Brinton & Gaskill) role playing involved issues such as grading, cheating, and plagiarism.

2. Protocol Materials

Protocol materials may be defined as films or videotapes illustrating educationally relevant concepts (Levinson-Rose & Manees, 1981). They may also include written materials as well as film or videotape recordings that illustrate education target skills or concepts (Sharp, 1981).

The use of model videotapes as a means of influencing behavior has been well documented (Bandura, 1965a; Bandura & Walters, 1963; Flanders, 1968; Ford, 1979; Hosford, 1980; Ivey, 1971). Bandura (1965b) demonstrated that adults may acquire attitudes, emotional responses, and complex patterns of behavior through exposure to pictorially presented models. Stage models had been used successfully in film-mediated, social-learning experiments with children (Bandura, 1969), in therapeutic settings (Kagan, Krathwohl, & Miller, 1963), group problem-solving settings (Walter, 1975), improving counselling interviewing skills (Kurpius, Froehle, & Robinson, 1980), and improving supervisors interpersonal skills in dealing with employees (Latham & Saari, 1979) (Sharp, 1981, page 492).

Sharp, using a true experimental study in which the viewing of a model videotape demonstrating effective lecturing was a prominent feature concluded that such viewing influenced the subsequent behaviors of the participants as assessed by trained observers. Several pre-experimental
and descriptive studies refer specifically to the use of protocol materials. As in the previous study demonstration videotapes were used by Barrus, Armstrong, Renfren and Garrard (1974) in addition to microteaching experiences. In the latter study the protocols demonstrated skills including those of response repertoire, questioning, increasing student participation, and presentation. Although no specific comment was made on the effectiveness of the protocol materials a "general decrease in student complaints about TA teaching" was reported.

Using the chalkboard, nonverbal cues, student initiated talk, and closure were the subject of protocols forming the basis of a program for geology TAs (Schade & Barthcolmew, 1980). The participants felt the program to have increased their awareness of the importance of the four techniques. As in the previous study the videotapes used in a similar study (Szymanski, 1978) were of experienced TAs in the classroom situation. The program was felt to lead to positive outcomes in the TAs' teaching. Damarin and West (1979) utilized videotapes of TAs' classes in progress in order to prepare novice foreign TAs. This correlational-based study was described as "generally successful".

White (1981) used seven modules at the core of a training course for chemistry TAs. These consisted of videotapes dealing with introduction to a course, performance
OBJECTIVES, question ing skills, reinforcement, testing, tutoring, microteaching, and interaction analysis. Additionally, written materials were provided to accompany the videotapes. In general participants felt that the modules were useful if not always directly applicable to the laboratory situation. The participants in the program were "rated highly in their teaching by their students at both mid-term and the end of term", although no specific data were reported. Three other studies (Bailey, 1977; Brinton & Gaskill, 1979; Flanigan, 1978) referenced the use of protocol materials, one involving the evaluation by the participants of a "well delivered videotaped lecture" (Brinton & Gaskill). All studies reported the programs to be well accepted by the participants. Filmstrip tape programs were used "once or twice" during a course described by Rose (1972) but were "originally prepared for elementary and secondary level teachers, however, and proved somewhat less exciting for the teaching assistants."

A recent review of research on improving college teaching concluded that the use of protocol materials has not been researched in postsecondary education (Levinson-Rose & Menges, 1981). This appears to be supported by this present review of the literature as far as the preparation of TAs is concerned although there is evidence for the effectiveness of protocol materials in other areas of teacher prepa-
Prior to 1981 no empirical data had been reported in support of the efficacy of protocol materials in bringing about change in teacher behavior.

Sharp (1981) identified a number of studies evidencing desired behavior change and concept acquisition in teachers as a result of exposure to protocol materials. Levinson-Rose and Manges (1981) also cited a number of similar studies, supporting the efficacy of protocol materials in teacher skill and concept acquisition and in behavior change that was related to pupil achievement. It was suggested that the combination of microteaching and protocol materials was more powerful than either approach alone.

The literature reviewed by Sharp suggested that "maximum transfer of modelled behaviors occurs when the setting of the model is as similar to the transfer setting as possible" (page 49). Consequently the protocol materials were designed with this in mind but Sharp concluded that the relevance to the viewer and the viewer's interest in the topic may not be important in effecting significant acquisition of the skills modelled (page 500). That is, the participants need not identify with the subject matter in order to acquire the skills modelled.

Sharp's point is well taken but there exists as much support for the protocol materials to be congruent with the
purpose of classes taught by TAs and the role expectations held for those TAs (e.g., Blizzard, Hogan & Roy, 1981, page 5; Rose, 1972). This may be particularly important for inexperienced TAs.

A further advantage of the use of protocol materials is that the trainee is neither required to participate in the videotaping of his behaviors or practice the behaviors before a group. In this respect, protocol training alone is less threatening to the participant but not necessarily as effective when used alone. Performance and feedback concerning that performance are perhaps the two most important of the phases of instruction as delineated by Gagne (1975).

One of the critical questions concerns the concepts that should be exemplified by protocol materials for use in the preparation of teaching assistants. In a survey of protocol materials (Cooper, 1975) evidence was presented for the acquisition of concepts as well as evidence of change in teacher behavior. Studies cited included: mastery of concepts which included encouragement, clarity, emphasis, feedback, and organization; recall and the ability to identify concepts as part of a teaching vignette; concept acquisition and transfer; the ability of protocols to teach concepts; effectiveness of protocols in the areas of lesson organization, using student ideas, questioning, praise, and corrective feedback; significant learning using protocols
as cognitive interaction, affective interaction, classroom management, and counselling. One study cited suggested that the viewing of films illustrating concepts about teaching behavior "leads to a demonstrable acquisition of those behaviors (as well as the acquisition of concepts about those behaviors)" (Cooper, page 72).

Cooper, in the summary of findings and recommendations, suggested that the skills and instructional concepts which research has identified as influencing teaching and learning should be prime material for protocols. Secondly, the relationship between the use of materials and change in student achievement was stressed. Cooper suggested that evidence of trainee learning should, when feasible, be collected through audio-and video-taping as being superior to the evidence produced by paper-and-pencil tests.

Sharp (1981) found existing audiovisual material to be inappropriate for use with inexperienced TAs and consequently produced his own protocol materials. A detailed description of the preparation of videotapes by and for TAs to use as protocol materials is given in a study involving twenty geology TAs (Schade & Bartholomew, 1980). The opportunity to view TAs in the same field enabled trainees to identify with the teaching event and "made a very positive impact on the teaching assistants" (page 101). This is supported by related literature that suggests maximum transfer
of modelled behaviors occurs if the setting for the model is closely similar to that of the transfer setting, that is, the one in which the trainee will function (Sharp, 1981).

Taking the concept of modelling one step further involves the use of self-as-model. This concept was applied to the production of protocol materials in an observational procedure for improving university instruction (Perry, Leventhal, & Abrami, 1981). Unfortunately, given the design, the results may be open to question. The treatment group and control group each contained only two participants one of whom in each was designated as initially "high effective" or "low effective". The two trainees after teaching a videotaped pre-treatment lesson were assisted in the construction of a master tape which would depict the trainee exhibiting the component sequence of the complex behavior designated as "effective teaching".

As the trainee role played, the three judges observed the sketch and then rated it on a 10-point scale. If the three judges rated the sketch being portrayed "very well" ($\bar{x} > 8$), the sketch was considered a "take" and was transferred to the trainee's master tape. If $\bar{x} < 8.00$, the trainee was informed and was given specific feedback for improving the portrayal of the sketch. This procedure was repeated until sketches for all four teaching behaviors were transferred to the master tape (Perry et al., 1981).
In other words, the trainee practiced until perfect, or at least nearly so. The master tape was then given to the trainee along with the pre-treatment video tape. Cognitive discrimination occurred by contrasting the effective teaching behaviors with the pretreatment presentation. The trainees were asked to "spend three and one half hours each week observing the tapes." A videotaped post-treatment lesson using equivalent students was taught by each participant and as before student ratings and answers to a multiple choice test on the content were collected. The results indicated that the treatment was effective for the initially high effective participant only. The low effective participant received lower ratings than either of the control participants on lecture value. An examination of the data presented shows that the low effective trainee had pre-treatment scores worse than any other participant for lecture value and teaching ability and did in fact make some gains post-treatment whereas the non-treatment trainees generally made losses. The non-treatment trainees also made losses on both measures of student achievement whereas the low effective trainee made one loss and one gain, not comparable in size to the two gains shown by the high effective trainee.

While the data may be of dubious value the study serves to illustrate two important points. Firstly, the experience
of taking part in a self-as-model experience and viewing self-as-model on protocol videotapes does seem to have been effective (contrary to the conclusions reported in the study) for both the high and low effective trainees compared to the non-treatment participants. Secondly, while the high effective trainee made better improvement than the low effective trainee, it is important to note that neither apparently received any feedback on their teaching nor any consultation while viewing the master tape. The importance of feedback and consultation will be addressed in subsequent sections.

In the self-as-model approach, it is held that the individual is more likely to learn behavior from observation of himself or herself effectively performing that behavior. The procedure has been used successfully in instructional settings to alter teaching behaviors (Staton-Spicer & Nyquist, 1979). Microteaching, which follows, has as its basis this self-as-model approach.

3. Microteaching

Introduction

Microteaching may be defined as a scaled-down lesson of five to ten minutes in length during which the teacher presents content but concentrates upon a specific teaching skill throughout the presentation. As originally conceived
at Stanford, microteaching was designed to meet a specific need. This was the provision of a reality situation for preservice teachers that would put them in contact with real students and themselves as teachers and familiarize them with their own behavior. The first technical skill used was "how to begin a lesson" and others were added as the research progressed. It is important to realize that these various skills were "under investigation". At the time it was a moot point as to whether these skills were worth learning. This was clearly realized by the then researchers and no implications otherwise are intended. Whatever the various sources of these skills and the reasons for inclusion in the research the technical skills themselves became a focal point of microteaching (Davis, 1981) although the originators foresaw "danger of locking in too early on a first alternative which arose purely out of chance and convenience" (Allen & Ryan, 1969, Page iii).

By and large, the original skills used at Stanford have seen little modification or development and tend to exist in isolation to one another. A more adequate formulation of the skills and a conceptual structure into which they fit may be a prerequisite for any further development of this aspect of microteaching that seems to be almost universally accepted as a fait accompli (Davis, 1981).
HcKnight and Bush (1977) commented on the great flexibility of the components within microteaching, permitting adaptation to a wide variety of teacher education programs. It was reported in their review of studies in microteaching that five to ten minutes seemed to be the acceptable limits, because there were no noticeable attempt to investigate alternative lesson lengths (McKnight & Bush). It has been suggested that:

Not only can the length of the lesson and the size of the group could be increased but the teaching skills practiced could be rigorously, but less narrowly defined; the focus upon not one particular skill but a group of skills. Such a graded increase would not only allow the combined practice of skills experienced previously but would permit the placing of more emphasis on skills such as lesson structuring, decision-making and classroom management which would become increasingly more relevant as school-based work approached (McIntyre, McLeod & Griffiths, 1977, page 251).

Studies Utilizing Microteaching

A number of descriptive and pre-experimental studies include the use of microteaching in programs to train TAs (Armenti & Wheeler, 1978; Barrus et al., 1974; Brinton & Gaskill, 1979; Crooks, 1980; Garland, 1969; Golman, 1975; Grasha, 1978; Mellon & Dence, 1971; Monson, 1979; Rose, 1972; White, 1981). In two of these (Armenti & Wheeler; Barrus et al.) microteaching formed the core of the experience for participants.
Two studies in TA training in the hard sciences support the efficacy of microteaching with feedback. In one (Bar-
rus et al.) a two hour microteaching laboratory was held weekly for TAs undergoing training in a chemistry setting. Each TA gave a five minute presentation and received an im-
mediate critique from the group. It was reported that ninety percent of the TAs supported the value of the micro-
teaching experience and that fewer complaints about TAs' teaching ensued from students. Armenti & Wheeler had each foreign TA in a physics setting give a thirty minute video-
taped lecture to the group. Immediately after the lecture feedback was provided by the professor and other TAs and the TA was required to produce a one-page critique as self-
evaluation of the lecture. Advantages for TAs to study speech and behavior patterns were emphasized.

Videotaping with feedback has also been used for role playing and laboratory introduction sessions. The ability to objectively compare tapes of an individual over a period of time was cited as being particularly useful. Although no supporting data were offered, it was concluded that par-
icipants began to show genuine and significant improvement in their teaching.

Microteaching experiences formed a major part of a pro-
gram for foreign TAs (Brinton & Gaskill) during which each TA made three videotapes All videotapings were before
peers, the first being a one-minute impromptu talk and the other two consisting of two three-minute explanations of concepts from their major field. The TA was allowed to review and evaluate this performance immediately afterward followed by critiques from the instructor and group, both orally and using a checklist completed for the TA presenting. Participant evaluations of the course rated these experiences highly, more so than any other part.

Videotaped peer teaching experiences formed a part of a number of other programs (Crooks, 1980; Garland, 1969; Goldman, 1975; Grasha, 1978; Mellon & Dence, 1971; Monson, 1970; Rose, 1972; Smith, 1974; White, 1981). All received immediate feedback and critiquing by the participants, oral or oral and written using checklists. Support for the experience was generally reported as positive.

Microteaching does not of necessity require the presence of videotaping equipment. Three studies (Bucknayer, 1972; Goepper & Knorre, 1980; Grasha, 1978) reported using the microteaching experience without videotaping. A generally positive response was reported in each instance, where the participant made a class presentation which was subsequently critiqued by the instructor and peers. Grasha initiated the technique by having his own presentations critiqued in order to model the kind of behaviors required.
These studies involving the training of TAs generally support the efficacy of the microteaching experience. They also tend to follow the generally accepted sequence established at Stanford, that of plan, teach, observe, critique, replan, reteach, reobserve, though the second lesson is not usually an improved version of the first one as originally conceived. Under the original conception the process may be repeated until a specified level of mastery was achieved.

Support for microteaching may be found in other literature than that dealing with TAs. In one review of the research on microteaching it was concluded that microteaching facilitates the acquisition of teaching skills and the development of fashionable attitudes toward teaching and that it does so in a relatively short time (Turney, Clift, Dunkin & Trial, 1973).

McKnight and Bush (1977) in reporting comparative studies of microteaching and other approaches to teacher training concluded that the superiority of microteaching was apparent. The concept of microteaching has been seen to represent a significant departure from traditional forms of teacher training (Brown, 1975). Inherent in microteaching is the opportunity to experience reality conditions that courses, readings, conversations and modelling alone do not provide and in an environment that is supportive and
conducive to the development of confidence. The positive attitudes expressed by TAs toward the microteaching experience in the literature was supported by Turney et al. who, stated that "most students acknowledge that any stress or artificiality are outweighed by the considerable advantages to be obtained from the experience" (page 10).

Microteaching is not without impact in the affective development of teachers. In general studies support the use of microteaching in improving teacher attitudes toward pupils in terms of a more interactive approach and the creation of a positive classroom atmosphere (Brown, 1975; Levinson-Rose & Manges, 1981).

It has been reported that microteaching measures and student-teacher self-concept were found to be unrelated (Freeman & Davis, 1974). Contrary findings by Stanton (1978) indicated that the basic requirement for improving the self-concept was present in the microteaching situation. Using the microteaching format in the development and field testing of a program to prepare teachers in interpersonal skills it was observed that teachers' affective skills may be increased via training procedures which include the identification of specific behavioral components of affective teaching, the observation of models performing the desired behaviors, and provision for immediate feedback on student performance of the behaviors (Sadker,
Sadker, & Strain, 1977). Teachers who think positively about themselves, who like themselves as people, are more able to encourage similar attitudes in those they teach than those Viewing themselves negatively (Combs, Blume, Newman & Wass, 1974).

Stanton (1978) further suggested that videotape feedback of teaching may be instrumental in the development of teacher self-concept. Feedback is considered an essential part of the microteaching experience and can have a special role in the preparation of foreign TAs as described below.

Video playbacks also permit the instructor to individualize his or her comments and thereby to reach students at their varying levels of skill and experience. Interestingly, many students are able to discern their own pronunciation errors. Correction, however, is very difficult, particularly in the context of a fast-moving oral presentation (Franck, Marion, & DeSouza, 1982, page 113).

The tapes were excellent for coaching the foreign students in their pronunciation—the key to the matter is self-awareness. The videotape experience forces us to become really aware of what he/she is saying and doing (Armenti & Wheeler, 1978, page 123).

Two studies (Armenti & Wheeler; Franck et al.) reported the efficacy of videotaping in terms of student increased self-awareness, self-confidence and improved effectiveness. Obviously such an approach is not limited in effectiveness to foreign TAs.
Several reviews of the state of the art in microteaching (McKnight & Bush, 1977; McIntyre, McLeod & Griffiths, 1977) addressed the importance of feedback. The latter referenced studies including audio as against visual feedback; face-to-face compared to video with face-to-face, and video with audio; and supervisor feedback compared to peer feedback and no feedback. Little difference between treatment groups was reported. The former (McIntyre et al.) reported on studies involving aspects of the feedback process, modelling, cueing, discrimination training, repeated presentation, and personal evaluation. It was found that students wished for a variety of feedback types involving both peers and supervisor, the latter being considered more valuable, in which the supervisor's and the student-teacher's perception of both strengths and weaknesses were considered.

In playback, instructor and analyst view the tape together. Both comment constructively on the teaching techniques observed. This two-person playback model is fairly common in teacher education programs outside of foreign languages. A review of research in teacher education by Fuller and Manning strongly supports the view that, "confrontation", i.e., identification of a discrepancy between the person's view of reality and that of some observer, .... or at least some focus, is necessary. Feedback that is not accompanied by some focus has been found to change behavior little if at all (Franck & Samaniego, 1981, page 273).
There is a clear trend for microteaching with discrimination training (i.e., practice in identifying the effective and ineffective use of the skill) to lead to superior performance. Wagner (1973) offered support for this in a study comparing groups on discrimination training, microteaching and more conventional preparation. The first group was found to be more student centered in their teaching than the other two groups which did not differ from one another. It was suggested that without discrimination training microteaching may be ineffective. As suggested in a recent review of research in college teaching (Levinson-Rose & Manges, 1981, page 415) the role of discrimination training may be critical.

Support was provided for the importance of personal consultation in the provision of feedback (Planigan, 1978; McKeachie et al., 1980). It was reported that simple knowledge about one's teaching effectiveness does not enable a teacher to improve as much as does the same information presented with encouragement and suggestions (McKeachie et al.). The person providing the feedback apparently needs to convey more than just the bald findings concerning effectiveness. Allen (1976) maintained that a training program would not succeed without the maintenance of a sympathetic and understanding atmosphere between the coordinator and the TAs (page 27). The instructors' behavior engendered in
the TAs a conviction that what they were doing was important and was seen as an important factor in the TAs' development by Armenti and Wheeler (1978). Grasha (1978), too, emphasized the importance of feedback and the establishment of a course atmosphere where feedback was accepted. Support for the importance of feedback in bringing about behavior change is also found in a non-training study by Pam-bookian (1974).

Microteaching provides the means of both experiencing the teaching situation and of obtaining immediate feedback about the experience. TAs desire some immediate practice during preparation but attempts to fully prepare a novice TA for teaching before meeting his first class may not be meaningful in the absence of some experience (Allen, 1976, page 26). Microteaching provides this experience and enables behaviors to be explored within the context of a reality situation.

Observation of oneself performing effective behavior is the optimum familiar situation. Thus, there is justification for using the individual as the model ("self-as-model") to maximize similarity and facilitate learning (Staton-Spicer & Nyquist, 1979, page 204).

Microteaching is usually but not always characterized by the presence of videotaping equipment and by the presence of peers playing the role of students. The setting which
microteaching takes place therefore is usually not the regular classroom with the regular students of the teacher being videotaped. Consideration will be given to the two most commonly reported means of observation of TAs teaching regular classes as reported in the literature. Essentially these involve the presence or absence of videotaping equipment.

Classroom Videotaping

Three true-experimental studies (Bray & Howard, 1980; Carroll, 1977; Dalgaard, 1976) used classroom observation of teaching through videotaping as part of their program. Bray and Howard (1980) assigned TAs to one of four treatment groups, two of which received three hours of individual instructor feedback and consultation using videotapes of one or more of their regular classes. Only the first of these two groups participated in a ten-hour training seminar. The third and fourth group received two hours of consultation and feedback (no videotaping) and no treatment of any kind respectively. There were no significant differences found between the first two groups. Hence the value of the ten hour training seminar component was not clear and it was suggested tentatively that videotaping with subsequent consultation is a most effective method of preparing TAs. Both groups receiving this treatment were rated higher by judges over the other two groups on behavioral dimensions.
This result tended to be supported by Carroll (1977) who assigned TAs to a treatment or control group who viewed videotapes of their classroom teaching in the presence of the instructor. Critiquing was provided by the instructor only for the TAs in the treatment group. These TAs later received significantly higher ratings from students on measures of student-centered teaching. Further support came from the study by Dalgaard (1976) who assigned TAs to a treatment group receiving training or a control group receiving no training. The training included the TA viewing videotapes of his classroom teaching with the instructor. Expert ratings of TAs were significantly higher at the end of the program for the treatment group.

In a quasi-experimental study (Koffman, 1974) the TAs in two of three groups made thirty-minute videotapes of their classroom teaching at the start and end of an eight-week period. Both groups of TAs reviewed the initial tapes with instructional specialists but only the first group continued to meet with the specialists who provided additional experiences designed to facilitate changes in instruction. This group showed positive changes in instruction and in "clarity", "education and feedback" and "relating to student responses". The third group, the control group, also made positive changes in "relating to student responses" but the second group regressed in this category. While the
study lends some support to those preceding it the results are not conclusive concerning TA behavior change.

Student volunteers for a minicourse taught by TAs in mathematics constituted students in the class for TAs who were videotaped (Damarin & West, 1979). Selected segments were then viewed and discussed by the whole group focusing on such skills as interpreting student remarks and questions or involving students in problem solving. The foreign TAs participating in the program generally indicated an improvement in their teaching as a result of participation. A number of descriptive and pre-experimental studies support the value of videotaping TAs in the classroom and providing feedback on the teaching (Allen, 1976; Barley, 1977; Barrus et al., 1974; Crooks, 1980; Manteuffel & Von Blum, 1979; Mozer & Napell, 1974; Szymanski, 1978). TAs generally reported such experiences as helpful and the effectiveness of the TAs' teaching was deemed to have increased as a result of the program. In some instances peer feedback did not accompany instructor feedback (Crooks; Manteuffel & Von Blum; Mozer & Napell).

It is unclear as to whether critiquing was used during a four week pilot program (Smith, 1974) in which TAs made a ten-minute videotape at the beginning and end of the program. Instruction in verbal interaction analysis was given and a programmed reading unit on effective questioning was
assigned prior to the second videotaping. Examination of the tapes revealed a general decrease in TA information giving and increase in asking questions. Student ratings were collected pre and post the second videotaping but the data were not reported.

In two true-experimental studies (Haber, 1973; Murphy, 1972) TAs in treatment groups experienced classroom observation without videotaping by instructors followed by consultation with the instructor and, in the case of the former, instruction in questioning techniques. Haber did not find significant changes in teaching performance whereas Murphy reported treatment TAs to be judged more successful in promoting interaction in the classroom. In a non-training true-experimental design study, TAs in the treatment group received no classroom observation but personal feedback on student ratings from and consultation with an expert teacher. These TAs were rated most highly at the end of the term by students. Other groups received either a computer print-out of student ratings or no information at all (McKeachie et al., 1980). Using a quasi-experimental design training study Yagbilian (1972) concluded that the students of TAs who received consultation about their teaching during two terms were rated as more satisfactory by their students than TAs who received no consultation.
A number of descriptive and pre-experimental training studies support the value of classroom observation with consultation (e.g. Allen, 1976; Armenti & Wheeler, 1978; Goeppeper & Knorre, 1980; Krockover, 1980; Manteuffel & Blum, 1979; Muhlstein & De Facio, 1974; Spears & Zollman, 1974; Szymanski, 1978). Allen indicated that all TAs rated the experience as useful and Manteuffel & Von Blum reported a significant improvement (not statistically tested) in TA performance as perceived by their students. Similar significant improvement in laboratory instruction was reported (Armenti & Wheeler; Muhlstein & DeFacio) as a result of programs involving classroom observation by the instructor and generally positive results were reported by Goeppeper and Knorre. In a survey involving TA supervision practices it was reported that the majority of the respondents considered direct classroom observation procedure, including videotaping, as the most effective (Schulz, 1980). Such procedures were also reported to be perceived as the means of evaluation most threatening to TAs.

It was concluded in a recent review of the research in improving college teaching that there was general support for feedback with consultation as an intervention (Levinson-Rose & Manges, 1981). Support for the provision of more than simply feedback may be found in a number of studies (e.g., Allen, 1976; Armenti & Wheeler, 1978; Bray & How-
ard, 1980; Brophy, 1979(b); Franck & Samaniego, 1981; McKee-chie et al., 1980; Miller, 1971; Pambookian, 1974).

**Intervention Strategies: Summary and Comment**

A multitude of studies support the use of modelling as an intervention strategy. Among these a few quasi- and true-experimental designs support the efficacy of modelling in improving student ratings (Costin, 1968; Hinofotis and Bailey, 1981) and improving TA behavior (Bray & Howard, 1980; Dalgaard, 1976; Murphy, 1972; Sharp, 1981) with one dissent (Haber, 1973). Live demonstrations involving faculty and TAs have been associated with outcomes including favorable reception by the TAs of the program (Armenti & Wheeler, 1978; Bailey, 1977; Brinton & Gaskill, 1979; Grasha, 1978; Monson, 1970; Rose, 1972; Szymanski, 1978) and improved student ratings (Goepper & Kmorre, 1980).

The use of protocol materials receives strong support from numerous studies including two true-experimental studies (Goodwin, 1982; Sharp, 1980). Sharp concluded that "viewing a model videotape did influence subjects micro-teaching lectures as assessed by trained observers." He suggested that the results indicate the potential efficacy of the approach at the post secondary level, an area in which the use of protocols has seen little research (Cooper, 1975; Levinson-Rose & Manges, 1981).
In these two last cited reviews (Cooper; Levinson-Rose & Mengers) evidence is presented for the acquisition of concepts through the use of protocol materials, although Levinson-Rose and Mengers also cite studies supporting behavior changes through the use of protocols alone and coupled with microteaching. Two studies (Schade & Bartholomew, 1980; Sharp, 1981) support the use of locally produced protocol materials and indicate, along with Szymanski (1978), that the closer the teaching depicted by the protocols to that of the trainee the better.

As may be seen from Table 1, the use of microteaching experiences in TA preparation is high and it may be considered well supported as an intervention strategy. The volume of studies includes several of pre-, quasi-, and true-experimental design. These, along with other descriptive studies and reviews (McIntyre et al., 1977; McKnight & Bush, 1977; Schulz, 1980) tend to emphasize the importance of feedback or consultation regarding observed teaching experience. In their review Levinson-Rose and Mengers (1981) concluded that microteaching shows promise for improving teaching.

Finally, several studies (Levinson-Rose & Mengers; McIntyre et al., 1977; Wagner, 1973) suggested that discrimination training may play a critical role in the acquisition of concepts and behaviors. The implication is that unless TAs are trained to recognize appropriate and inappropriate
behaviors they may be unable to change their performance to demonstrate the former.

**Outcome Measures**

**Introduction**

A preliminary review of the literature revealed four major categories of outcomes resulting from studies concerned with TA preparation. This section will focus on these which are: change in TA knowledge; change in student ratings of TAs; change in TA behavior; change in the achievement of TAs' students. Confirmation for this categorization may be found in reviews by Levinson-Rose and Menges (1981) and Carroll (1980) although the latter briefly addresses the small number of studies involving changes in TA and student attitudes the results of which are inconclusive.

Table 1 below lists empirically-based studies concerned with producing change in TA behavior as a result of training. An empirical as distinct from a descriptive study is one reporting actual data or statistics even though statistical analyzes may not have been carried out on the data. The design of each study is indicated, pre-experimental, quasi-experimental, or true-experimental and the change variable described as much as the literature permits.
The measure used to determine the change was that of ratings by non-student observers of either classroom teaching or some proxy of classroom teaching. Additionally, claims by the author(s) of statistical significance are shown and the principal intervention strategies are indicated.

Similarly, in Table 2, studies are listed in which ratings by students formed the measure of change in TA behavior. As before, these ratings were based on observed classroom teaching or some proxy.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Design</th>
<th>Variable Descriptor(s)</th>
<th>Significance</th>
<th>Intervention Strategies Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen</td>
<td>1976</td>
<td>P</td>
<td>Organization, presentation, attitude toward students and students' problems, helpfulness, motivation</td>
<td>Yes</td>
<td>Seminars, classroom videotaping, playback with instructor feedback</td>
</tr>
<tr>
<td>Anderson</td>
<td>1973</td>
<td>Q</td>
<td>More student response and teacher questioning, positive reinforcement and silence, less teacher lecturing</td>
<td>Yes</td>
<td>Seminars, microteaching</td>
</tr>
<tr>
<td>Arbes &amp;</td>
<td>1974</td>
<td>T</td>
<td>Ability to facilitate group discussion</td>
<td>NT</td>
<td>Protocol videotapes, role playing</td>
</tr>
<tr>
<td>Kitchener</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray &amp;</td>
<td>1980</td>
<td>T</td>
<td>Organization, instructional objectives, managing classroom interactions, utilizing feedback effectively</td>
<td>Yes</td>
<td>Seminars, videotaping with feedback and consultation</td>
</tr>
<tr>
<td>Howard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carroll</td>
<td>1976</td>
<td>T</td>
<td>Accepting student feelings, giving praise, using student ideas, asking questions</td>
<td>Yes</td>
<td>Videotaping and critiquing, workshops</td>
</tr>
<tr>
<td>Dalgaard</td>
<td>1976</td>
<td>T</td>
<td>Stating aims and objectives, involving students</td>
<td>Yes</td>
<td>Seminars, videotaping and feedback</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Method</td>
<td>Indirectness, flexibility</td>
<td>Yes/No</td>
<td>Training/Analysis</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>--------</td>
<td>---------------------------</td>
<td>--------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Daniels</td>
<td>1970</td>
<td>T</td>
<td></td>
<td>Yes</td>
<td>Training in Flanders Interaction Analysis</td>
</tr>
<tr>
<td>Goodwin</td>
<td>1982</td>
<td>T</td>
<td>Lesson objectives, use of</td>
<td>Yes</td>
<td>Workshops, protocols, microteaching</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>questions, closure, eye</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>contact, variability, ac-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>knowing student</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>contributions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haber</td>
<td>1973</td>
<td>T</td>
<td>Class questioning tech-</td>
<td>No</td>
<td>Training in Flanders Interaction Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>niques, giving objective</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hinofotis</td>
<td>1981</td>
<td>P</td>
<td>Vocabulary, grammar, eye</td>
<td>Yes</td>
<td>Videotaping, role playing, seminars</td>
</tr>
<tr>
<td>&amp; Bailey</td>
<td></td>
<td></td>
<td>contact, clarity of ex-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pression, ability to</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>relate to students, de-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>velopment of explanation,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hockett</td>
<td>1972</td>
<td>P</td>
<td>Less teacher control,</td>
<td>Yes</td>
<td>Seminars</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>more individual interaction,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>more high level question-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koffman</td>
<td>1974</td>
<td>Q</td>
<td>Use of student ideas, or-</td>
<td>NT</td>
<td>Videotaping with consultation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ienting statements (e.g.,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>summary, introduction)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murphy</td>
<td>1972</td>
<td>T</td>
<td>Praise and encouragement,</td>
<td>Yes</td>
<td>Seminars, microteaching class observation, individual</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>asking more questions</td>
<td></td>
<td>conferences</td>
</tr>
<tr>
<td>Rhyne</td>
<td>1973</td>
<td>P</td>
<td>More time with students,</td>
<td>Yes</td>
<td>10 hours of seminar</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>more convergent and diver-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>gent questions, more</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>indirect talk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Year(s)</td>
<td>Design</td>
<td>Findings</td>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>Rodriguez</td>
<td>1982</td>
<td>T</td>
<td>Wait time, verbal reinforcement, summaries, acknowledging correct answers</td>
<td>Yes, Microteaching and individual conferences</td>
<td></td>
</tr>
<tr>
<td>Sharp</td>
<td>1980, 1981</td>
<td>T</td>
<td>Lesson objectives, use of questioning, closure, eye contact, variability, acknowledging student contributions, and others</td>
<td>Yes, Protocol videotapes, workshop, microteaching, modelling</td>
<td></td>
</tr>
<tr>
<td>Smith</td>
<td>1974</td>
<td>P</td>
<td>Asking questions or requiring information about content, subject, or procedure</td>
<td>NT, Programmed instruction package (written)</td>
<td></td>
</tr>
<tr>
<td>Tubb</td>
<td>1975</td>
<td>T</td>
<td>More indirect teaching, less lecturing</td>
<td>Yes, Workshop</td>
<td></td>
</tr>
</tbody>
</table>

Key to experimental designs (Campbell & Stanley, 1963): P = pre-experimental design; Q = quasi-experimental design; T = true-experimental design. NT indicates that data were not tested for statistical significance. I indicates that the findings were inconclusive.
### TABLE 2

**CHANGE IN TA BEHAVIOR AS MEASURED BY STUDENT RATINGS**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Design</th>
<th>Variable Descriptor(s)</th>
<th>Significance</th>
<th>Intervention Strategy Descriptor(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexander &amp; Davis</td>
<td>1970</td>
<td>P</td>
<td>Overall effectiveness</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>Allen</td>
<td>1976</td>
<td>Q</td>
<td>Effectiveness in discussion of content</td>
<td>NT</td>
<td>Seminar, classroom videotaping, playback with instructor feedback</td>
</tr>
<tr>
<td>Arbes &amp; Kitchener</td>
<td>1974</td>
<td>T</td>
<td>Ability to facilitate group discussion</td>
<td>Yes</td>
<td>Protocol videotapes, role playing</td>
</tr>
<tr>
<td>Bray &amp; Howard</td>
<td>1980</td>
<td>T</td>
<td>Unspecified &quot;specific&quot; behaviors</td>
<td>Yes</td>
<td>Seminars, videotaping with feedback and consultation</td>
</tr>
<tr>
<td>Buckenmeyer</td>
<td>1972</td>
<td>P</td>
<td>Unspecified &quot;effectiveness&quot;</td>
<td>NT</td>
<td>Lesson preparation, demonstration teaching, seminars</td>
</tr>
<tr>
<td>Carroll</td>
<td>1976</td>
<td>T</td>
<td>Agreement between objectives and tests</td>
<td>Yes</td>
<td>Workshops, videotaping and critiquing</td>
</tr>
<tr>
<td></td>
<td>1977</td>
<td></td>
<td>Agreement between objectives and teaching</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Overall effectiveness</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Costin</td>
<td>1968</td>
<td>Q</td>
<td>Student-teacher rapport</td>
<td>Yes</td>
<td>Seminars, demonstration lessons with critiquing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Skill, structure, feedback, interaction</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Dalgaard</td>
<td>1976</td>
<td>T</td>
<td>Stating aims and objectives, involving students</td>
<td>No</td>
<td>Seminars, videotaping and feedback</td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>Method</td>
<td>Item</td>
<td>Yes/No</td>
<td>Training Methods</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
<td>--------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Goodwin</td>
<td>1982</td>
<td>T</td>
<td>Global effectiveness item</td>
<td>No</td>
<td>Workshops, protocols, microteaching</td>
</tr>
<tr>
<td>Hinofotis &amp; Bailey</td>
<td>1981</td>
<td>P</td>
<td>Vocabulary, grammar, eye contact, clarity of expression, ability to relate to students, development of explanation, and others</td>
<td>Yes</td>
<td>Videotaping, role playing, seminars</td>
</tr>
<tr>
<td>Hockett</td>
<td>1972</td>
<td>P</td>
<td>Less teacher control, more individual interaction, more high level questioning</td>
<td>Yes</td>
<td>Seminars</td>
</tr>
<tr>
<td>Kanaga</td>
<td>1979</td>
<td>T</td>
<td>Effective leadership in small group processes</td>
<td>I</td>
<td>Role playing</td>
</tr>
<tr>
<td>Keye</td>
<td>1980</td>
<td>P</td>
<td>Foreign TA language skills</td>
<td>Yes</td>
<td>Classroom presentations</td>
</tr>
<tr>
<td>Koffman</td>
<td>1974</td>
<td>Q</td>
<td>Clarity, evaluation and feedback, relating to student responses</td>
<td>I</td>
<td>Videotaping with consultation</td>
</tr>
<tr>
<td>Lewis &amp; Orvis</td>
<td>1973</td>
<td>Q</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Manteuffel &amp; Von Blum</td>
<td>1979</td>
<td>P</td>
<td>Knowledge of course material, communication skills, stimulation of student interest, overall ratings</td>
<td>NT</td>
<td>Workshop, seminars, classroom observation and/or videotaping</td>
</tr>
<tr>
<td>Prentice</td>
<td></td>
<td>P</td>
<td></td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>Tubbs</td>
<td>1975</td>
<td>T</td>
<td>General teaching ability</td>
<td>Yes</td>
<td>Workshop</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>----</td>
<td>--------------------------</td>
<td>-----</td>
<td>----------</td>
</tr>
<tr>
<td>Wortman &amp; Willis</td>
<td>1976</td>
<td>P</td>
<td>Organization, effective communication, genuine interest in helping students, fair grading</td>
<td>NT</td>
<td>Seminars, course preparation with consultation</td>
</tr>
<tr>
<td>Yaghlian</td>
<td>1972</td>
<td>Q</td>
<td>Interpersonal skills</td>
<td>Yes</td>
<td>Workshops, consultation while teaching</td>
</tr>
</tbody>
</table>

Key to experimental designs (Campbell & Stanley, 1963): P = pre-experimental design; Q = quasi-experimental design; T = true-experimental design. NT indicates that data were not tested for statistical significance. I indicates that the findings were inconclusive.
1. Change in TA Knowledge

There appears to have been little research in the area of change in TA knowledge as a result of training. Apparently only one quasi-experimental study has directly measured cognitive outcomes related to TAs. Participant bias and possible interaction effects may make the validity of the findings questionable. Anderson (1973) found treatment TAs making significantly different cognitive gains over a control group in familiarity with the use of behavioral objectives and statistical procedures after a training program over a fifteen week period.

2. Change in TA Behavior

As may be seen from Table 1 strong support may be found for behavior change observed by non-student observers. Fourteen of the studies reported statistically significant change, three did not test for change but concluded change to be found, and only one (Haber, 1973) found no statistically significant change in TA behavior. Rodriguez (1982) found that significant differences observed after seven days for more verbal reinforcement and use of summaries had disappeared after thirty days.

The significance level set by Murphy (1972) at 0.10 yielded significant differences on measures of Flanders Interaction Analysis Categories (PIAC). A similar study
(Carroll, 1976, 1977) of nineteen novice psychology TAs randomly assigned to a training or a control group also used FIAC. At the 0.06 level the treatment group had significantly higher Indirect/Direct ratios. That is, the teaching of the treatment group TAs was more student-centered than that of the control group TAs. Supportive findings were made by Tubb (1975) for eight calculus TAs and Daniels (1970) for eight mathematics TAs using FIAC. Dalgaard (1976) reported significant differences from observations of trained raters using the Teacher Performance Appraisal Scale. Only one study (Haber) failed to find significant difference between groups receiving (a) feedback on classroom performance and instruction in effective questioning techniques, (b) feedback but no instruction, and (c) no-treatment.

In a study utilizing three experimental treatment groups and a control group (Bray & Howard, 1980) it was reported that self-reported changes in TA behavior were supported by observer ratings. Trained TAs improved significantly in their skill, rapport with students, and overall teaching ability. Analysis of FIAC ratios found that training enabled TAs to adopt a more interactive mode of teaching using praise, making use of student ideas, and utilization of questioning, rather than use of lecturing and criticism.
Sharp (1981) randomly assigned thirty-seven TAs representing seven university departments to treatment or control groups. Participation was voluntary and all TAs had less than one year of teaching experience and no formal teacher training. The treatment group took part in a workshop during which effective lecturing was modelled. Pre-test and posttest consisted of microteaching lessons taught by all TAs and rated by trained observers. Modelling was found to produce a significant effect on the acquisition of modelled behaviors.

3. Change in Student Ratings of TAs' Behavior

Five of the eight pre-experimental and one of the five quasi-experimental studies of TA training failed to test for statistically significant changes in TA behavior as measured by student ratings. The remainder reported significant change though the findings of Hoffman (1974) were inconclusive. Although student ratings showed a general improvement, it must be remembered that these studies used at best either a pre-post one group design or non-equivalent control groups. Without random assignment of TAs, initial bias between treatment and control groups cannot be discounted and the TAs entering the treatment group by self-selection or similar means may do so as a result of a greater interest in or experience of teaching. If pretesting did not expose significant differences between groups,
the significant differences on post-tests could still be attributed to bias in the group compositions. Extraneous (uncontrolled) variables peculiar to participants pose threats to internal validity in such a design (e.g., prior experience, spoken English level, attitudes toward teaching).

Two of the true-experimental design studies did not produce statistically significant difference in student ratings of the TAs. Goodwin (1982), who had reported significant change as rated by non-student observers, was not able to report significant differences for student ratings. Costin (1968) was only able to find significant difference for student-teacher rapport, not for skill, structure, feedback, or interaction. Carroll (1976, 1977) found significant improvement for TAs' overall effectiveness and for agreement between course objectives and tests but not for agreement between objectives and the TAs' teaching.

4. Change in the Achievement of TAs' Students

Two quasi-experimental studies offer conflicting evidence for the effectiveness of TA training as measured by student achievement. Koffman (1974) assigned thirteen TAs in English to a control group and two treatment groups. All received feedback from instructional specialists but only one group received continuous assistance toward improvement
from these specialists. Non-significant differences in student achievement were found over an eight week period for the assisted group over the others. In a comparison of student performance from fall to winter for the students of seven economics TAs in a training program performance was found to increase as a result of the program (Lewis & Orvis, 1973).

Four true-experimental studies (Arbes & Kitchener, 1974; Boeding & Vattano, 1976; Daniels, 1970; Tubb, 1975) are reported in the literature. Daniels assigned eight TAs randomly to two groups one of which received training in Flanders Interaction Analysis techniques (FIAC) without finding any significant difference in the achievement of the TAs' students during the quarter. He did find significant difference in student achievement in terms of the academic major of the TAs. Those TAs majoring in mathematics education achieved significantly higher than those majoring in mathematics. This seems to suggest the Flanders techniques above are not sufficient to bring about changes in student achievement or that an identified interest in teaching may account for differences.

In a non-training study using a three-treatment design McKeachie et al. (1980) found the effects of consultation concerning student ratings upon effectiveness as measured by student achievement to be inconclusive. In only one of
the four classes where teachers received consultation were measures of students' psychological thinking significantly higher. A similar finding was reported in a non-training study by Miller (1971) in which TAs were randomly assigned to one of three courses and then subdivided on the basis of having shown favorable or unfavorable attitude toward feedback so that each course had a feedback component and a no-feedback component for both unfavorably and favorably responding TAs. Only in one of the three courses was the student achievement significantly different (improved) for those TAs (both favorable and unfavorable concerning feedback) who received feedback during the semester.

Miller further reported that the mean final examination scores were higher for those students whose instructors received feedback from students' ratings during the semester than for those students whose instructors did not. Scrutiny of the data presented indicates evidence for this in only four of the six subgroups. Differences in student achievement were found on recall items for students of TAs trained in Flanders Interaction techniques or Polya's Heuristic teaching by Tubb (1975). Students of trained TAs scored significantly higher than students of TAs who received no training.

Utilizing Education Through Student Interaction (ETSI), a systematic approach to classroom discussion based on a
model developed by Hill (1969), and a Lecture Discussion Method (LDM), no significant differences were found in achievement in a comparison of students of undergraduate TAs trained in using the two methods (Boeding & Vattano, 1976).

Significant findings resulted from a training study (Arbes & Kitchener, 1974) utilizing four undergraduate facilitators for discussion groups in a psychology class. The review of assigned articles was led by the facilitators who were chosen from the previous class yet unaware "as to what aspects of the article the instructor believed important". The students of the groups with facilitators performed significantly better than students in the control group but only on part two of the final test which covered material specifically from the discussion groups.

In accord with the findings of Carroll (1980), the true-experimental design studies (Arbes & Kitchener; Boeding & Vattano; Daniels; Tubb) involving student achievement as a measure of training effectiveness tend to be narrower in topic than other studies. Also clear support for significant changes in student achievement is not forthcoming. Three studies (Arbes & Kitchener; Lewis & Orvis; Tubb) do support positive significant change, two (Boeding & Vattano; Daniels) do not and one (Koffman) is inconclusive.
While it is clear that programs of TA training can effect the learning of TAs' students the best means of assessing the effects of training is debatable. It now becomes appropriate to review literature related to such assessment.

**Literature Related to the Evaluation of Program Outcomes**

A number of possible approaches allow for the evaluation of the effective teaching of TAs and by doing so provide evaluation techniques for the effectiveness of the TA preparation program. Schulz (1980) suggested certain components for a "comprehensive evaluation program". The first involved the establishment of specific criteria on which the evaluation will be based. Without this specificity, particularly in self-reports from TAs taking a training program, the data represent little more than a "happiness index" with the associated problems of reliability (Levinson-Rose & Menges, 1981).

The TAs indicated that they had benefited substantially and that they would highly recommend training for future TAs. Data of this type indicate that the participants were probably well satisfied with the training programs, but these data alone do not yield much valid information on specific changes...that may have occurred as a result of training (Carroll, 1980, page 169).
Self-evaluation (e.g., the use of a rating scale to assess one's own abilities) has been used as a means of introducing students to the concept of more stringent evaluation by others (Manteuffel & Von Blum, 1979) or as a means of diagnosing TA needs (Staton-Spicer & Nyquist, 1979). Obviously self-assessment is not without merit but the specification of the instruments used and the purpose for which the data are collected play an important part in determining that merit.

1. Both immediate and delayed tests should be made, since without opportunity for continuing practice-with-feedback the postcourse level of skill mastery is likely to decline; (2) If participants' self-assessments are to be accurate, they should refer to specific behaviors, those behaviors should have been assessed during instruction, and participants should have had opportunity to compare their performance with an external criterion; (3) finally, if sessions aim at attitude change and if they are evaluated through participants' self-reports, the sessions should include exercise or discussions that ensure active encounter with a variety of views held by other members (Levinson-Rose & Henges, 1981, page 410).

Schulz (1980) recommended that audio or videotaped lessons should form the basis for self-evaluation and evaluation by peers. It was reported in a recent survey of inservice training programs for TAs of basic communication courses that few departments used any type of self-evaluation or peer evaluation (Yoder & Hugenberg, 1980). Schulz
also found little provision for self-evaluation using a standard response form for departments (9 of 167 responses; 53% response rate). Only four of the responding departments indicated self-evaluation and peer evaluation as effective supervision procedures.

Marsh & Overall (1979) investigated faculty perceptions of potential bias in student ratings. Although it was reported that faculty were skeptical about the accuracy of student ratings, the study offered evidence that student ratings are valid and relatively free of bias. This tends to be supported by Nevill, Ware, and Smith (1978) who found that students judged teaching assistants and faculty members in similar manner both in terms of the ratings themselves and the conceptual framework within which the decisions were made. Further validation for the efficacy of student evaluations is provided by Lambeth and Kosteki (1982) who found high correlation between student and TA ratings of faculty.

A study involving the influence on student ratings of TAs' traits found some factors related to evaluations of TA effectiveness (Bos, Zakrajsek, Wolf, & Stoll, 1980). Sex of the TA and discipline of undergraduate study tended not to influence student ratings. TAs holding a master's degree tended to be rated higher than those holding a bachelor's and TAs in their late twenties tended to be rated higher.
than those in their early twenties. The differences were highly significant (p at 0.000) on measures of organization and preparation but must be interpreted with caution in the case of age as the number in each cell was below ten (using on-way ANOVA). TAs with previous contract teaching experience in elementary or secondary school were consistently and significantly (at the 0.000 level) rated higher in effectiveness than those with no experience. Presumably those with contract teaching experience had undergone some form of teacher training prior to that experience. There was a significant trend for TAs in their second year of TA experience to be rated as more effective than those in their first or post-second year. Once again, numbers were small but there seems to be some evidence to support the notion of learning on the job followed by increasing staleness in the event of inservice support not being forthcoming. The literature generally does not offer any conclusive evidence in this respect.

The notion that more teaching experience ought to be related to higher measures of teaching competence is not documented by this study, nor does the literature support this premise. Instead there is some contradictory data which leaves the assumption open to debate. Centra (2) reported that the highest ratings occurred within an experience range of three to twelve years while Remmers (9) found that teachers with over eight years of experience were rated better than teachers with less than five years. Guthrie (7) and Hawkins and Stoops (6) found no relationship between teaching experience
and teaching effectiveness. Turner (10) on the other hand, noted that on teaching task performance teachers were most effective during the first three years of experience and then leveled off after that (Bos et al., 1980, page 185).

Using a Chi-square test for correlations of TA characteristics with student performance, Monts and Pickering (1981) concluded that TA background effects were significantly small to be not considered as having significant impact on student achievement in the laboratory. It was concluded that the TA's ability to motivate students had more effect on their achievement than the TA's experience, undergraduate school, age, or marital status. It was suggested that similarity in age and marital status between TA and student might contribute toward motivation. The number of subjects in the study, two TAs each with twenty students, for each of four years of comparison, was small. Whitely and Doyle (1979) in a study of the reliability and generalizability of student ratings concluded that most ratings are highly generalizable, showing no relationship to the background characteristics of student sex, ability, year in school, or whether the course was required.

It was further concluded (Whitely & Doyle) that only some ratings are related to student learning and that certain aspects of generalizability and validity vary with the instructor's role. The ratings of teaching assistants on
overall effectiveness correlated with how much their classes felt they had learned but were not correlated significantly with final examination scores. Student-perceived learning was also significantly related to ratings of TAs on items of "motivation of student interest". The data suggested a need to examine student-perceived and tested learning separately for the TAs but not for the faculty. Findings were seen to support the multifaceted nature of effective teaching although the similarity of specific factors for TAs and professors indicated that "the facets of effective teaching are consistent across different levels of instructors".

Marsh and Overall (1972) presented evidence for the existence of distinctive factors related to different dimensions of teaching quality. This study found that the same nine factors influenced ratings for TAs and faculty. These factors were: learning/value instructor enthusiasm; organization; group interaction; individual rapport; breadth of coverage; examination/grading; value of assignments; and workload difficulty.

Student ratings and instructor self-ratings were shown to have good convergent reliability in a study over a two semester period of seventeen TAs (Braskamp et al., 1979). The student ratings were more consistent over the two quarters than the self-ratings.
Student ratings were more consistent over the two semester period than were the self-ratings. Even though the student ratings of instructors improved for the spring semester, these ratings correlated highly with ratings from the fall semester. Consistency of student ratings has been found in previous studies (Costin, Greenough, & Menges, 1971; Feldman, 1977). Greater student stability can be expected since each student rating score represents the class mean, a composite rating which would have greater reliability than a single self-rating (Braskamp et al., 1979, page 304).

It was also found that student ratings of the instructor's skill was among the best predictors of final student examination success, although the only statistically significant predictor was that labelled "teacher control", that is, the instructor's organization of the lesson. The findings were found to support the validity of student ratings and also indicated that self-ratings can be a useful source of evidence. It was suggested that each source, student ratings, self-ratings, and student achievement, represent a different perspective on teacher effectiveness and needs to be carefully considered as to how it can contribute to the evaluation process. Hardy (1983), while finding overall TA ratings significantly related to certain TA abilities, was unable to find any relationship between student achievement and TA overall ability. Schulz (1980) suggested that mid-term ratings by students can provide formative feedback for instructors.
A number of reasons for the possible ineffectiveness of mid-term ratings in bringing about changes in instructor behavior have been suggested. These include failure to compare instructor ratings to the mean or median and hence establish some norm from which an instructor may assess his or her effectiveness and failure to provide consultative services for instructors. The latter refers to the need for feedback with instruction enabling the instructor to change his or her behavior. A final reason suggested involves the lack of specificity in the teaching dimensions addressed by the rating instrument. The instrument needs to target specific instructional behaviors in order to provide useful feedback. Such behaviors should be empirically related to effective teaching, of course.

Effective teaching, or effectiveness of TA preparation has also been evaluated in terms of student achievement as described previously. Two studies using measures of student achievement are worthy of note for their approach to the data collection.

Whitely and Doyle (1979) attempted to eliminate grading bias among eleven TAs by having each TA grade a single question on the examination across all 485 students. In a similar attempt to eliminate bias in another study the measures of student learning took the form of tests administered as part of a larger battery of tests (McKeachie et
These were given to all students in a setting independent of those used for grades by the instructor. This study found consistency over time (one year) for the relationship between ratings of teacher effectiveness and student achievement.

We have on previous occasions argued that course final examinations are not good measures of teaching effectiveness, even though in using them one makes the proper assumption that the ultimate criterion of teacher effectiveness is student learning. Unfortunately, if we are comparing two methods of instruction or two teachers, final examination performance is an insensitive method of doing so. Students are so highly motivated for grades that they may well compensate for any deficiencies in the instruction they have received by extra preparation for the final examination. Thus, as reviews of teaching research indicate (e.g., McKeachie & Kulik, 1975), it is unusual to find experiments comparing teaching methods in which significant difference on a final examination are revealed, except in cases in which the final examination has been contaminated by the inclusion of items on which only one group of students has had previous practice (McKeachie, Lin & Mendersson, 1978, page 352).

Miller (1971) has suggested that the general conclusion in the literature at that time of no significant relationship between student achievement and ratings of the teacher by students may have resulted from lack of control in the design of the studies. Consequently Miller was careful to utilize TAs of similar academic and teaching experience.
presenting similar course content to students assigned randomly to discussion sections of the same class. No significant differences were found in the end-of-semester TA ratings between TAs who did and did not receive midsemester feedback of ratings, nor was the academic performance of the students of the treatment TAs superior in two of the three courses. Nevertheless, the study serves to emphasize the importance of control of extraneous variables in the research design.

Another approach that has been used is to give students a common test the content of which has not been specifically taught to by the instructors. Such a test represent a measure unbiased with respect to the specific content emphasized by different teachers (McKeachie et al., 1980). Clearly, consideration needs to be given to potential bias when teaching effectiveness is measured by student achievement.

**Outcome Measures: Summary and Comment**

The conclusions that may be drawn from the above review generally agree with those found in two recent reviews of the research literature (Carroll, 1980; Levinson-Rose & Manges, 1981).

Concerning change in TA knowledge as a result of training, the findings from a single study by Anderson (1973)
are inconclusive. The control group used was non-equivalent and the data were in the form of self-report. Studies of matched or equivalent groups of participants and non-participants is a minimum requirement in order to preserve integrity in the research design (Carroll, 1980). Matching TA treatment and control groups or pairs may also serve to overcome a dilemma that is as much ethical as methodological (McGahie & Mathis, 1977). To randomly assign students who will become TAs to a treatment group and a control group, that is, one in which no instruction is given, is difficult to defend ethically. It would seem that some measure of TA knowledge as a result of training should be a pre-requisite for measures of performance following training, knowledge of the outcomes desired in a training program serves as a necessary but not sufficient condition for the desired performance in the classroom setting. The evaluation of trainee performance alone as a measure of the effectiveness of training is not of itself sufficiently valid given that some participants in the training may not have assimilated the knowledge needed to be able to perform successfully.

In comparing treatment and control groups the presence in the treatment group of subjects who have not assimilated knowledge from the training could prove a confounding variable particularly if the groups were compared using group
means for treatment and control groups. Steps need to be taken to establish if the treatment group members have assimilated the necessary knowledge. Unless the extent of the assimilation is established this confounding variable may make the comparison of treatment and control groups pointless.

Scrutiny of Table 1, above, leads one to concur with Carroll (1980) and Levinson-Rose and Menges (1981), that, in general the research indicates that training programs designed to bring about change in TA behavior are able to do so. Except for Rodriguez (1982) there is little evidence for either the permanence or impermanence of the behaviors of the trained TAs. A study over time of the frequency of use of such behaviors may prove to be enlightening. As both Levinson-Rose and Menges and Carroll indicate, a design using treatment and control groups either randomly constituted or consisting of equivalent subjects is to be recommended though infrequently encountered.

The results of training programs designed to bring about behavioral changes in TAs should be directly observable. The presence or absence of the desired behaviors is but one measure of the effectiveness of the treatment. Unless the changes in TA behavior can be associated with improvement in student outcomes the true value of a training program is not established (Cooper, 1975; Kanaga, 1979; McKeachie et
al., 1978, 1980). The question arises: Did the acquisition of the behaviors mean that the TAs were more effective teachers?

While the effectiveness of the intervention strategy employed in a given training program may be considered a function of the presence or absence of the desired behaviors it cannot be considered a measure of the effectiveness of instruction. The two most commonly reported means of assessing the impact of trained TAs upon the effectiveness of instruction is through student ratings and measures of student achievement.

Carroll (1980) has indicated that the studies in most true-experimental designs tended, unlike quasi-experimental designs, to concentrate on a narrower range of topics and that few of the former demonstrated significant differences on student variables. Of the five true-experimental design studies cited above three had a broad range of topics. Four of the studies reported significant changes in student ratings as did all four of the quasi-experimental design studies cited. It would seem reasonable to say then that there is some evidence that student ratings can provide a measure of the effectiveness of TA training. Of the nine studies, six experimental design studies related the training to student achievement, three of them reporting significant gains.
The literature is generally inconclusive on the effectiveness of training programs as measured by improved student achievement. The studies in this area (Arbes & Kitchener, 1974; Boeding & Vattano, 1976; Daniels, 1970; Koffman, 1974; Lewis & Orvis, 1973; Tubb, 1975) have tended to be narrow in their topics focusing on some particular training approach such as Flanders Interaction Analysis techniques (Daniels, 1970; Tubb, 1975), Education Through Student Interaction (Boeding & Vattano, 1976) or feedback (Koffman, 1974) rather than the broader, more general topics dealing with effective teaching, itself a somewhat nebulous concept.

Finally, subject to temporal constraints, there is considerable merit in a follow-up approach to student achievement. Where courses taught by TAs are prerequisite for other courses, as is often the case with many service courses in various departments, a random selection of students from these courses might be made. Students from the groups of both treatment and control group TAs would be required, with attention to the equivalence of the group so formed.

The achievement of these randomly selected students in subsequent courses could then be compared within the particular courses. Over a period of time these data may prove more illuminating than comparison of group means and end of term examinations in the courses taught by TAs. The
use of group means for comparison requires a large number of groups, preferably of equal numbers. End-of-term examinations for between-group comparisons are open to bias as described earlier.

If effective teaching is to be measured in terms of student achievement then steps need to taken in the research design to reduce bias in the gathering of data (McKeachie et al., 1978, 1980; Whitely & Doyle, 1979). The multifaceted nature of teacher effectiveness must be recognized (Whitely & Doyle), so there is a need to identify specific behaviors where these exist (Schulz, 1980) that are related to the concept of effective teaching and investigate their relationship to student achievement. The identification of such behaviors is the purpose of the next section of this literature review.

**Literature Related to Training Program Content**

Ervin and Huyskens (1982) surveyed the perceived needs of 204 TAs and 99 faculty teaching foreign languages at four universities. The participants responded to a total of twenty-nine items, "key priorities for training", identified earlier as a result of statements collected from a group of 75 TAs. Considerable agreement was found as to what was most important in a training program and what was least important.

For inexperienced TAs the most important priorities were:
1. Gaining experience and/or self-confidence
2. Learning practical teaching methods and techniques (also rank second by experienced TAs)
3. Making class interesting (ranked fifth experienced TAs)
4. Making best use of class time (also ranked fifth by experienced TAs)
5. Making presentations
6. Lesson organization and planning

Experienced TAs, except where noted above, tended to focus more on the skills specific to the subject matter. For inexperienced TAs least important (selected) priorities were (commencing with the least):

29. Improving my command of English (same for experienced TAs) (sixty-two percent of the sample were non-native English speakers)
28. Dealing with shy, slow, uncooperative students (twenty-second place for experienced TAs)
25. Teaching adult beginners (twenty-sixth for experienced TAs)
23. Learning about my students and relating to them (twenty-first place for experienced TAs)

In follow-up interviews of fourteen TAs at two universities, perceptions of the most important topics in TA training focussed on: needing and using specific methods, tech-
niques and drills; getting to know the surroundings in terms of texts, other TAs, professors' styles and expectations; seeing others teach; lesson planning; presenting grammar; and teaching culture. TAs reasserted that topics relating to the reality of their situation were of most importance to them (Ervin & Muyskens, 1982, page 342).

A number of studies have attempted to identify teaching behaviors that constitute teaching methods and techniques have been identified in a number of studies. In a study of foreign TAs there was considerable agreement between undergraduate student raters rating TAs' performance in role-play situations before and after training and ratings by TA trainers (Hinofotis & Bailey, 1980). The importance of good pronunciation, the development of explanation, the flow of speech, and clarity of expression were rated highly by both groups. By comparison delivery items such as eye contact, other nonverbal aspects, confidence in manner and presence were ranked much lower.

The question of how important these delivery items might be as behaviors aiding the communication of information was not addressed although in responding to a strongly-agree/strongly-disagree part of the questionnaire, students and trainers agreed on the importance of eye contact when talking directly to students. Strong agreement was also expressed for a TA being able to paraphrase ideas or use syn-
onyms to explain technical concepts or assignments and even stronger agreement was expressed for TAs giving examples and illustrations without being asked. These last two may be considered aids to clarity which is supported at least by undergraduates rejecting the contention that "a TA who is interesting but disorganized is better than one who is boring but clear" (Binofotis & Bailey).

In a study of overt teaching behaviors and their relation to TA effectiveness as measured by student ratings, Braskamp et al. (1979) found high correlation for clarity and rapport items ($r = 0.65$ and $0.64$, $p < .001$). Rapport, defined as the quality of interpersonal relations between teacher and students was characterized by: addressing students by name; showing concern for student progress; showing strong interest in the subject matter; giving preliminary overviews; praising good student ideas; using graphs etc. to facilitate the explanation of concepts; soliciting student questions and comments; stressing important points; using eye contact. Telling jokes, stating one's own views, and using gestures were not found to be related to effective teaching as measured by student ratings. As the average class size of the study was below twenty it may be that attention-getting behaviors like these may have greater importance for large class teaching (Braskamp, et al.).
Clarity, defined as the ability to explain concepts or principles in a clear, straightforward way, was associated most strongly with the following behaviors: giving concrete examples of abstract principles; giving several examples of each concept; giving a preliminary overview of each lecture; speaking expressively or dramatically (Braskamp et al.).

Considerable interest has been shown in the clarity variable over the past twenty years. Teacher clarity found a prominent place in the mind of the educational community when Rosenshine and Furst (1971) placed it at the top of a list of the five most promising variables related to effective teaching. Their findings were based on a review of ten studies over the period 1963-1970. Subsequently, some doubt was cast upon the validity of the findings by Heath and Nielson (1974) who suggested that variables in a number of the studies reviewed had been assigned the label "clarity" unjustifiably. It was further suggested that the claim by Rosenshine and Furst of a significant relationship between the clarity variable and at least one criterion measure of pupil gain for seven of the studies was not supported in the original data by five of those studies. In recently re-examining six of these seven studies with colleagues the writer must agree with the findings of Heath and Nielson regarding significance claims. Only in four of
the ten studies reported did the original investigators report significance and these, along with several others, were not direct studies of clarity.

It is not therefore possible to say that the early studies, all of which were correlational, showed a certain link between teacher clarity and student achievement. Little was learned from these studies in terms of knowledge of what constitutes teacher clarity. Nevertheless, the early studies and the review by Rosenshine and Furst provided the impetus for research on this variable.

At The Ohio State University studies commenced in 1974 attempting to identify the lower-inference (easily observable) teacher behaviors associated with clear teachers as perceived by their students (Cruickshank, Myers, & Boenjak, 1975). This led to an empirical investigation of teacher clarity (Bush, Kennedy, & Cruickshank, 1977) using a set of 110 relatively low-inference behaviors compiled from the previous study. Results from this study suggested that a number of specific behaviors could be labeled as prime discriminators between clear and unclear teachers. The 1977 study took place at the junior high school level in Cleveland and was later successfully replicated with improved instrumentation and design in Ohio, Tennessee, and Australia using a total of 1263 students (Kennedy, Cruickshank, Bush, & Myers, 1978). These studies moved the con-
struct of clarity closer to a point at which an operational definition could be made and toward identifying the critical lower-inference behaviors associated with the construct.

Gloeckner (1983), using an experimental design, was able to show that teacher clarity behaviors can be taught. The treatment group receiving the training in specific low-inference behaviors significantly out-performed the control group in the use of those behaviors. Gloeckner was unable to find a significant difference in the achievement of the students in the classes taught by the trained and untrained teachers who were preservice teacher education majors at The Ohio State University.

Williams (1983), also using preservice education majors, found support for the stability over time in terms of frequency of occurrence of the majority of the teacher clarity behaviors identified below in a study by Hines (1981). That is to say, the frequency of use of high-inference, intermediate-inference, lower-inference measures of clarity was stable over time and different in terms of usage for teachers considered to be high, intermediate or low on clarity. Given the small n for clear teachers (n=11) and unclear teachers (n=4) Williams did not perform analyses attempting to relate the level of use of behaviors to student achievement.
Brophy (1979, (a), (b), (c)) found support in numerous studies for the effective teacher being one who could organize and structure a lesson well. Brophy concludes that "in general, students taught with a structured curriculum do better than those taught with more individualized or discovery learning approaches, and those that received much of their instruction directly from the teacher do better than those expected to learn on their own or from one another" (Brophy, 1979, (a), (b), (c)). Teachers who were more effective in their teaching were found to be more thorough and systematic in their teaching (Brophy, 1979 (b)). The studies reviewed were primarily not at the college level. Nevertheless, one implication made was that findings concerning such instruction do generalize to higher grade levels and different kinds of students but only to the extent that basic skill mastery is the primary goal (Brophy, 1979 (b)).

An evaluation of TA performance using responses from 2,258 undergraduates (Meredith & Bub, 1977) identified three major factors. Factor one, relational impact, included: a sense of humor by the TA; the encouragement of discussion; apparent interest in the students; tolerance of different opinions; stimulating a sense of challenge; insight and discovery; being easy to talk to outside of class; effectiveness in leading discussions; having speech
and mannerisms free from distractions; and raising challenging questions or problems for discussions.

Factor two, classroom organization, included: the TA being well organized and prepared, making clear the aim of each lab/discussion section; clarifying assigned readings; being helpful in reviewing tests; responding helpfully to content questions; and being able to explain difficult concepts in terms understandable by the students. Factor three, learning outcomes assessed by student responses to questions concerning perceived cognitive and effective growth, and factor two were both related statistically ($r=0.76$ and $0.65$ respectively, $p<0.01$) to relational impact which "appeared to measure the interpersonal relationship between the TA and the students" (Meredith & Bub, 1977).

Hurray (1983) reported that "university lecturers receiving low, medium, and high ratings from students differed significantly in the frequency with which they exhibited various specific, low-inference teaching behaviors. Three factors which differed significantly across all groups were Clarity, Enthusiasm, and Rapport. Significant group differences were found in areas which included expressive speaking, showing strong interest in the subject, moving about while lecturing, using humor, showing facial expressions, using graphs and diagrams, stressing important
points, giving multiple examples, signalling transition points, asking questions of students, addressing students by name, offering help with problems, avoiding eye contact (inverse relationship), giving preliminary overviews of the lecture, and encouraging questions and comments. Murray suggests that a certain level of teacher clarity may be a prerequisite for being perceived as enthusiastic (page 141).

Murray also noted that several of the behaviors cited above were identified in other studies. Mintzes (1979) identified speaking expressively, giving a preliminary overview, using multiple examples, and addressing students by name as being significantly correlated with instructor ratings. Tom and Cushman (1975) found significant correlation between student self-ratings of learning and twenty-eight low-inference teacher behaviors including teacher vocal expressiveness, signaling of topic transition, and use of real-life examples. Non-significant correlations were found for movement and gesture and giving a preliminary overview. Cranton and Hillgartner (1981) also reported teacher questioning, praising students, and elaboration of student responses to be related to instructor ratings.

In a landmark study Hines (1981) investigated the relationship between teacher clarity and student achievement and satisfaction. The study suggested that teacher clarity
bore a substantial and significant relationship to student achievement and that it is:

a meaningful multi-dimensional construct, which can be operationalized in terms of specific lower-inference teacher behaviors, generalizable across at least two educational levels—the junior high school and undergraduate college level (page 330).

The three underlying behavioral dimensions referred to by (Hines, page 330) are reproduced below with their respective primary, lower-inference teacher behaviors associated with each dimension of clarity.

A. Provides for Student Understanding and Assimilation of Instructional Content:
1. Answers students' questions;
2. Asks questions to find out if students understand;
3. Repeats things when students do not understand;
4. Explains something and then stops so that students can ask questions;
5. Explains things simply;
6. Teaches at a pace appropriate to the topic and to students;
7. Gives students enough time for practice;
8. Gives students a chance to think about what has been taught;
9. Stays with the topic until students understand;
10. Shows similarities and differences between things;
B. Explains/Demonstrates How to Do the Work by Use of Examples:
1. Uses examples when explaining;
2. Works examples (e.g., on the chalkboard) and explains them;
3. Shows students how to do the work by use of examples;
4. Teaches step-by-step;

C. Structures Instruction and Instructional Content/Presents Content in a Logical Sequence:
1. Presents content in a logical manner;
2. Points out what is important for students to learn;
3. Informs students of course/lesson objectives;
4. Tells students what they are expected to know or should be able to do on completion of instruction;
5. Summarizes the material presented in class;
6. Distributes time adequately over the topics covered during the course.

Behaviors discriminating between clear and unclear teachers were identified by students in the college setting. Hines then used a clinical setting to determine the strength of the relationship among teacher clarity measures and student achievement and satisfaction. Preservice teachers taught videotaped lessons to small groups of approximately five peers. Using the videotapes, observers recorded the frequency of occurrence of the lower-inference clarity behav-
iors. The results showed that all clarity measure (intermediate and high as well as lower-inference) across all sources of measurement (observers, learners, and self-ratings) were significantly related (p at 0.05) to both student achievement and satisfaction.

Path analyses were performed to establish if the effects of clarity were mediated by student perceptions. The results showed that student perception of teacher clarity affected the students' degree of satisfaction with the teaching. Furthermore, students taught by teachers who were observed by external observers to be clear teachers showed higher levels of achievement.

It was reported that the relationship between teacher clarity and student satisfaction was "essentially completely explained through students' perception of clarity". This was interpreted to be that students would tend to feel more satisfied with their learning if they perceived their teacher to be clear. Students' perception of clarity was found to moderately mediate the effect of clarity on student achievement. It was therefore concluded that:

if a teacher exhibited clear teaching behavior during instruction, these results suggest that it was possible for students to achieve, although their perception of the level of teacher clarity exhibited might not necessarily be congruent with the actual level of clarity behavior exhibited by the teacher (Hines, 1981, page 336).
The importance of the quality and quantity of the low-inference teacher clarity behaviors was discussed. The frequency of the behaviors was not in itself found to be critical in forming an accurate picture of clear teaching behavior but the quality of certain behaviors was. These included teachers answering students where the frequency was less a critical factor than the "appropriateness or relevance of the answers" and the relevance of the examples the teacher used in explaining as well as the frequency with which they were used. Some behaviors of necessity occurred with limited frequency, for instance, provision of a summary of the material presented in the class but these "are not necessarily less important to clear teaching than others that do not occur very frequently" (Hines, page 332).

It is perhaps through the quality of the interaction with students that teachers facilitate the understanding of broad concepts rather than the ability merely to deal with a specific problem. Tubb (1975) has shown a relationship to exist between teacher questioning techniques and student problem solving ability. Concerns have been expressed in diverse fields and at different levels of college instruction that the provision for conceptualization is crucial in instruction (Damarin & West, 1979; Peters, 1981; Beif, 1981).
The Hines study also provided support for the role of organization as cited earlier (Brophy, 1979 (a), (b), (c); Hinoftis & Bailey, 1980; Meredith & Bub, 1977) in perceptions of effective teaching as measured by student ratings of achievement. Teacher clarity in conjunction with organization and student opportunity for learning criteria material was found to bear an even stronger relationship to student achievement and satisfaction than was observed with teacher clarity independently (Hines, 1981). The provision of opportunity for the student to learn the required content was shown to account for a "sizable" amount of additional variance in the student achievement and satisfaction after teacher clarity had been partialled out.

Finally, it was established that the perceived adequacy of teacher preparation was a very strong discriminator between clear and unclear teachers. Hines concluded that "teachers perceived to be more clear were also perceived to be more adequately prepared for teaching by observers, students and teachers" (page, 338). It was suggested that the more adequately prepared teacher presented a more organized lesson during which the teaching was more clear than the less adequately prepared teacher.

The conclusion reached by Hines may come as no surprise to a good number of the members of the teaching profession. The importance of the findings lies in the large step that
has been taken toward the provision of principles of teaching that have general application. These principles are sometimes referred to as generic and an example is given below:

If teachers study and plan materials to be taught in order to reduce vagueness of presentation, student achievement will increase. This is true when the teacher is the only source of information about the topic—as in a lecture on a special area (Smith, 1980, page 93).

Brophy (1979 (c)), on the other hand, has suggested that there does not appear to be any universal teaching competencies that are appropriate in any and all teaching circumstances and that there may only be a few truly universal instructional principles. Recognizing that context (grade level, subject matter, student characteristics, etc.) is important Gage (1978) suggested that research on effective teacher behavior can yield some main effects that will have general value for all teaching, whatever the grade level, subject matter, or character of the student. As Hines (1981) concluded regarding the operationally defined teacher clarity behaviors above:

It is expected that teachers who frequently exhibit these behaviors in instructional situations will be perceived by their students to be more clear. In addition, the findings also show that teacher clarity bears a substantial and significant relationship to student achievement and satisfaction (Hines, 1981, page 338).
Teacher Nonverbal Behavior

A recent review of research in education suggests that student perceptions of teacher effectiveness may be influenced by teacher nonverbal behavior.

In an observational study of teacher proximity, use of space, and student perception, Hesler (1972) found that students perceived college teachers who spent more time sitting on, behind, or beside their desks as less warm, friendly, and effective than teachers who spent more time away from the desk (Moolfolk & Brooks, 1983, page 117).

Close teacher behaviors seemed to be associated with warmth, friendliness, approval and liking in the few studies cited. As these studies dealt largely with students below high school grade generalization to higher levels of education may be inappropriate.

Studies have indicated that encouragement may take different forms depending on the ability level of the students (Brophy, 1979 (c)). Teacher effectiveness with high ability students seems to be associated with challenge and demand for superior work and occasional use of criticism plus maintenance of a rapid work pace. Effective teaching with low ability students seems to be associated with a less demanding, more encouraging atmosphere with frequent praise, less criticism and more wait time after questions and use of hints or rephrasing. This has some significance for the
teaching of freshman classes. In a study supporting the
validity of student ratings (Braskamp et al., 1980) it was
found that student ratings of teacher's "control" of the
class discussions was the only statistically significant
predictor of the class achievement. The author went on to
describe what may well be a situation typical of a TA led
college recitation class.

In this instructional setting in which
the students' academic performance and
thus their grades were primarily based
on their level of mastery of the objective exams prepared by someone else,
the instructor's ability to take charge
of the discussion, provide drill type
questions and engineer the discussion
to maximize student achievement of the
subject matter as outlined in the
course syllabus may be the best in-
ductional strategy (Braskamp et
al., 1979, page 305).

Brophy (1979 (c)) suggested that teacher vocal asser-
tiveness might play a role in effective class management.
There does not appear to be support at the college level
where impact upon students may be different although it has
been shown that adults rely more on facial expression and
voice tone to determine the meaning of incongruent communi-
cations. Children tend to attend to the words and voice
tone but less to facial expression for the same purpose.

Breed (1971) found that subtle differences in eye con-
tact had no significant impact on audience retention of
lecture material. More obvious behaviors such as giving
considerable or no eye contact did significantly influence retention for female students in small groups. This seems to support an earlier finding (Hinofotis & Bailey, 1980) that eye contact is important when talking to students, the implication being directly to the individual rather than the group.

In general, findings from studies dealing with proxemics suggest that students seated in the front and center of a classroom are likely to be most participative (Woolfolk & Brooks, 1983). Such students are perhaps better located for eye contact with the teacher and this kind of visual attention and involvement may influence the student behavior. It has been shown that students so located look at the teacher more and write more. Such students have also been associated with positive teacher impressions of students (Woolfolk & Brooks, 1983). Moderate or low verbalizers among students have been shown to avoid these locations and teachers tend to exhibit a less permissive and interactive verbal style with these students, communication being more one-way (Brooks, Silvern & Wooten, 1978).

Content of Studies Involving TAs

Support for the importance of a number of the concepts identified above—rapport, organization, use of time, reinforcement, non-verbal communication, and clarity of in-
struction—may be found in a number of studies related to TA preparation. These are reviewed briefly below.

Bray and Howard (1980), using a true-experimental design, reported that trained TAs improved significantly in their skill, rapport with students, and overall teaching ability. These TAs were further found to make more use of praise, of student ideas, and questioning rather than lecturing and criticizing. It was suggested that the TAs developed a more indirect or student-centered style of teaching "although student ratings of teaching evidenced positive improvements, dramatic changes were not found overall" (page 68). No attempt was made to associate changes in TA behavior with student achievement.

In a correlational study by Damarin & West (1979) the teaching style of a group of foreign mathematics TAs most highly correlated with the program overall grade. Teaching style was defined by: organization; ability to involve students in discussion and problem solving; forcefulness in presenting new material; and "related abilities". Low grades were attributed among other things to TAs': lack of clarity; tendency to lecture over the heads of the audience; and failure to pay attention to the needs and questions of the class.
Support for change toward more student-centered teaching by trained TAs was found at the 0.06 level in a true-experimental study (Carroll, 1977) as was improved student ratings of the trained TAs at the 0.10 level. The use of indirect teaching skills was positively correlated (r=0.53, p at 0.02) with student ratings of effectiveness. No changes in student achievement were measured. The above program (Carroll, 1977) also emphasized active listening skills, establishment of a conducive classroom atmosphere, awareness of nonverbal communication, and provision of closure.

A quasi-experimental study by Anderson (1973) supported the acquisition of interactive behaviors by trained TAs. The treatment group TAs demonstrated significantly more student response, teacher questions, positive reinforcement, use of silence and less teacher lecture but not significantly more use of student ideas.

As may be seen from Table 1 above, a number of empirical studies support the inclusion of interactive skills designed to increase student participation. Descriptive studies (Bailey, 1977; Barrus et al., 1974; Mozer & Kappell, 1974; Spears & Zollman, 1974) also offer support. Specific skills include reinforcement, the use of silence, vocal variation, eye contact, and gestures.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Design</th>
<th>Construct</th>
<th>Measure</th>
<th>Related Variable</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Braskamp</td>
<td>1979</td>
<td>C</td>
<td>Classroom Organization</td>
<td>Student ratings of TA &quot;teacher control&quot;</td>
<td>Student achievement</td>
<td>Yes</td>
</tr>
<tr>
<td>Bray &amp; Howard</td>
<td>1980</td>
<td>T</td>
<td>Rapport</td>
<td>Observer ratings of TA behavior</td>
<td>TA training</td>
<td>Yes</td>
</tr>
<tr>
<td>Gloeckner</td>
<td>1983</td>
<td>T</td>
<td>Clarity</td>
<td>Observer ratings of TA behavior</td>
<td>Teacher training</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Student achievement</td>
<td>No</td>
</tr>
<tr>
<td>Hines</td>
<td>1981</td>
<td>T</td>
<td>Clarity</td>
<td>Observer, student, and teaching ratings of teacher behavior</td>
<td>Student satisfaction and student achievement</td>
<td>Yes</td>
</tr>
<tr>
<td>Hino&amp; Bailey</td>
<td>1981</td>
<td>P</td>
<td>Clarity</td>
<td>Observer and student ratings of TA behavior</td>
<td>TA training</td>
<td>Yes</td>
</tr>
<tr>
<td>Mintzes</td>
<td>1979</td>
<td>C</td>
<td>Clarity</td>
<td>Student ratings of TA behavior</td>
<td>Student ratings of TA effectiveness</td>
<td>Yes</td>
</tr>
<tr>
<td>Meredith &amp; Rub</td>
<td>1977</td>
<td>C</td>
<td>Relational impact</td>
<td>Student ratings of TA behavior</td>
<td>Student achievement</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Classroom Organization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murray</td>
<td>1983</td>
<td>C</td>
<td>Clarity</td>
<td>Observer ratings of teacher behavior</td>
<td>Student ratings of overall effectiveness</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Key to experimental designs (Campbell & Stanley, 1963): P = pre-experimental design; C = correlational study; T = true-experimental design.
Table 3 summarizes a number of college level studies addressing specific constructs identified above. As may be seen, the clarity construct has received more than a little attention and appears to bear a significant relationship to instructor effectiveness. Rapport, perhaps synonymous with Meredith and Bub's (1977) "relational impact" also receives some support for being related to instructor effectiveness.

Support for the twenty lower-inference teacher clarity behaviors cited above (Hines, 1981) to be related to student achievement is suggested by the summarized findings in Table 3. Weaknesses in design, as identified by Gloeckner (1983), possibly precluded a significant relationship being found in the Gloeckner study.

The content of any training program should be chosen in order to meet the needs of the program participants. In the case of would-be TAs some clearly perceived needs emerge. These include exposure to techniques and practices that will equip the new TA for survival in the classroom. How to organize and plan a lesson, and the development of self-confidence through practice situations appear of primary importance (Ervin & Muyskens, 1982). More experienced TAs are likely to express more concern for the development
of their students and less for the acquisition of "how-to" behaviors than inexperienced TAs. Such findings are in accord with the developmental levels of teaching as delineated by Fuller (1969).

Student-teacher rapport-building behaviors such as the use of eye contact, addressing students by name, and encouraging and praising student contributions are identified generally as desired TA behaviors. For foreign TAs, clear pronunciation, flow of speech, and general clarity of expression seem to be of paramount importance.

Few studies have investigated the constructs identified in Table 3 at the college level. Hines (1981) concluded that the behaviors she described will be exhibited by teachers who are perceived as more clear at the college level. This suggests that TAs specifically trained in these behaviors should be more effective in the classroom and that this effectiveness may be measured by gain in the achievement of their students. To date no such study appears to have been undertaken.

**Foreign Teaching Assistants**

To judge from the number of training programs designed to prepare foreign TAs (FTAs) there appears to be a heightened awareness of concerns in this area. Prior to 1979 no
programs specifically targeting FTA training were found in this review.

Between that date and 1982 nine were identified, three reporting empirical data and the remainder descriptive in nature. In addition, eleven FTA-related non-training program studies were identified.

There has been a marked increase in the enrollment of foreign graduate students over the last decade in U.S. colleges (Thomas & Richardson, 1978). The Institute of International Education Statistics reported in 1979 that the number of foreign students enrolled was almost a quarter of a million. Many of these students support themselves as TAs in basic undergraduate courses (Franck & De Sousa, 1982). The Ohio State University employs almost two thousand TAs. Approximately twenty-five percent are foreign TAs, the majority from countries where English is not the first language. While the percentage of FTAs varies from department to department it can be quite high; in Mathematics and Statistics the percentage exceeds sixty (Office of the Vice President for Personnel Services, O.S.U., 1984).

As Hood & Beardon-Anderson (1979) have pointed out, the increasing number of FTAs at U.S. universities provides opportunities for U.S. students to benefit from contact with
people from other cultures. They further stated that FTAs may, as a result of their cultural background, not share U.S. attitudes toward teaching and that this may be an inhibiting factor in their ability to communicate in the classroom. Sadow and Maxwell (1983) stated the position very clearly:

Trained in autocratic systems where personal contact between professor and student is the exception rather than the norm, they are ill-prepared for the constant and often intense interaction between teacher and student found in the American discussion section; problem-set session, laboratory, or review class. Ironically, as teaching assistants (TAs), these graduate students will rarely be called upon to lecture, a task which many could handle adequately (Sadow & Maxwell, 1983, page 250).

It is contended by Sadow and Maxwell that unless the foreign TA has an understanding of the dynamics of the American university class, his or her chances for success are severely limited (page 250). There is often an implied requirement for foreign TAs firstly to prove themselves and secondly to adjust to U.S. ways of behaving and thinking (Mestenhauser, 1981). Clearly, there are socio-cultural and academic differences in the university system that may cause communication breakdowns in the classroom (Blizzard, Roy, & Hogan, 1981; Damarin & West, 1979; Dege, 1981; Meleis, 1982; Rice, 1979). The question arises as to the value of exploring in any depth these differences with foreign TAs.
Confronting the issue of cultural differences is a risky business and creates denial, suspicion and unease. But pretending that the issue doesn't exist does not make it go away; it merely creates a "hidden agenda" which interferes with successful communication and (in the classroom) learning. (Dege, 1981, page 7).

The possible reason for American students' adverse reaction to foreign TAs has been explored. Westenhauser found that it was often unreasonably justified in that the TAs did not "speak English like we do", could not "behave objectively and rationally as U.S. students do", and could not be "understood" (when asked purposefully confusing and complex questions). Orth (1983), in a study of undergraduate reactions to foreign TAs, concluded the negative evaluations of TAs' speech often were based in social mythology rather than linguistic reality. It would appear that the problem, where extant, does not always rest with the TA.

"When I ask them, (my class) whether they have problems understanding me, all I get is blank faces and no answer," Doshi said. "Do they not care if they cannot understand what I say?" Doshi also said few of his students take advantage of office hours to ask questions or express problems. "I sit there during my office hours and few students come; they only bother to see me around midterm or final," Doshi said (The Lantern, 1982 (a); page 1).

Obviously, it is important to consider cultural factors other than language when preparing or appointing TAs. Nev
Nevertheless, it is language that is primarily focussed upon in cultural considerations.

Among a number of recommendations coming from a forum concerning TA preparation was one that would require all prospective TAs to be interviewed in person or over the telephone with emphasis on ability to communicate and proficiency in the English language (The Lantern, 1982 (c), page 4; The Graduate Voice, 1983 page 1,4). In a survey by Schulz (1980) of foreign language departments it was reported that of 167 departments responding not one required scores on a target-language proficiency test. Jones (1979) reported that the tests of English used at "a large Eastern university" did not measure speaking proficiency directly at all. Cake and Menasche (1982) described the testing procedure at the University of Pittsburgh as basically consisting of a structured interview using a rating instrument. On the basis of the ratings by several observers the prospective TA was assigned to received tutoring assistance, to take a class to improve the English level, or to receive no help if the level was satisfactory.

The question of testing prospective foreign TAs in English proficiency was addressed in a survey of fifty-one United States universities (Thomas & Richardson, 1978). These were public or private four year program institutions, with a total enrollment of 15,000 or above, a total
foreign enrollment of five hundred or greater, and a total foreign graduate enrollment of two hundred and fifty or more.

With a return rate of 62.8% it was found that the Test of English as a Foreign Language (TOEFL) was the most preferred measure of an individual's English proficiency. This measure is based on the individual's proficiency in written English and consequently some departments indicated interviews were given instead of TOEFL (twelve departments) or in combination with TOEFL (six departments). It was reported that individuals responding to the survey "felt it was of prime importance that an individual be able to effectively communicate in the classroom" (Thomas & Richardson). Warner (1982) reported finding TOEFL scores to be unrelated to communicative ability. Hardy (1983), in studying the relationship between 1,500 student test scores and TA evaluations, found the ratings of TAs to be related to their ability to communicate effectively, to stimulate interest, and their knowledge of the subject matter (p=0.05) as measured by student ratings. Bailey (1982) found 392 students surveyed to feel that their understanding of the subject matter was negatively influenced by low levels of English ability.

Hinofotis, Bailey, and Stern (1981) stated the oral English proficiency of foreign TAs to have been identified as
a major problem in undergraduate instruction. The level of proficiency in English does not necessarily reflect the level of knowledge in the subject area, of course.

Plenty of intelligent TAs have serious difficulty communicating with their students. Sometimes these people come from foreign countries and lack sufficient English skill. Others have difficulty conveying ideas because they understand the subject at a higher level and have problems explaining these concepts in simple terms (The Lantern, 1982 (c), page 4).

An implication here is that communication problems are not peculiar to foreign TAs but merely exacerbated by language deficiencies. A recent study of American undergraduates' reactions to the communication skills of foreign teaching assistants compared priorities of ten undergraduates with those of three teachers of English as a second language and three trainers of TAs. The undergraduate raters were drawn from mathematics classes because "courses in math and the sciences often employ non-native teaching assistants" (Hinofotis & Bailey, 1980). Considerable agreement was found. Pronunciation was rated as first priority by both sets of raters with the ability to develop an explanation ranked third by both closely followed by flow of speech. Clarity of expression ranked much lower for the undergraduate raters than the other raters who placed it second. The reverse was true for TAs' ability to relate to students.
Using a separate instrument supporting evidence was provided when both teacher/trainer and undergraduate groups strongly supported the notion that foreign students applying for TA positions should not be required to pass an oral English proficiency examination. Both groups also disagreed with a statement that non-native speakers of English should not be given TA positions. Halting speech was distracting to undergraduates but was not similarly considered by the other raters whose mean was "no opinion". Good pronunciation was considered more important than good vocabulary by both sets of raters.

An instrument designed to measure proficiency in spoken English has been reported (Clark & Swinton, 1980). This instrument, the Test of Spoken English (TSE), was used with sixty non-native English-speaking TAs in order to investigate its scores as predictors of student ratings of the TAs. It was found to be highly predictive of ratings of the TAs speaking proficiency in classroom lecture settings and in in-class question-answer situations, as well as of their communicative effectiveness in one-on-one conversational situations such as student-teacher interchanges in tutorial or laboratory sessions or in after-class or office-visit settings.

It may be quite strongly inferred from these analyses that administration of... the TSE to applicants for teaching assistant positions can provide appreciably greater prediction of their
probable communication performance in English-speaking situations associated with their instructional assignments than is available through biographical data concerning the nature and amount of their prior English study (Clark & Swinton, 1980, page 45).

If proficiency in spoken English is viewed as a major factor in the general concept of teacher effectiveness, then applying such a test to prospective TAs whether foreign or non-foreign could have potential for reducing problems encountered in the use of graduate students as instructors. It was reported that some degree of relationship was found between TSE scores and ratings of general teaching effectiveness, though not at the same high level of prediction as found for ratings of speaking proficiency (Clark & Swinton, 1980).

The potential importance of a TA's verbal ability is suggested in a study by Vecchio & Costin (1977) of TAs in psychology. It was observed that the perceived skill of the TA in the course most highly correlated with TA scores on the Graduate Record Examination—Verbal (see description of this test under Instrumentation), r=0.34 (non-significant). Step-wise regression analysis revealed that student achievement correlated most strongly with the TA's hours of psychology taken as an undergraduate (beta weight =0.736, significant at the 0.02 level) and the TA's score on the Graduate Record Examination—Advanced (see Instrumentation
for a description) (beta weight =0.595, significant at the 0.10 level).

The perceived TA skill was measured by responses from undergraduate students to items on a five point scale from "almost always occurred" to "almost never occurred". These items addressed: interesting delivery of material; stimulation of students' intellectual curiosity; clear teacher explanations; and teacher skill in observing student reactions. The fifth item, dealing with rating of overall teaching ability was coded from "outstanding and stimulating" to "poor and inadequate" on a five point scale.

Negative affect (instructor coldness toward students) was measured on a scale as described above. Responses to two questions addressing students' lack of involvement in or reaction to the class, and two questions addressing teacher lack of flexibility in evaluation of students' achievement and indications of unwillingness to discuss student statements measured this variable.

TA scores on the Graduate Record Examination—Verbal were found to correlate with two measures. The first (p=.10, r=-.33) on measures of TA negative affect indicated that the TA's coldness toward the students was related to lower TA verbal ability. Greater TA skill (p non-significant, r=.34) was related to greater TA scores on the GRE-Verbal.
TA skill and TA negative affect, taken together were seen to reflect the level of student satisfaction with the TA. Vecchio and Costin concluded that the GRE-Advanced score may be a predictor of student achievement and that GRE-Verbal may be a predictor of student satisfaction. Isaacson, McKeachie, and Mulholland (1963) define TA "culture" as "polished, artistic, imaginative, and effectively-intelligent". To the degree that the GRE-Verbal is indicative of such culture, Vecchio and Costin concluded that their study supported that of Isaacson et al. who found the instructor's "culture" most highly correlated with his/her overall effectiveness.

Training Programs for Foreign TAs

There have been twice as many programs for training foreign TAs reported this decade to date than in the previous decade. Unfortunately, very little empirical data is available supporting the effectiveness of the training. Himofotis and Bailey (1981) reported finding significant improvement in TAs' performance as rated by undergraduate raters ($p=0.005$) and teacher/TA instructor raters ($p=0.0005$) after a forty-hour program in oral communication.

Oral and listening skills were emphasized as well as presentation techniques and communication strategies in a
one-week orientation program described by Brinton and Gaskill (1979) which included TAs who were native Americans. Extensive use was made of videotaping experiences and one conclusion reached was that the program's greatest value lay in its opportunity for foreign TAs to meet other foreign TAs.

Concern that foreign TAs should interact more with native Americans was expressed by Cake and Menasche (1982) who used role playing and microteaching with feedback extensively in a one-term course which emphasized clear communication, fluency of speech, and a cultural orientation. As with Rice (1979) who described an orientation over one semester no data of any kind regarding the effectiveness of the programs were offered. Both Rice and Cake and Menasche emphasized socio-cultural and academic differences in the university system that may contribute to breakdown in communication in the classroom.

Cross-cultural teaching effectiveness was the focus of a forty hour program reported by Keye (1980) who found that training did affect the improvement in language skills, teaching effectiveness, and cross-cultural awareness of TAs in the U.S. classroom setting. Dege (1981), in sessions over a ten week period, focussed on the cross-cultural component. She reported that the provision of opportunity for the FTAs to discuss their individual problems was considered a highlight of the program.
The importance of immediate feedback and consultation was emphasized by Russo (1982) who used live demonstrations of teaching and coordinator observation of TAs teaching. Cake and Menasche (1982) also stressed the importance of feedback on FTA performance. Further they strongly suggested that the ability to speak English fluently is not the answer to producing effective TAs in the classroom as it is only one factor in socio-cultural inhibitors of clear communication. Rice (1979) similarly stated that a TAs linguistic ability is not the only factor in determining the ability to teach clearly.

Clear communication in English, cultural differences in teaching, the utilization of audiovisual teaching aids, and the use of videotaping experiences with immediate feedback and consultation were described by Franck and DeSousa (1982) as components of an FTA training program. While reporting the success of the program, it must be remembered that the data are in the form of participant self-report. Damarin and West (1979) identified a number of problems during FTA training. These included the FTAs' lack of clarity in the classroom and an "apparent inability to think on their feet in English". The authors concluded that many cultural factors other than language were related to the difficulties foreign TAs experienced in interacting with students in U.S. classrooms.
The transition from a society with a caste system or with a highly selective educational tracking system to an open-admission university must be a difficult one. While all the TAs in the experimental course adjusted remarkably well to the Ohio State situation, the difficulty of the transition was apparent in some (but not all) cases in a variety of ways: (a) difficulty in selecting and adhering to an appropriate level, (b) unrealistic expectations with respect to students' background knowledge as evidenced by presenting skeleton derivations and expecting students to fill in gaps mentally, (c) derogatory remarks to the instructors and among themselves concerning students' levels of achievement and motivation, and (d) occasional obvious anger at students for not understanding (Damarin & West, 1979, page 497).

Foreign TAs: Summary and Comment

It would appear to be erroneous to consider the problems peculiar to foreign TAs in U.S. classrooms as resulting purely or even largely from language difficulty. Oral proficiency in English is obviously important but constitutes only one factor in communicative ability that essentially has its bases in cultural background.

 Culturally related problems such as attitude toward students, teaching techniques in terms of a didactic or interactive approach, and expectations of student and instructor roles are no less important than those arising purely from linguistic difficulty. There are dangers inherent in con-
frenting the cultural difference issue but the consensus strongly supports including such a component in programs designed to prepare foreign teaching assistants for U.S. classrooms. If the emphasis is on what is culturally appropriate in a given context rather than categorizing cultural differences in teaching and learning styles as right or wrong, it would seem that TAs, both foreign and native to the U.S., could benefit from exposure to the issue.

The language-based part of the problem seems to focus upon clear pronunciation and on being able to clearly explain material. There is some evidence of a relationship between a TA's verbal ability and the TA's effectiveness as an instructor. Evidence also exists of a relationship between the TA's effectiveness and his or her negative affect or degree of coldness toward the students. This perceived coldness might be synonymous with poor rapport or act as an inhibitor of good rapport. This negative affect has been reported as a factor in student perceptions of satisfaction. Possibly the TA's affect has language ability as a prerequisite or it may be that knowledge of expected teaching style in U.S. classrooms could lead to improved affect. In other words, it may be a necessary but not sufficient condition for the TA's English to be of certain level. It may also require an attitudinal change toward teaching style.
If the above is correct, it becomes important to measure not only TAs' English ability, but the relationship between levels of ability and behavior in the classroom. Presently, the most frequently used measure of English ability is the Test of English as a Foreign Language (TOEFL). This written test does not measure the spoken English ability of the testee. The Test of Spoken English, like TOEFL, a product of the Educational Testing Service, would appear to provide a more accurate means of assessing an individual's ability in spoken English.

In addition to a cultural orientation, a component of TA training programs appearing most frequently is that designed to increase the communicative ability of the TA participants. Clarity of communication received some emphasis as did the use of videotaping with immediate feedback and consultation as an intervention strategy. Experience in the use of audiovisual teaching aids also received support for inclusion in FTA training programs.

Research into TAs levels of spoken English and their performance in the classroom is sparse. Only one study attempted to statistically relate TAs' spoken English level to measures of teaching effectiveness. The measures cited were student ratings rather than student achievement. There have apparently been no quasi-experimental or true-experimental designs used in research studies investigating the preparation of foreign TAs.
Conclusion to the Review of the Literature

The purpose of the preceding literature review was to address a number of key issues in the design and evaluation of a study concerning the preparation of foreign TAs. It has been suggested (Carroll, 1980) that the design of a training program for research should be an iterative process in which the results of previous research should be incorporated cumulatively in the design of the new program. Consequently the literature reviewed included: intervention strategies reported in studies of TA preparation and related literature; outcomes of TA training programs reported in studies of TA preparation; the evaluation of TA preparation programs; and both studies of TA preparation and related literature dealing with the content of training programs. As the outcomes to be researched were those resulting from a program designed specifically to prepare foreign TAs for teaching in U.S. classrooms, special focus was directed to the literature dealing with the concerns related to foreign TAs and their preparation.

In reviewing the literature concerning the training of TAs one can only echo the findings of Levinson-Rose and Manges (1981) that the quality of the material is not high. In fact, since the review by Carroll (1980) the quantity also seems to have diminished, Levinson-Rose and Manges re-
porting only one additional empirical study. Not including that study, the author has identified ten empirical studies over the period 1978-1983. Carroll (1980) cited twenty-four such studies up to and including 1977. He also stated that he had identified forty-eight descriptive studies. In the period 1978-1983 the author identified and examined twenty such studies along with nine prior to that period. The trend appears to be for less rather than more TA training studies, empirical or otherwise, to appear in the literature. Given the increasing concern for the level of instruction by TAs as expressed in Chapter I, this is somewhat surprising.

While the descriptive studies generally speak very favorably of the efficacy of TA training in bringing about desired changes in TA behavior and in increased teaching effectiveness they fail to present any empirical data in support of their findings. The few empirical studies on the other hand are far more conservative in their findings, often contradicting one another.

Nevertheless, there is evidence that TA training can produce changes in TA behavior and that these changes may be detected by observers and students. Few studies have attempted to relate or been successful in relating TA training to the achievement of students taught by the TAs.
Reviewing the related literature suggests certain lower-inference teacher clarity behaviors to be related to student achievement. Training in these behaviors could form the content for a TA training program. Additional behaviors associated with student-teacher rapport are also supported as potential content for such a program.

There is evidence that TA effectiveness in the classroom may be related to verbal ability in English for foreign TAs. The sex and background of TAs appear to have little mediating effect in student ratings except for background in the case of foreign TAs. Focus on the concerns and training of foreign TAs provides additional behaviors related to the use of audiovisual teaching aids, pronunciation, and voice level. There is also strong support for the inclusion of a cultural component in any training program for foreign TAs.

The literature generally supports the use of microteaching experiences with feedback as an effective intervention strategy. Support is also found for protocol materials in the form of locally-made videotapes as an effective way of bringing about behavior change particularly when such strategy is coupled with microteaching and discrimination training.
The research design needs to be carefully conceptualized in order to avoid the effect of potentially confounding variables. TA assimilation of behaviors at the cognitive level is no sure evidence of performance of the behaviors in the classroom. The design therefore requires the investigator to examine the TAs under both the conditions of cognitive assimilation of the behaviors and their performance in the classroom. The performance of the behaviors does not in itself completely indicate successful training. Such training may be considered especially successful if achievement gains can be found among the students of the trained TAs.

Given a program designed for foreign TAs, particular care would be required in matching in order to control for potentially confounding variables. Similarly, if measures of undergraduate student achievement are to represent the dependent variable, steps must be taken to control variables that might constitute extraneous variables in terms of measures of student achievement such as different undergraduate groups receiving lectures from different faculty.

It has been strongly recommended that any training program should be designed as a result of the findings of previous research upon TA training. Bearing this and the preceding conclusions in mind, a training program was designed for the training of novice foreign TAs. This program forms
the treatment for the research study described in Chapter III.
Chapter III

RESEARCH PROCEDURES

This chapter provides a description of the procedures employed to obtain answers to the research questions of this study. It commences with a brief overview of the principal objectives of the study and the context in which it was conducted. The research procedures are then addressed under the headings: (a) sample; (b) design; (c) instrumentation; and (d) data analysis procedures.

Overview

The principal objective of this study was to determine if the undergraduate students of treatment TAs who undertook a summer training program achieved significantly better on end-of-term examinations than the undergraduate students of an equivalent group of (control) TAs who did not participate in the training programs. The secondary objectives of this study were: (a) to determine if the lower-inference teacher behaviors used in the study could be consistently identified by undergraduate students and the extent to which the variable is reliably measured; (b) to determine if the assimilation by the treatment TAs of certain lower-inference teacher behaviors is significantly related to the
achievement of their undergraduate students; (c) to determine if the degree of demonstration of certain lower-inference teacher behaviors by the treatment and control TAs is significantly related to the achievement of their undergraduate students; (d) to determine if there is a significant relationship between the achievement of the undergraduates of the TAs and the TAs' scores on tests of spoken and written English; (e) to determine if there is a significant relationship between the degree of performance of the lower-inference teacher behaviors by the TAs and their scores on tests of spoken and written English.

In order to attain these objectives the study was conducted in three phases. Table 4 illustrates the phases and their purpose.
<table>
<thead>
<tr>
<th>Phase</th>
<th>Time</th>
<th>Purpose</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>b) Production of protocol videotapes.</td>
<td>b) Videotapes of TAs teaching 100 level course material.</td>
</tr>
<tr>
<td>II</td>
<td>Summer 1983</td>
<td>Treatment and examination.</td>
<td>TA training program with examination of assimilated knowledge at conclusion.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Training of raters and observers.</td>
<td>Raters trained to identify behaviors on videotape used for TA examination; observers trained to use Report of Observed Teacher Behaviors instrument.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collection of SPEAK data.</td>
<td>Measures of TAs' spoken English ability.</td>
</tr>
<tr>
<td>III</td>
<td>Fall 1983</td>
<td>Treatment effect measurement.</td>
<td>Classroom observation of TAs' teaching; collection of end-of-term undergraduate achievement scores on examination.</td>
</tr>
<tr>
<td>Phase</td>
<td>Description</td>
<td>Purpose</td>
<td>Department</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>I</td>
<td>a) Undergraduate students in 100 level courses</td>
<td>Pilot-testing Report of Observed Teacher Behavior instrument for reliability.</td>
<td>Math.</td>
</tr>
<tr>
<td></td>
<td>b) Undergraduate students in 100 level courses</td>
<td>Re-testing of amended Report of Observed Teacher Behaviors instrument for reliability</td>
<td>Stat.</td>
</tr>
<tr>
<td>II/III</td>
<td>a) Novice TAs teaching 100 level courses</td>
<td>Treatment group undergoing Summer training program and subsequent classroom observation.</td>
<td>Math.</td>
</tr>
<tr>
<td></td>
<td>b) Novice TAs teaching 100 level courses</td>
<td>Control group not receiving training but receiving subsequent classroom observation.</td>
<td>Stat.</td>
</tr>
</tbody>
</table>
Sample

The subjects involved in each phase of this study are described in Table 5. In Phase I, the development of the observation instrument, the sample consisted of undergraduate students in the departments of Mathematics and Statistics at The Ohio State University. This university is the major comprehensive post-secondary institution in the State of Ohio and offers baccalaureate and advanced degree programs in a wide variety of disciplines. It has an open admissions policy for undergraduate students and therefore its population consists of students from varied socio-economic, ethnic, and cultural backgrounds, as well as varying ability levels. As the TAs involved in the study were to teach undergraduates in these departments at the 100 course level, two such typical classes, one from each department, were selected. The instrument was piloted with undergraduates in a recitation section in Statistics 125 and Mathematics 151.

The use of these undergraduates served several purposes. It provided a means of determining both the internal consistency and the reliability over time (one week) of the instrument. After adjustment, the final instrument was re-tested with two different groups of undergraduate students in 100 level courses in Mathematics and Statistics. The undergraduate responses to the items on the instrument were
correlated from one application to the next item by item for each of the classes of undergraduates involved in the pilot testing of the instrument. This enabled the items to be scrutinized with a view to removing or amending the item. Further, a coefficient of internal validity was calculated for every application of the instrument. This coefficient, a Cronbach alpha, was calculated for the total instrument and for each of the five sections. The implementation and subsequent analysis of results for the initial and final instruments are described under Design and Data Analysis Procedures respectively.

The target population for this study was involved in Phases II and III and consisted of novice foreign TAs enrolled as graduate students at The Ohio State University. The sample consisted of sixteen graduate students, eight (four treatment and four control) from the Mathematics Department and eight Statistics Department. Four of the students from each department were enrolled in a summer training program that was specifically designed to prepare novice foreign TAs in these departments. The term novice indicates that the prospective TA had not taught before in English nor had received any formal teacher training.

Even though three of the four students from the Department of Statistics were English-speaking Americans the requirements under the definition of novice TAs were not vio-
lated as these graduates had not taught before nor had formal teacher training. Although the program was designed specifically for novice foreign TAs it was felt that no loss would accrue from the presence of American students in the program and that some benefit to the foreign students was likely. While the American participants met the requirements under the definition of novice (Chapter I) they were obviously not foreign. This latter tends to explain some of the unique results reported below in Chapter IV.

The two departments were selected for the following reasons: (1) there is a high incidence of foreign TAs in them, (2) there is a commonality of subject matter, and (3) there are usually many prospective novice TAs available during the summer quarter for inclusion in a training program. Unfortunately there was not a high incidence of novice foreign TAs available during the summer in question. Consequently, the size of the sample for treatment was eight, smaller than anticipated, and included three American novice TAs.

Eight novice TAs who did not participate in the training program were selected to form a control group from among those TAs who were not present during the summer but arrived on campus prior to the commencement of teaching in the following quarter. These included three Americans to match the three who had undergone training. Limitations to
this form of selection are discussed under Limitations above. Four TAs were chosen as controls in each department each being matched with a TA in the treatment group. The criteria for matching were as follows: (1) each TA should be novice as in the definition; (2) each pair should be of similar ethnic origin; and (3) each pair should be of equivalent ability in spoken English. The means of ascertaining the last are described under Design in Phase III. Sex was not considered as a criterion variable for matching pairs of TAs as the literature does not support it as a significant variable (Bos, Zakajsek, Wolf, & Stoll, 1980).

**Design**

This section provides a description of the research methodology and data-gathering procedures employed in each phase of the study.

**Phase I: Construction of the Observation Instrument and Production of the Protocol Videotapes**

This phase addressed Research Questions 2 and 3 concerning the measurement of the reliability and internal consistency of the instrument to be used in collecting data concerning the teaching behavior of the TAs. The sources of the items contained in this instrument are presented later in the Instrumentation section of this chapter. The instrument, the Report of Observed Teacher Behaviors, was administered as a
questionnaire, prior to the quarter in which the TA training program took place, to two classes of undergraduate students at the 100 course level, one group in the Statistics Department and one in the Mathematics Department.

Each student completed the full instrument twice, the occasions being separated by one week. The questionnaire (Appendix A) was administered by the researcher using a standard set of procedures for each occasion to ensure uniformity in instrument administration. At the beginning of the administration session, students were informed that the purpose of administering the instrument was to reliability-test the instrument and not to evaluate the teacher. Total time for administration of the instrument on each occasion was fifteen minutes.

As described below under Phase III, this instrument, after analyses for reliability, was amended in consultation with the potential observers and then given to two further groups of undergraduate students as before for further reliability tests. The results for all implementations of the instrument with undergraduate students are reported in Chapter III under Results.

Production of Protocol Videotapes

During this phase, graduate students who had been involved in a TA training program taught by the researcher assisted
in the production of videotapes designed to demonstrate the presence or absence of certain lower-inference teacher behaviors. The episodes were scripted by the graduate students who were from the Statistics Department in collaboration with the researcher and were videotaped in the Teacher Education Laboratory at The Ohio State University. Episodes varied from fifteen to thirty minutes in length and depicted a typical recitation session in which a TA taught mathematical content at the 100 level and answered questions posed by students, peers in this instance.

**Phase II: The Treatment, Rater-Observer Training, and SPEAK Test of Spoken English Implementation**

This phase addressed Research Question 4 which relates to the power of the treatment in terms of the knowledge of the lower-inference teacher behaviors that were assimilated by the treatment group TAs as a result of the training program. These data were collected in the form of a paper and pencil test. This test consisted of a videotape previously unseen by participants of an actual recitation class from the previous year. The test took place in the week prior to the start of Fall Quarter teaching and was preceded by a review session. The reason for this review was to reinforce the knowledge assimilated during the summer program and to take into account the otherwise six week gap between the treatment and the start of the Fall Quarter. It was
Deemed extremely important that the lower-inference teacher behaviors be addressed in review as close to the start of teaching as possible. Collection of these data permits an investigation of the relationship between participants' level of spoken and written English and assimilation of knowledge of the lower-inference teacher behaviors at a given point in time immediately prior to the initiation of teaching.

Participants were permitted to view the 30 minute videotape of a TA teaching undergraduate students in mathematical content at the 100 course level, twice. Participants were asked to list and describe those lower-inference behaviors exhibited by the videotaped TA and those behaviors that were absent. The responses of the participants were compared with those of raters who had previously viewed the videotape and followed the same procedures as required of the participants. These expert raters were TAs trained in the identification of the lower-inference teacher behaviors. These raters identified the behaviors on the examination videotape after training and an index of interrater reliability was calculated as described under Data Analysis Procedures. These raters also formed the observation team for observing TA teaching in the quarter following the training program.
At about the same time as the training program participants took the test they and the control group of associates had their spoken English level assessed. In all, 31 TAs from the two departments took the SPEAK test (see below under Instrumentation) including the 16 TAs in the study. The data were used in a number of analyses as described below under Data Analysis Procedures.

Rater-Observer Training

Over a three-day period, prior to formal observation of a videotape selected for the purpose of testing observer ability, the three observers met with the researcher for training purposes. The first step consisted of discussing the behavioral definitions (Appendix B) of each of the twenty-two items on the Report of Observed Teacher Behaviors Questionnaire (Appendix C) in order to reach consensus as to their meaning.

The Report of Observed Teacher Behaviors consisted originally of 27 behavioral items, 20 of which were drawn from the findings of a study by Hines (1981). Operational definitions associated with 10 of these 20 items were provided by Hines (p. 388). Seven of these 10 were amended slightly in an attempt to ensure observer understanding, one was amended in detail, and two remained unchanged. For the remaining 10 of Hines and the seven from elsewhere, defini-
tions were prepared by the researcher in collaboration with the trained observers.

As a result of analyses from pilot-testing and discussion with the potential observers, the instrument was reduced from 27 to 22 items. Item twenty in the original was considered to be sufficiently similar to item six as to make differentiation difficult. These two items were combined into a final item number six with an appropriately amended behavioral definition. Similarly, original items 15 and 14 resulted in final item 11. Original item nine was not considered to be sufficiently observable to be lower-inference and was considered to be accounted for by the behavioral definition of final item three. Original items 11, 12, and 13 were similarly reduced to final item 10.

Following discussion with the potential observers, amendments were made to Hines' definitions and behavioral definitions for the remaining 12 items were decided. Some minor adjustments to the statements on the original instrument in its questionnaire form were made prior to the second implementation with the undergraduate students (e.g., changing "if students understand" to "if students understand content").

In order to facilitate replication of this study, great care was taken in behaviorally defining the items on the
Report of Observed Teacher Behaviors instrument. Levinson-Rose and Menges (1981) in their review of the research on college teaching pointed out the usefulness of clearly operationalized variables when comparing studies. In general, the behavioral definitions for the lower-inference teacher behaviors used in this study focused on verbal teacher behaviors in an attempt to reduce the need for higher-inference based decisions by observers.

During subsequent sessions observers-in-training recorded the presence or absence of such behaviors using this instrument with a modified scale while watching videotapes of TAs teaching. The instrument is described under Instrumentation and can be found in Appendix D. Observers then compared their reports and examined the sources of disagreement. The modified scale for the instrument allowed for the recording of the absence or presence of the behaviors and in the case of certain ones for the frequency of occurrence. Recording in these sessions was at first collaborative and finally independent.

The final step, once good independent agreement seemed to have been reached, was to view and independently rate two selected videotapes, previously unseen, of TAs teaching. One of these videotapes was used for examining the treatment TAs for the assimilated knowledge of the lower-inference teacher behaviors and was described above under Phase II.
Measures of interrater reliability were computed for each of the videotapes as described under Data Analysis Procedures. The obtained indices for these videotaped lessons were found to be high. The raters-observers were consequently deemed to be trained.

Observation Procedures

Following training, the three trained observers were randomly assigned pairs of TAs each consisting of a TA from each of the treatment and control groups. Each observer then observed his or her assigned TAs three times over a six week period commencing the third week of the Fall Quarter. The observers were required to observe the same pair commencing either the third or fourth week of the quarter and thereafter every two weeks. As a result of the three observations of each TA the frequency of the use of the lower-inference behaviors as exhibited by each TA over the six week period was established. The analysis of these data is described under Data Analysis Procedures.

Phase III: Collection of Observation, Achievement, and Written English Data for Principal Analyses

This phase concerns itself with seven important areas of the study: a) The Major Research Question which relates to the success of the training program in producing TAs whose teaching as measured by their undergraduate students'
achievement is statistically significantly better than that of TAs who did not take the training program; (b) Research Questions 5 and 7 which relate to the relationship between the undergraduate students' achievement and (1) the levels of their TAs' assimilated knowledge and (2) the TAs' demonstrated performance of the lower-inference teacher behaviors addressed in the training program; (c) Research Question 6 which addresses differences between the treatment and control TAs in the use of the lower-inference behaviors during classroom observation over a six week period; (d) Research Question 9 which addresses the relationship between the achievement of the TAs' undergraduate students and the TAs' scores on tests of spoken and written English; (e) Research Question 8 which addresses the relationship between the performance of the lower-inference behaviors and the TAs' scores on tests of spoken and written English; and (f) Research Question 10 which addresses the relationships between TAs' end of training examination scores and scores on tests of written and spoken English.

Observation data were collected by three observers trained by the researcher. The observers were graduate students who had previously undertaken the TA training program and were currently working as TAs in the Statistics Department at The Ohio State University. The observers refrained from professional and social contact with the TAs being studied.
Data concerning the achievement of the undergraduate students of TAs were collected as a result of the final examinations given at the end of the quarter in each department. As all TAs could not be assigned to the same course or instructor, steps were taken to control for these eventualities during analysis and these are described under Data Analysis Procedures.

TA pairs were assigned as recitation leaders to specific courses within each department. Such assignment was arbitrary depending upon the needs of the department. Consequently, not all pairs were assigned to the same undergraduate course nor to a single professor lecturing for the course. Details of the distribution of TA pairs may be seen in Table 6 under Results.

Data concerning the spoken English ability of the TAs were collected from three sources. These were: TAs' scores on the Graduate Record Examination (GRE), on the Test of English as a Foreign Language (TOEFL), and on the Speaking Proficiency English Assessment Kit (SPEAK). They are described under Instrumentation below and the analyses performed with the data are described below under Data Analysis Procedures.

**Instrumentation**
In this study the major dependent variable was the achievement of the undergraduate students of the TAs. The major independent variable whose relationship to the dependent variable was investigated was the treatment effect of the training program undergone by half of the TAs in the study.

Secondary independent variables affecting the dependent variables that were studied were as follows: (1) the level of assimilated knowledge of certain lower-inference teacher behaviors by the trained TAs collected through a paper and pencil examination; (2) the degree of demonstration of those lower-inference behaviors in the classroom by all TAs involved in the study as recorded through the use of the Report of Observed Teacher Behaviors instrument; and (3) the level of spoken and written English ability of all TAs involved in the study as measured by the three instruments, SPEAK, GRE, and TOEFL. These three instruments also constituted independent variables during analyses investigating their relationship to the degree of demonstration of the lower-inference behaviors in the classroom by the TAs.

Undergraduate Achievement

The measure for this dependent variable was the results of the end-of-quarter examination for the students of the TAs in the study. Within each course, Mathematics 151, Statistics 125, and Statistics 133, the undergraduate students
received a common examination set by the lecturing faculty member(s). The examination for Mathematics 151 took the form of problems which were graded by the TAs, each taking a problem across the whole examination. The examinations in Statistics 125 and 133 took the form of multiple choice questions. It may therefore be considered that grader bias was minimal at least in the Statistics Department. The researcher was unable however to control the content validity of the examinations nor was it possible to control for spuriousness resulting from uncontrolled variables such as failure of some recitation TAs to cover all of the material examined. In general, the probability of such an event is not high given the structuring of the courses.

Data were collected directly from departmental records at the conclusion of the quarter in which the TAs first taught. These data may be found under Results (Chapter IV) and were subjected to the analyses described below under Data Analysis Procedures.

**Examination of Assimilated Knowledge**

Data for this variable were collected by means of a paper and pencil test six weeks after the conclusion of the training program and just prior to the TAs commencing teaching. A videotape of a TA teaching for which consensus had been reached by the researcher and three trained raters
on the presence or absence of certain lower-inference teacher behaviors was then studied twice by the training program participants under closed-book examination conditions.

Data collected from participants on the absence or presence of the lower-inference teacher behaviors were compared to that determined by the researcher and trained raters. Where agreement was reached the TA was assigned one point. The score for each TA was out of a possible total of fifty points and is reported under Results.

Report of Observed Teacher Behaviors

The initial instrument piloted with two classes of undergraduate students consisted of (a) an introductory page outlining the purpose of the instrument and giving general instructions for completing it and (b) a second page containing twenty-seven items. The complete instrument may be found in Appendix A.

The twenty-seven items were grouped under five different dimensions or headings. The first twenty items falling under three different headings were drawn from a study (Hines, 1981) which identified three underlying teacher behavioral dimensions defining the clarity variable at the undergraduate college level.
Ten, four and six lower-inference teacher behaviors associated with these three dimensions respectively were identified. These were represented as items 1-10, 11-14, 15-20 on the initial Report of Observed Teacher Behaviors questionnaire. The behavioral definitions for these items are given in Appendix B. Behavioral definitions for items not defined in the Hines study were formulated by the researcher in collaboration with the observers-in-training. The twenty items referenced above were supported by a number of other studies (Meredith and Bub, 1977; Mintzes, 1979; Goepper and Knorre, 1980; Krockover, 1980; Murray, 1983).

**Items 21-24, under the heading:**

*Establishes Rapport with the Students:* were drawn from two studies of teaching at the college level (Mintzes, 1979, Meredith & Bub, 1977) The behavioral definitions for these items are given in Appendix B.

**Items 25-27, under the heading:**

*Uses Communication Skills* were drawn from several studies dealing with TAs in general and foreign TAs in particular (Buckemayer, 1972; Dalgaard, 1976; Goepper & Knorre, 1980; Golmon, 1975; Krockover, 1980; Szymanski, 1978; Hinofotis & Bailey, 1980; Bailey, 1977). The behavioral definitions for these items are given in Appendix B.
As a result of the analysis of responses by the undergraduate students on two occasions separated by one week and discussion with the observers-in-training some adjustment was made to the instrument. This resulted in a final instrument consisting of twenty-two items (Appendix C). The subsequent adjustments were described above under Phase II of Design.

The adjusted instrument used by the trained observers differed from the original in two respects. Firstly, the original instrument (Appendix A) used a five-point scale from "never" to "very often" whereas the final instrument (Appendix D) used by the observers consisted of a frequency scale for eleven of the twenty-two items and a zero/one scale for the remainder, indicative of behaviors absent or present during the lesson.

The second difference involved the abandonment of subheadings dividing the items into five groupings. This was at the request of the observers who desired the items to be set out in an order judged by them to be most conducive to facilitating the observation of the behaviors. As may be seen from Appendix D, this was done without renumbering the items as the observers had come to know the behavioral definitions by the number attached to the item.

*Speaking Proficiency English Assessment Kit (SPEAK)*
This instrument is a product of the Test of English as a Foreign Language (TOEFL) program, which is sponsored by the Educational Testing Service. SPEAK is described as "(A) n off-the-shelf product for administering and scoring the Test of Spoken English (TSE) developed in response to the interest expressed by many institutions in an instrument to assess the spoken English proficiency of foreign teaching assistants and other international students who are already a part of the student body" (TOEFL Program, 1982, page 5).

The kit was purchased through the Office of the Dean of Graduate Studies at The Ohio State University. Following the procedures laid down in the directions the test was implemented by the researcher and scored by trained scorers in the office of the American Language Program at The Ohio State University.

The data obtained provide measures of the examinee's pronunciation level, spoken grammar level, spoken fluency, and overall spoken comprehensibility. The first three items are recorded using a four-point scale from zero to three and the last item a five-point scale from zero to three hundred. This last score was that used in the analyses involving spoken English ability in this study.

Graduate Record Examination (GRE) Aptitude Test
This test, developed by the Educational Testing Service (1980) provides measures of developed verbal, quantitative reasoning, and analytical reasoning abilities. It is a paper and pencil test using written instructions and is frequently used for admission requirements to graduate schools and departments. The population against which the test is standardized is American though the test is commonly used for admission purposes for foreign nationals whose background and language are different from American Nationals (TOEFL Research Reports, 1982, page 1).

Such data are usually obtained through testing the applicant at a designated Educational Testing Service site and are then transmitted to the institution to which the applicant has applied. These data were therefore obtained from departmental records.

Test of English as a Foreign Language (TOEFL)

This test, developed by the Educational Testing Service (1981), is designed for use as a measure of English proficiency for individuals whose native language is not English and is frequently used by graduate schools and departments when applicants' English language proficiency is questionable (TOEFL Research Reports, 1982).

TOEFL yields measures of listening comprehension (ability to understand English as it is spoken in the United States), structure and written expression (mastery of structural and gram-
matical points in standard written English), and reading comprehension vocabulary (word knowledge and understanding of a variety of reading materials). Scores on the three sections are provided along with a total score (TOEFL Research Reports, 1982, page 1).

This test is standardized on a population consisting predominantly of foreign nationals whose native language is not English. Testing is carried out by the Educational Testing Service at designated sites and the results conveyed to the institution to which the applicant has applied. These data were therefore obtained from departmental records.

Data Analysis Procedures

Following data collection for each phase of this study, obtained responses to the instruments were prepared for statistical analysis procedures. Appropriate steps were taken for handling missing data during statistical analysis (see Chapter IV). Descriptive statistics were computed for data from each instrument and these statistics are reported in Chapter IV.

The data analysis procedures specific to each phase of the study are described below. The results obtained from these procedures are reported in Chapter IV.

Phase I: The Observation Instrument
The analysis procedures undertaken in this phase of the study provided answers to Research Questions 2 and 3.

Research Question 2: Can the lower-inference teacher behaviors selected for study be identified in the classroom by undergraduate students with consistency?

This question addresses the reliability over time of the Report of Observed Teacher Behaviors questionnaire. To determine the test-retest reliability of the instrument two groups of undergraduate student (total n=83) in 100 level courses in Mathematics and Statistics respectively at The Ohio State University were given the instrument on a test-retest basis with an interval of eight to ten days between applications. The two resulting data sets comprised pairs of student responses from each application of the instrument. Where a student had not responded to one of the two applications the data were excluded from analyses. This left a usable n of 61. The coefficient of stability was calculated by correlation for the two sets of scores item by item in order that consideration could be given to removing items showing very low correlations between applications.

In the interest of rigor the two sets of scores were correlated item by item for each of the 27 items. As de-
scribed under Design, subsequent adjustments to the instrument resulted in the final consisting of 22 items. This instrument was then used with two new groups of undergradu-
ates (total n=47) as before. Usable returns numbered 39. Results of the applications are reported in Chapter IV.

Research Question 3: Do the lower-inference behaviors selected for study represent reliable measures of the variable clarity?

This question addressed the internal consistency of the instrument.

Within each of the five dimensions of the instrument, reliability estimates were computed using "Subprogram Reli-
ability" in the Statistical Packages for the Social Science (SPSS) Update (Hull and Nie, 1979). The reliability coeffi-
cient estimated in each case was Cronbach's Coefficient Alpha. This determined the instrument's internal consist-
tency for each of its five dimensions and for the overall instrument.

After adjustment the reduced instrument was retested with two new groups of undergraduate students in Mathematics and Statistics. Results for the original and final in-
strument are reported in Chapter IV.

Phase II: Training Program Data
(a) **Interrater reliability for the examination videotape.**

In order to determine which lower-inference behaviors were or were not evidenced on the examination videotape trained observers using a check-list of the lower-inference behaviors watched the tape twice. The observers independently checked those behaviors present and the check-lists were then compared using a correlational technique. A correlation matrix was constructed for the ratings of the three observers and the researcher. Almost perfect agreement was found with only very small differences existing on two of the 22 items of the instrument for two of the observers.

(b) **Program participants' examination responses.**

The analysis procedures undertaken in this phase of the study provided answers to Research Question 4.

Research Question 4: Will the participant novice TAs (treatment group TAs) in the training program assimilate the knowledge concerning the lower-inference teacher behaviors?

The written responses of the participants identifying those lower-inference behaviors present and absent on the examination videotape were compared for each participant with those agreed upon by the trained raters. For each lower-inference behavior correctly identified as present or ab-
sent the examinee was assigned one point. Where the testee identified lower-inference behaviors as both present and absent the testee was assigned no point for the item.

Analyses directed at finding the relationship between participants' assimilated knowledge of the lower-inference behaviors and the achievement of their undergraduate students are addressed under the Phase IV heading below.

**Phase III: Principal Analyses**

This phase dealt with seven important questions in the study including the Major Research Question. Table 7 pictorially represents the analyses carried out in this and preceding phases. Each area is addressed below.
<table>
<thead>
<tr>
<th>Phase</th>
<th>Purpose</th>
<th>Analyses</th>
</tr>
</thead>
</table>
| I     | Identification by undergraduates of the lower-inference teacher behaviors: internal reliability and consistency over time of the instrument | Cronbach Alpha  
Pearson Correlation |
| II    | Interrater reliability | Pearson Correlation |
| III   | a. TA training and undergraduate achievement | T-test for each TA pair; dependent variable achievement |
|       | b. TA knowledge and undergraduate achievement | T-test; dependent variable achievement |
|       | c. TA training and uses of lower-inference behaviors | Chi-square for binary data, t-test for frequency data; dependent variable behavior |
|       | d. TA behavior use and undergraduate achievement | T-test; achievement as dependent variable |
|       | e. TA English ability and use of behaviors | ANOVA, per-behavior summed over three observations as dependent variable |
Table 7 (continued)

f. TA English ability and undergraduate achievement  ANOVA, achievement as dependent variable by trained or untrained

g. Relationship between English ability scores (TOEFL, GRE-V, SPEAK) and TA course exam score  Pearson Correlation
(a) **Major Research Question:** Will the training program designed as the treatment for novice teaching assistants in the Departments of Statistics and Mathematics produce treatment TAs whose teaching is statistically significantly better than that of the control TAs as measured by the achievement of the TAs' undergraduate students?

This question addressed the relationship between undergraduate student achievement and the treatment. Its intent is to find if the achievement of undergraduate students of TAs experiencing the training program is statistically significantly superior to that of undergraduate students of TAs not experiencing the training program.

Using a one-tailed t-test analyses were carried out comparing treatment and control TA matched pairs using the raw undergraduate achievement scores. Results are reported in Chapter IV. The significance level was set at 0.05.

Alternate analyses could take the form of combining undergraduate scores and working with overall means. Given the small number of TA recitation classes involved, it was decided to treat the analyses as exploratory and utilize those analyses described above.

(b) **Undergraduate Achievement and TA Assimilated Knowledge**

This analysis provided answers to Research Question 5.
Research Question 5: Will there be a statistically significant difference in the achievement of the undergraduate students of the treatment group TAs who do and who do not assimilate the knowledge concerning the lower-inference teacher behaviors?

Within each course (Stat. 125, Stat. 133, Math. 151) a t-test was applied using TA scores on the training program examination designated as high (above 45) or low (below 46) as the independent variable and undergraduate achievement scores on the course examinations as the dependent variable.

(c) Differences in Use of Behaviors by Treatment and Control TAs

The analyses provided answers to Research Question 6.

Research Question 6: Will there be a statistically significant difference in the demonstration of the lower-inference teacher behaviors by the treatment and control group TAs over a six week time period of classroom observation?

In order to investigate differences in the use of behaviors between the treatment and control groups of TAs, two analyses were carried out for the three sets of observations made over the six week period. A two-way contingency table (Chi-square) was constructed for each of the 12 behaviors recorded on a binary scale for each time of observation
This produced 36 Chi-square tables. For the 10 behaviors recorded on a frequency of use scale a t-test was used for each behavior for each time of observation. This produced 30 analyses. The analyses were also performed on each of the 22 behaviors separately using data combined over the three observations.

Table 8 shows the analyses pictorially. The results are reported in Chapter IV. Using the Report of Observed Teacher Behaviors instrument data were collected on each of the TAs in the treatment and control groups as a result of three visits to each TA's classroom by a trained observer who did not know if the TA was treatment or control. The instrument collecting this data, the Report of Observed Teacher Behaviors Questionnaire, consists of five dimensions containing a total of twenty-two items (Appendix D). The instrument was used three times in the observation of each TA teaching his or her recitation group. Of the twenty-two items, ten used a frequency count and twelve used a 0/1 (no/yes) count.

(d) Undergraduate Achievement and TA Performance of the Lower-inference Behaviors

The analyses provided answers to Research Question 7.

Research Question 7: Is there a significant relationship between the behaviors used by the TAs and the achievement scores of the TAs' undergraduate students?
In order to test for significant relationships, a t-test was used. The dependent variable was the undergraduate achievement scores and the independent variable was high or low usage of the lower-inference behavior in question. Tests were performed item by item for the 22 items involved.

For those 10 behavior items recorded at each observation using a frequency scale the frequency was summed across the three observations selecting the appropriate level for the high or low use categorization. Similarly, an analysis was carried out for these behaviors combined by cluster for the five sections of the instrument.

For those 12 behavior items recorded at each observation using a binary scale similar summation was made across observations for each of the twelve behaviors. A total sum of 0-1 was categorized as low use and 2-3 as high use. Table 9 represents the analyses. A similar analysis was carried out for these behaviors combined by cluster for each section of the instrument.
### TABLE 8

**CHI-SQUARE AND T-TEST ANALYSES**
**FOR TA USE OF BEHAVIORS**

#### a. Binary data*

<table>
<thead>
<tr>
<th></th>
<th>Low use of behavior item</th>
<th>High use of behavior item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained (n=8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untrained (n=8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*behavior items 5, 6, 10, 11, 13, 14, 15, 18, 19, 20, 21, 22

#### b. Frequency data**

<table>
<thead>
<tr>
<th></th>
<th>Behavior use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trained (n=8)</td>
<td></td>
</tr>
<tr>
<td>Untrained (n=8)</td>
<td></td>
</tr>
</tbody>
</table>

**behavior items 1, 2, 3, 4, 7, 8, 9, 12, 16, 17

Analyses were performed for each behavior separately for each observation and combined over three observations.
TABLE 9

T-TEST FOR UNDERGRADUATE ACHIEVEMENT AND TA USE OF BEHAVIORS

<table>
<thead>
<tr>
<th>High use*</th>
<th>Undergraduate achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low use*</td>
<td></td>
</tr>
</tbody>
</table>

* Each behavior tested separately summed across the three observations. In addition, the behaviors tested by section of the instrument (e.g., items 1,2,3,4,7,8,9 constitute section 1 for the frequency data)
### TABLE 10

ANOVA FOR USE OF BEHAVIORS AND THE ENGLISH ABILITY OF TRAINED AND UNTRAINED TAS

<table>
<thead>
<tr>
<th></th>
<th>Trained (n=8)</th>
<th>Untrained (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High ability</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Low ability</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

*Analysis carried out per behavior item summed over three observations, and per section of the instrument (e.g., items 1, 2, 3, 4, 7, 8, 9 constitute section 1 for the frequency data)
**TABLE 11**

ANOVA FOR UNDERGRADUATE ACHIEVEMENT AND THE ENGLISH ABILITY OF TRAINED AND UNTRAINED TAS

<table>
<thead>
<tr>
<th>Ability</th>
<th>Trained (n=8)</th>
<th>Untrained (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High ability</td>
<td>UA*</td>
<td>UA*</td>
</tr>
<tr>
<td>Low ability</td>
<td>UA*</td>
<td>UA*</td>
</tr>
</tbody>
</table>

*undergraduate achievement*
(e) **TAs' Demonstrated Behavior and Scores on Language Tests**

Analyses provided answers to Research Question 8.

Research Question 8: Is there a statistically significant relationship between the TAs' scores on tests of spoken and written English and their demonstrated performance of the lower-inference teacher behaviors?

To examine this relationship between TA behavior as the dependent variable and English ability as the independent variable, an analysis of variance technique was used. TAs were assigned as high or low in English ability and as treatment or control. For each behavior item the frequency or binary score was summed across the three observations and the analyses were carried out separately for each behavior item. Table 10 pictorially represents the analyses. Where appropriate, data were summed by section of the instrument for analysis. Not all of the five sections (see Appendix D) contain sufficient data on the frequency or binary scales to warrant summing.

(f) **Undergraduate Achievement and TAs' Scores on Language Tests**

Analyses provided answers to Research Question 9.
Research Question 9: Is there a significant relationship between the TAs' scores on tests of written and spoken English and the achievement of their undergraduate students?

In order to test for significant relationships between the TAs' ability on tests of English and the achievement of their undergraduate students, an analysis of variance was carried out with achievement as the dependent variable and TA English ability as the independent variable which was categorized as high or low. The dependent variable was subdivided into cells representing undergraduates of trained and untrained TAs in order to examine interaction effects between training and English ability. Table 11 represents the analyses.

Scores on the SPEAK test were taken directly from the results provided by the office of the American Language Program at The Ohio State University. The data for the TOEFL and GRE-Verbal scores were provided by the Educational Testing Service and obtained through the departmental records. Results are reported in Chapter IV.

(g) TAs' Demonstrated Behavior and Scores on Language Tests

Analyses provided answers to Research Question 10.

Research Question 10: What is the relationship between TAs' scores on tests of written and spoken English and scores on the training program examination?
Using a correlational technique (Pearson r) the relationship between TAs' examination score in the training program and scores on the SPEAK test of spoken English, the Test of English as a Foreign Language, and the GRE-Verbal was investigated.

The data for the SPEAK test scores were provided by the office of the American Language Program at The Ohio State University and the scores on the Test of English as a Foreign Language (TOEFL) and the Graduate Record Examination-Verbal (GRE-Verbal) were provided from departmental records.

Summary

The purpose of this chapter was to describe the analyses performed in order to answer the questions posed in the study. Chapter IV which follows presents the results of these analyses.
Chapter IV
RESULTS AND CONCLUSIONS

Introduction

This chapter presents the results of the analyses described in Chapter III above. It follows the same three phases as previously.

Phase I: The Observation Instrument

Research Question 2: Can the low-inference teacher behaviors selected for study be identified in the classroom with consistency by undergraduate students?

This question addressed the reliability over time of the instrument whose development is described in Chapter III. As described there, the original instrument, consisting of twenty-seven items was administered to two classes of undergraduate students at the 100 course level (n=38 and 21 usable returns) in the departments involved in the study. Eight to ten days later the instrument was re-administered and correlations calculated for each of the twenty-seven items on the test-retest responses.
For the twenty-seven items the Pearson $r$ ranged from 0.4 to 0.8, 75% exceeding the 0.5 level. See Chapter III for analysis details.

As described in Chapter III modifications were made to the instrument which resulted in a reduction to twenty-two items. The instrument was then tested as before with two different classes of undergraduate students (n=23 and 16 usable returns). Once again the Pearson $r$ ranged from 0.4 to 0.8 with 75% of the correlations again exceeding the 0.5 level.

It was concluded from this rigorous test of the instrument that its reliability over time was satisfactory. Research Question 2 was therefore answered in the affirmative.

Research Question 3: Do the lower-inference behaviors selected for study represent reliable measures of the variable clarity?

In order to answer this question, reliability indices (Cronbach Alpha) were calculated for each occasion of administration of the instrument, which was four both for the original and amended instrument. In each case an index was calculated for the total instrument. In every case the alpha for the full instrument was in the range of the 0.87 to
the 0.94 level. For the final instrument, the indices for
the five sections were in the range of 0.66 to 0.85.

It was concluded, given these indices, that the final
twenty-two item instrument reliably measured the variable
clarity. Research Question 3 was answered in the affirma-
tive.

Phase II: Training Program Data

a) Interrater reliability

Three graduate students were trained as observers as de-
scribed in Chapter III. After training, the observers and
the researcher independently observed and rated a videotape
of a TA teaching. The instrument used was the Report of
Observed Teacher Behaviors described in Chapter III. The
observers rated the TA on twenty-two behaviors.

A Pearson correlation was performed on the ratings of
the three observers and the researcher. The percentage of
agreement on the twenty-two behaviors was found to range
from 98% to 100%.

(b) Training program participants' examination responses.

The same videotape as rated by the observers and researcher
was shown to the trained TAs at the conclusion of the
course as delineated in Chapter III. TAs were required to
identify the presence or absence of the twenty-two behaviors which were the focus of the course. The purpose was to provide an answer to Research Question 4.

Research Question 4: Will the participant novice TAs (treatment group TAs) in the training program assimilate the knowledge concerning the lower-inference teacher behaviors?

Utilizing the consensus rating from the observers and researcher the participant examination responses were scored out of a total of fifty points as described in Chapter III. Of the eight TAs who took the course, two TAs scored forty-one points and the remainder fifty. It was therefore concluded that the TAs had generally assimilated the knowledge of the lower-inference teacher behaviors such that they could identify their presence or absence on a videotape of a TA teaching. Research Question 4 was answered in the affirmative.

Phase III: Principal Analyses

1. Major Research Question: Will the training program designed as the treatment for novice teaching assistants in the Departments of Statistics and Mathematics produce treatment TAs whose teaching is statistically significantly better than that of the control group TAs as measured by the achievement of the TAs' undergraduate students?
Analyses focussed on the TA pairs, each consisting of one trained and one untrained TA. These pairs were independent. Consequently a one-tail t-test was applied to each pair using a statistical package supplied by Statistical Analysis Systems (SAS). The package accounted for unequal numbers of undergraduate students in the TAs classes when analyzing data and the appropriate t and p values were used in reporting the findings. The dependent variable for the analysis was undergraduate raw achievement scores on the end-of-term examination and TA training was the independent variable.

Significant differences were found at the 0.05 level for pairs 3, 6 and 8. In the instance of pair 8, the two groups of undergraduate students were not taught by the same professor. The different faculty involved therefore may be considered to be an uncontrolled (extraneous) variable which may have affected the results.

While the significant difference in achievement shown by the undergraduates of the TAs in pairs 3 and 6 favors the trained TA it provides insufficient evidence such that the Major Research Question can be answered in the affirmative. Nevertheless, the findings, which are shown in Table 6, provide grounds for further analyzes in this area. These are discussed Chapter V.
<table>
<thead>
<tr>
<th>TA PAIR</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Course</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Common course exam</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Same faculty member for students of each TA of pair</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Number of undergraduate students in class</td>
<td>31</td>
<td>18</td>
<td>32</td>
<td>19</td>
<td>30</td>
<td>38</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>Classes of each TA of pair in matching time slot</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Same observer for each TA of pair</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Each TA of pair matched on ethnicity</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Each TA of pair matched on spoken English ability</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TA previous teaching experience in U.S. classrooms</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TA experienced formal preparation</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Undergraduate mean achievement</td>
<td>102.00</td>
<td>100.06</td>
<td>98.38</td>
<td>90.11</td>
<td>139.00</td>
<td>127.76</td>
<td>114.41</td>
<td>111.67</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>14.31</td>
<td>39.28</td>
<td>15.02</td>
<td>24.73</td>
<td>22.91</td>
<td>27.03</td>
<td>22.56</td>
<td>23.68</td>
</tr>
<tr>
<td>t value</td>
<td>0.2724</td>
<td>1.3201</td>
<td>1.8183</td>
<td>0.3162</td>
<td>-0.6771</td>
<td>1.6682</td>
<td>0.5337</td>
<td>-1.8798</td>
</tr>
<tr>
<td>p value</td>
<td>0.42</td>
<td>0.95</td>
<td>0.03</td>
<td>0.37</td>
<td>0.25</td>
<td>0.05</td>
<td>0.30</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Research Question 5: Will there be a statistically significant difference in the achievement of the undergraduate students of treatment group TAs who do and do not assimilate the knowledge concerning the lower-inference teacher behaviors?

A simple t-test was used to test for differences. The dependent variable was undergraduate raw scores on the end of term examinations and the independent variable, TA assimilated knowledge as measured by the training program examination, was categorized as high (score > 45) or low (score < 46) level.

No significant differences were found between the achievement scores of the undergraduate students of the high assimilation TAs (n=6) and low assimilation TAs (n=2). Table 12 shows the means, standard deviation, t and p values. Research Question 5 could not be answered in the affirmative.
TABLE 12

UNDERGRADUATE ACHIEVEMENT SCORES AND
TA ASSIMILATED KNOWLEDGE ON TRAINING
PROGRAM EXAMINATION

<table>
<thead>
<tr>
<th>TAs</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (n=2)</td>
<td>135.15</td>
<td>28.40</td>
<td>-0.9702</td>
<td>0.3352</td>
</tr>
<tr>
<td>High (n=6)</td>
<td>142.31</td>
<td>35.30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Research Question 6: Will there be a statistically significant difference in the demonstration of the lower-inference teacher behaviors by the treatment and control group TAs over a six-week time period of classroom observation?

Twenty-two behaviors were observed on three different occasions over six weeks at two week intervals. Twelve behaviors used a binary (no/yes) observation scale and ten used a frequency scale. A Chi-square test was used for analyzing each of the former and a one-tail t-test for each of the latter. Analyses were carried out for each of the 22 behavior items separately for each of the three observations and combined over the three observations.

"Binary" Behaviors

The Chi-square analyses failed to reveal any significant differences in the use of the behaviors recorded as binary data by the treatment and control group TAs. Certain of these behaviors were used by almost all of the sixteen TAs and certain others by practically none of the TAs. These behaviors are listed below.

Behaviors used infrequently by all TAs:

#13 Informs students of the course/lesson objectives.
#14 Tells students what they are expected to know or should be able to do on completion of instruction.
#15 Summarizes material presented in class.
#18 Encourages student participation.
#22 Makes use of a variety of teaching aids.

Behaviors used frequently by all TAs:

#5 Explains things simply.

#6 Teaches at a pace appropriate to the topic and to the students.

#10 Shows students how to do the work by use of examples.

#11 Teaches step-by-step.

#19 Looks at the class when teaching.

#20 Pronounces words clearly.

#21 Can be clearly heard in the classroom.

"Frequency" Behaviors

A t-test applied to each behavior item using data summed over the three observations during the six week period identified four behaviors which the 8 treatment TAs performed significantly more frequently than the 8 control TAs. Table 13 shows the means, standard deviations, t values (adjusted for unequal n when appropriate) and p value for these items. The behaviors were:

#2 Asks questions to find out if students understand.

#4 Explains something and then stops so that students can ask questions.

#7 Provides time for students to practice.

#17 Praises student contributions.
Summing the mean use of each of the 10 behaviors measured by frequency data for the treatment and control group separately for each of the three observations enabled a grand mean representative of the use of these over time to be calculated. These are shown below in Table 14.
TABLE 13

BEHAVIORS* USED SIGNIFICANTLY MORE OFTEN BY TREATMENT TAS \((n=8)\) THAN CONTROL TAS \((n=8)\)

<table>
<thead>
<tr>
<th>Item</th>
<th>TAs</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>treatment</td>
<td>24.63</td>
<td>11.93</td>
<td>1.7626</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>14.50</td>
<td>11.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>treatment</td>
<td>16.38</td>
<td>10.28</td>
<td>1.7003</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>9.25</td>
<td>5.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>treatment</td>
<td>20.88</td>
<td>14.57</td>
<td>2.0682</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>8.50</td>
<td>8.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>treatment</td>
<td>10.25</td>
<td>12.98</td>
<td>1.7110</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>2.13</td>
<td>3.44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Summed over three observations
TABLE 14

GRAND MEANS PER OBSERVATION OF THE TEN
BEHAVIORS MEASURED USING A FREQUENCY
SCALE

<table>
<thead>
<tr>
<th>Observation</th>
<th>All TAs</th>
<th></th>
<th></th>
<th>Foreign TAs</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment TAs (n=8)</td>
<td>Control TAs (n=8)</td>
<td>Treatment TAs (n=5)</td>
<td>Control TAs (n=5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>34.25</td>
<td>23.64</td>
<td>19.88</td>
<td>8.98</td>
<td>19.60</td>
<td>12.05</td>
</tr>
<tr>
<td>2</td>
<td>31.25*</td>
<td>16.64</td>
<td>18.50*</td>
<td>10.46</td>
<td>22.00</td>
<td>11.53</td>
</tr>
<tr>
<td>3</td>
<td>28.50</td>
<td>12.92</td>
<td>19.40</td>
<td>13.06</td>
<td>22.60</td>
<td>5.27</td>
</tr>
</tbody>
</table>

* Significant difference at p=0.05
As may be seen, the use of the combined behaviors is relatively stable for the control group of 8 TAs over the six-week period, and always less than that of the treatment TAs. The use of behaviors by the eight treatment TAs gradually decreases over that period, the difference between the usage by the two groups gradually diminishing. The difference was only significant at $p=0.05$ for the second observation made in the fourth or fifth weeks. This difference involves all sixteen TAs of which six were native American. When the same analysis was carried out using only the five pairs of foreign TAs, no significant differences were found. Details may be found in Table 14.

The number of subjects involved is small but these results nevertheless suggest that training can impact on the level of use of some behaviors. It is possible therefore to answer Research Question 6 in the affirmative but only with respect to certain behaviors.

Research Question 7: Is there a significant relationship between the behaviors used by the TAs and the achievement scores of the TAs' undergraduate students?

As the analyses used in the Major Research Question could not establish TA training and undergraduate achievement to be significantly related, the analyses for Question 7 used a t-test. Raw undergraduate achievement scores were used as the dependent variable and level of use of behavior by the 16 TAs as the independent variable.
Data for each behavior item measured on a frequency scale were combined over the three observation sessions and an arbitrary decision was made as to "high" and "low" use of the behavior item given the distribution of use by the 16 TAs. Items 8, 9 and 16 exhibited little variation such that no designation could be made.

Table 15 shows the designated cut-off points for the remaining items, the n for each cell, the F value, p value, and direction of difference. For example, high use of behavior item 3 was set at greater than a total of five times over the three observations. Low use was set at zero. The number of undergraduates in the two cells was 100 and 92 respectively. The analysis showed that the undergraduates in the high use cell performed significantly better. The table also shows the results when items were combined to reflect the sections of the instrument. An attempt was made to analyze the data by course (Stat. 125, Stat. 133, Math. 151) but the prevalence of empty cells rendered the findings effectively unusable.
<table>
<thead>
<tr>
<th>Behavior item</th>
<th>High use (N)</th>
<th>Low use (N)</th>
<th>F value</th>
<th>p value</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&gt;15 (146)</td>
<td>≤10 (100)</td>
<td>0.64</td>
<td>0.4242</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>&gt;21 (195)</td>
<td>≤21 (143)</td>
<td>0.23</td>
<td>0.6336</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>&gt;5 (100)</td>
<td>≤0 (92)</td>
<td>10.90</td>
<td>0.0011</td>
<td>UG(H)&gt;UG(L)</td>
</tr>
<tr>
<td>4</td>
<td>&gt;20 (109)</td>
<td>≤6 (111)</td>
<td>27.78</td>
<td>0.0001</td>
<td>UG(L)&gt;UG(H)</td>
</tr>
<tr>
<td>7</td>
<td>&gt;21 (112)</td>
<td>≤5 (144)</td>
<td>25.71</td>
<td>0.0001</td>
<td>UG(L)&gt;UG(H)</td>
</tr>
<tr>
<td>12</td>
<td>&gt;3 (122)</td>
<td>≤0 (86)</td>
<td>4.70</td>
<td>0.0313</td>
<td>UG(H)&gt;UG(L)</td>
</tr>
<tr>
<td>17</td>
<td>&gt;8 (119)</td>
<td>≤0 (107)</td>
<td>26.66</td>
<td>0.0001</td>
<td>UG(L)&gt;UG(H)</td>
</tr>
<tr>
<td>Section 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items 1,2,3,4</td>
<td>&gt;79 (111)</td>
<td>≤46 (128)</td>
<td>15.77</td>
<td>0.0001</td>
<td>UG(L)&gt;UG(H)</td>
</tr>
<tr>
<td>7,8,9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 12</td>
<td>&gt;3 (122)</td>
<td>≤0 (86)</td>
<td>4.70</td>
<td>0.0313</td>
<td>UG(H)&gt;UG(L)</td>
</tr>
<tr>
<td>Section 4:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Items 16,17</td>
<td>&gt;6 (117)</td>
<td>≤0 (135)</td>
<td>17.12</td>
<td>0.0001</td>
<td>UG(L)&gt;UG(H)</td>
</tr>
</tbody>
</table>

UG=undergraduate achievement; (H)=high behavior use; (L)=low behavior use; N= number of undergraduates
The results available from Table 15 suggest the use of certain behaviors to be related to undergraduate achievement. With the exception of item 3 (repeats things when students do not understand) and item 12 (points out what is important for students to learn), greater undergraduate achievement appears related to less use of three items. The items are:

#4 Explains something and then stops so that students can ask questions.
#7 Provides time for students to practice.
#17 Praises student contributions.

Performing the same analyses for the 10 foreign TAs found non-significant differences on all of the behavior items recorded using a frequency measure. It would appear that any differences in use of the behaviors was due to the presence of the native American TAs.

The use of the behaviors recorded as binary data was uniform across the TAs. Consequently, it was not possible to carry out the analyses proposed.

Research Question 8: Is there a statistically significant relationship between the TAs' scores on tests of spoken and written English and their demonstrated performance of the lower-inference behaviors?
Of the 16 TAs involved, scores on TOEFL and GRE-V were available for only 4 and 6 TAs respectively. Consequently analyses were performed using only the spoken English scores based on SPEAK. For purposes of analysis, the American TAs were assigned a score of 300 which represents speaking English "like a native American". Analysis revealed that the scores on SPEAK were not significantly different for the treatment and control groups of foreign TAs.

A two-way ANOVA was used to test for the relationship between SPEAK scores, training, and the performance of the behaviors. A median score of 245 on the test SPEAK was set as the break point for "high" and "low" spoken English ability. The dependent variable, the behavior item treated as frequency data, was summed across the three observations. Significant findings were made for three items. Figures 1 through 3 illustrate the relationship between training, spoken English level, and the use of the behavior respectively for the 16 TAs. As may be seen from Figures 1 through 3, the interactions for behavior items 2 and 7 are disordinal. The interaction for behavior item 17 is close to ordinal. Except for the last, this suggests that the relationship between the dependent variable and one of the independent variables will be influenced by the second independent variable as it interacts with the first independent variable.
Figure 1. Representation of the relationship between use of behavior 2 and spoken English level for trained and untrained TAs.

<table>
<thead>
<tr>
<th></th>
<th>Trained (n=8)</th>
<th>Untrained (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Level</td>
<td>34</td>
<td>11</td>
</tr>
<tr>
<td>Low Level</td>
<td>16</td>
<td>18</td>
</tr>
</tbody>
</table>
Figure 2. Representation of the relationship between use of behavior 7 and spoken English level for trained and untrained TAs.

<table>
<thead>
<tr>
<th></th>
<th>Trained (n=8)</th>
<th>Untrained (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Level</strong></td>
<td>34</td>
<td>8</td>
</tr>
<tr>
<td><strong>Low Level</strong></td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 3. Representation of the relationship between use of behavior 17 and spoken English level for trained and untrained TAs.

<table>
<thead>
<tr>
<th></th>
<th>Trained (n=8)</th>
<th>Untrained (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Level</td>
<td>19</td>
<td>2.5</td>
</tr>
<tr>
<td>Low Level</td>
<td>1.5</td>
<td>1.75</td>
</tr>
</tbody>
</table>
For behavior 2 (asks questions to find out if students understand), there was a significant interaction between training and spoken English level (p=0.025). The use of the behavior was related to training, but the kind of relation depended on the level of spoken English. Under the condition of high English ability, the behavior was used more by trained TAs but under the condition of low English ability the behavior was used more by untrained TAs.

For behavior 7 (provides time for students to practice), there was a highly significant interaction between training and spoken English (p=0.004). The use of the behavior was related to training, but the relation depended on English ability. Under the condition of high English ability, the behavior was used more by trained TAs but, under the condition of low English ability, the behavior was used more by untrained TAs.

For behavior 17 (praises student contributions), there was a significant interaction between training and spoken English (p=0.042). The use of the behavior was related to training, but the relation depended on English ability. Under the condition of high English ability, the behavior was used more by trained TAs but under the condition of low English ability the behavior was used more by untrained TAs.
Analyses of the data for only the 5 pairs of foreign TAs failed to reveal any significant differences in behavior use for any item either among the 10 TAs or between the 5 trained and 5 untrained TAs. For the purpose of the analysis, the SPEAK score was set at high (>235) and low (<200). When the data were summed by section of the instrument, significant differences were found between the trained TAs for section one (Items 1, 2, 3, 4, 7, 8, 9). High spoken English ability was related ($p=0.02$) with high use of the behavior.

A further analysis was carried out on these data. This took the form of a Pearson correlation between the SPEAK scores and each item of behavior summed over the three observations. Matrices were produced for SPEAK and the ten frequency behaviors for all 16 TAs and for the 8 trained and 8 untrained TAs separately.

For all 16 TAs, only item 7 (provides time for students to practice) correlated significantly with SPEAK scores ($r=0.5$, $p=0.04$). For the 8 trained TAs, SPEAK correlated significantly with item 7 ($r=0.9$, $p=0.005$) and item 2 (asks questions to find out if students understand) ($r=0.9$, $p=0.004$). None of the items correlated significantly with SPEAK scores for the 8 untrained TAs. Collapsing the items by instrument section (section 1 includes items 1, 2, 3, 4, 6 & 9) produced a significant correlation with SPEAK scores ($r=0.9$, $p=0.006$) only for the trained TAs. An examination
of the matrices reveals items 2, 4, 7, and 17 to be highly interrelated for the trained TAs but not for the untrained TAs.

For the 10 foreign TAs, significant correlations between SPEAK scores and items summed over the three observations were found in only two instances. The 5 trained TAs' SPEAK scores correlated highly with item 2 (asks questions to see if students understand) (r=0.9, p=0.04). The SPEAK scores of the untrained TAs correlated highly and negatively with item 4 (explains something and then stops so that students can ask questions) (r=-0.9, p=0.06). In general, the untrained foreign TAs' SPEAK scores correlated negatively with most of the behaviors measured on the frequency scale while this was only so for three of the ten items for the trained foreign TAs.

Research Question 9: Is there a statistically significant relationship between the TAs' scores on tests of written and spoken English and the achievement of their undergraduate students?

As explained above, scores on tests of written English were not available in sufficient quantity for use in analyses. Consequently the scores on the SPEAK test of spoken English were used as the independent variable in a one-way ANOVA in which undergraduate achievement was the dependent variable.
Spoken English ability was set as high (SPEAK score>245) or low (SPEAK score<246) and the analysis was carried out separately for all sixteen TAs, the eight treatment group TAs, and the eight control group TAs.

In all instances, the results showed that the higher undergraduate achievement was associated with the lower SPEAK score. Table 16 shows the means, F and p values for each of the analyses. The results suggest that a significant relationship does exist between a TAs' spoken English ability and the achievement of the TAs' undergraduate students. The relationship appears to be an inverse one. When the same analyses were performed using only data involving the foreign TAs, with SPEAK set at >235 for high and <200 for low, no significant differences in undergraduate achievement were revealed for the 10 TAs, the 5 treatment TAs, or the 5 control TAs. Apparently the inverse relationship described above was due to the presence of American TAs with assigned SPEAK scores.
<table>
<thead>
<tr>
<th>Number of TAs</th>
<th>High Level</th>
<th>Low Level</th>
<th>F value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate Achievement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>113.87</td>
<td>134.35</td>
<td>33.43</td>
<td>0.001</td>
</tr>
<tr>
<td>(n=175)</td>
<td>(n=163)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8T*</td>
<td>112.31</td>
<td>136.90</td>
<td>35.10</td>
<td>0.0001</td>
</tr>
<tr>
<td>(n=102)</td>
<td>(n=83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8C*</td>
<td>115.42</td>
<td>131.80</td>
<td>7.00</td>
<td>0.009</td>
</tr>
<tr>
<td>(n=73)</td>
<td>(n=80)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* T=treatment TAs; C=control TAs; n= number of undergraduates
Research Question 10: What is the relationship between TAs' scores on tests of written and spoken English and scores on the training program examination?

Once again, for the five foreign and three American TAs in the training program, only the scores on the test of spoken English (SPEAK) were available for use in the analyses for the foreign TAs. A Pearson correlation revealed a non-significant correlation of 0.47 between the training program examination scores and the spoken English scores for the eight treatment using a maximum SPEAK score of 300 for the American TAs. It was therefore concluded that a relationship between these scores could not be found. For the 5 foreign TAs alone, a non-significant relationship, weaker ($r=0.24$) than for all 8 TAs was found.

The ten foreign TAs involved in this study were drawn from a group of thirty-one TAs in the Mathematics and Statistics Departments. As SPEAK scores were available for all of these TAs and TOEFL and GRE-Verbal scores for many of them a Pearson correlation was used to investigate the relationship among the SPEAK, TOEFL, and GRE-Verbal scores.

It was found that GRE-Verbal scores correlated significantly and fairly strongly with SPEAK scores ($r=0.60$, $p=0.008$, $n=18$) and TOEFL scores ($r=0.61$, $p=0.02$, $n=15$). No such strong linear relationship was found between TOEFL and
SPEAK scores ($r=0.34$, $p=0.14$, $n=20$). This suggests that as a proxy for spoken English ability GRE-Verbal may be better than TOEFL scores.

Summary

This chapter has presented the results of the analyses performed as described in Chapter III. Chapter V which follows discusses the implications of these findings.
Chapter V
DISCUSSION, SUMMARY, AND RECOMMENDATIONS FOR FURTHER RESEARCH

Introduction

The purpose of this chapter is to discuss the results of the analyses reported in Chapter IV. The discussion commences with the major research question and proceeds question by question through to question ten. The discussion is followed by summary and recommendations for further research and the chapter ends with a short section titled Concluding Remarks which attempts to relate the findings to the original purposes of the study.

Major Research Question

The principle question of this study concerned the efficacy of a TA training program in producing TAs (treatment group) whose teaching was more effective than untrained (control group) TAs as measured by the achievement of the undergraduate students of the TAs. Designed specifically for novice foreign TAs the program nevertheless included three native
Americans when it was offered during the Summer Quarter of 1983. Only five foreign TAs were enrolled in the program, a number falling far below that anticipated, as a result of circumstances beyond the researcher’s control.

Each of the eight participants in the training program were matched with a graduate student (TA) who had not taken the program. These TA equivalent pairs were matched as closely as possible on ethnic background (e.g., European with European, Asian with Asian, Oriental with Oriental) and spoken English ability. Equivalent pairs were then assigned as TAs in 100 level courses in either Mathematics or Statistics.

Not all pairs could be assigned to a single professor lecturing to the undergraduate students of the TAs in the given pair. This was the case for two of the five pairs of foreign TAs (Pair 5 and 8). The presence of different professors violated the study design, constituting uncontrolled variables and thus causing the results to be treated with extreme caution. Of the remaining equivalent pairs of foreign TAs, analyses showed the undergraduate students of the trained TAs to achieve significantly better for two of the three pairs (Pairs 3 and 6).

An examination of Table 6 in Chapter IV will reveal that Pairs 3 and 6 the untrained TA of the pair had had previous
teaching experience in U.S. classrooms, though none of the TAs had ever received formal training. The same was true for Pair 7 where no significant difference was found. As discussed below, the control TA had a further advantage (if previous experience is an advantage) in terms of spoken English ability. No significant differences were found for the undergraduates of the three pairs of American TAs, none of whom had taught in U.S. classrooms previously.

Any conclusions drawn from these results must be tentative given the low success rate overall. While it is not possible to conclude that the treatment produces more effective TAs in general, it is possible that training contributed to the higher achievement in the two instances described above for foreign TA pairs.

Discussion

It must be remembered that undergraduate achievement was measured on the examinations at the end of Fall Quarter 1983, four months after the conclusion of training. The effect of training as a variable becomes harder to detect as time passes, time becoming an intervening variable. As no reinforcement of the behaviors learned in the training program was offered during the quarter of teaching it is possible for the behaviors to become extinguished early and hence have no impact on the achievement of the students.
Research Question 6, below, threw some light on this possibility.

It is interesting to note that in the two instances where professors constituted an uncontrolled variable (Pairs 5 and 8) three professors were involved, one of them teaching the control group undergraduates of both TA pairs. In both instances the undergraduates of the control TAs achieved higher than those of the treatment TAs although only in one instance was this difference statistically significant. This suggests that different faculty may constitute an important variable in the learning process.

Also of interest is the question of the effect upon undergraduate achievement of the combined variables: faculty member; TA training; and TA spoken English level. As may be seen from Table 6 in Chapter IV, for foreign TA pairs the undergraduate students of the treatment TA achieved significantly better when the faculty variable was controlled and TA spoken English was equivalent for each member of the TA pair (Pairs 3 and 6). When the faculty variable was not controlled (Pair 5) or spoken English was not equivalent but in favor of the non-treatment TA (Pair 7) or both these conditions existed (Pair 8) the undergraduate achievement was either not significantly different or in favor of the non-treatment TA (Pair 8). The data are not sufficient to do other than suggest that the training may
go some way toward nullifying the effect of low spoken English ability but not when coupled with an imbalance in the variable constituting the faculty member lecturing the undergraduates.

The Observation Instrument

Beyond the principle question which attempted to establish causality the remaining questions were exploratory in nature. Research Questions 2 and 3 addressed the construction and testing of the observation instrument while Research Questions 4-10 dealt with further questions involving the TAs and their undergraduate students.

Research Questions 4-10: Conclusions and Discussion

TA Assimilated Knowledge and Undergraduate Achievement

Research Questions 4 and 5 dealt with the extent to which training program participants assimilated knowledge from the training program and the relationship between that knowledge and the achievement of their undergraduate students. As only two of the eight participants failed to score full marks on the training program examination, scoring 41 out of 50, it was considered that in general a cognitive mastery of the lower-inference teacher behaviors had been achieved. Relating TA knowledge to undergraduate achievement was made difficult given the small number of
TAs involved (n=8) and the disproportionate numbers of undergraduates in the high and low cells based on TA knowledge.

It was not possible to find a significant relationship between the two variables using a simple t-test. A more equal division between high and low TA knowledge may have enabled some relationship to be found. Such a division, on the other hand, would have tended to indicate a lack of effectiveness in the training program in equipping the TAs with a knowledge of the behaviors. This would have complicated subsequent analyses, additional difficulties arising from small n in cells reflecting high or low knowledge levels.

**TA Use of Behavior and Undergraduate Achievement**

Research Questions 6 and 7 addressed the extent to which trained and untrained TAs used the lower-inference variables and the relationship of the use to the achievement of the undergraduate students. Analyses revealed differences between the American TA pairs and the foreign TA pairs.

It was found that, in general, statistical differences between treatment and control TAs in the use of behaviors occurred for the American and not for the foreign TA groups. No relationship between TA use of behaviors and undergraduate achievement was found for the foreign TA
group but the use of certain behaviors were found to be related inversely to undergraduate achievement when the American TA group was included in the analyses. Three behaviors (Items 4, 7, and 17) were used more frequently by the trained American TAs (Explains something and then stops so that students can ask questions; Provides time for students to practice; Praises student contributions). Item 2 (Asks questions to find out if students understand) was not used significantly more by trained American TAs and was not related significantly to undergraduate achievement. As may be seen from Table 15, higher undergraduate achievement was associated with more use of Items 3 and 12 (Repeats things on request when students do not understand; Points out what is important for students to learn). The use of these two behaviors was low generally but the relationship is clear.

The American TAs used more of certain behaviors than the foreign TAs. The American TAs also experienced lower undergraduate achievement than the foreign TAs. The question occurs: are the two directly related or is there an intervening variable at work?

The behaviors described as Items 4, 7, and 17 are behaviors involving direct interaction between the TA and the undergraduate. They are behaviors during which the TA either solicits undergraduate contribution (Item 4), explicitly requires undergraduate contribution (Item 7), or re-
sponds as a result of undergraduate contribution (Item 17). It might be that these behaviors constitute a break in the flow of information from the TA to the undergraduate. In this sense, the behaviors constitute interruptions to learning. The implication here is that the undergraduate students do not want "involvement" but "information". The relationship of Items 3 and 12 tends to support this implication, these behaviors being of a direct information-giving kind.

The courses involved in this study were at the 100 level and as such are basic service courses. If the objectives for these courses address factual recall rather than the development of critical thinking or problem solving skills, it may well be that the behaviors discussed above are in fact inappropriate for the level. It must be remembered that the study linking these behaviors to student achievement (Hines, 1981) did so in a laboratory setting.

Alternative explanations exist for the greater use of these behaviors to be associated with lower undergraduate achievement. The relationship between the two variables may be a curvilinear one and the American TAs were using the behaviors too frequently. The American TAs may simply be ineffective teachers for reasons not apparent in this study when compared to the foreign TAs. It is also possible that the American TAs' classes contained a dispropor-
tionate number of weak undergraduate students. This may have occurred accidentally or because these students transferred out of recitation sections led by foreign TAs. Finally, a compensatory mechanism may be in effect in the sections conducted by foreign TAs. The undergraduates there may feel pressed to take remedial action as a result of the TAs' failure to sufficiently bring about the undergraduates understanding of the content. Such failure may result from an inability to communicate clearly through poor spoken English ability, inability to use the lower-inference teacher behaviors, or a combination of both.

The analyses failed to find significant differences in the use of behaviors between trained and untrained foreign TAs except for Item 1 during the first observation. This item, (Answers students' questions) was used significantly more by the non-trained TAs. If this item is seen as a reflection of the need by undergraduates to have the material explained more clearly it may be that experience plays a factor in determining the behaviors used in subsequent weeks. A large number of such questions early in the quarter might cause the TA to be more careful in his or her explanation of the material at later sessions.

Setting this particular item aside, it would appear that training in the use of these behaviors did not for the foreign TAs result in significantly different use of the be-
haviors when compared to the control TAs, even though the training program examination demonstrated cognitive mastery of the behaviors. Again, experience may have caused the control TAs to rapidly adopt these behaviors to a level at least not significantly different from the trained TAs. It is also possible that a less interactive, that is, a more didactic approach may be better for the undergraduates in the given classroom situation.

An alternative explanation is that given the cultural background of many of the foreign TAs the interactive mode of teaching as promoted by the training program may be too alien for the TAs to demonstrate in actual classroom performance. Of the ten foreign TAs it is reasonable to say that at least six were not familiar from experiences in their own country with the interactive mode of teaching.

An equally plausible explanation may be that of the effect of undergraduate feedback upon TA behavior. Behaviors that the undergraduates overtly or covertly reject in the classroom may be subject to extinction especially in the absence of any positive reinforcement from outside of the classroom. Given that interactive behaviors require some fluency with language such behaviors are more likely to be extinguished rapidly for the foreign TAs. One foreign TA offered supporting evidence for this. The TA stated that the TAs tried hard to involve the undergraduate
students but in the face of non-response, temporal constraints, the pressures in their own graduate programs, and no support following training it was very difficult to continue practicing what had been learned.

Analyses showed that this gradual decline in the use of these behaviors was to be found among the trained American TAs who presumably do not have the handicap of speaking in a foreign language. In their case the experience factor, in the form of feedback from the undergraduates, may account for the decline. Nevertheless, for all 16 TAs over the six-week period of observation the overall use of the ten frequency-scaled behaviors was greater for the trained TAs, gradually diminishing as time went on.

The twenty-two behaviors observed consisted of twelve measured using a binary scale and ten using a frequency scale. Many of the former behaviors were used to the same degree by all TAs, either infrequently or frequently over the three observations.

Those behaviors infrequently used (Items 13, 14, 15, 18, 22) represent behaviors that are advanced in the sense of being beyond the mere delivery of content matter. They represent a more sophisticated approach to teaching than that demonstrated by novice TAs, involving the use of advance organizers, summative review, explicit declarations
of intent to involve students, and the use of presentation aids beyond merely chalk and blackboard. The usual argument from novice TAs supporting the non-use of these behaviors is that time constraints preclude them. The frequently used behaviors rated on a binary scale (Items 5, 6, 11, 19, 20, 21) generally related to the actual explanation of content for the given level of instruction and to the TA's voice level and pronunciation.

TA English Ability, Behavior, and Undergraduate Achievement

Research Questions 8 and 9 addressed the relationship between the level of spoken English of the TAs and the variables of TA behavior and undergraduate achievement. Once again the analyses revealed that differences were associated with the presence of native American TAs.

Essentially, the American TAs accounted for the only significant differences in behavior use which was for Items 2 (Asks questions to see if students understand), 7 (Provides time for students to practice), and 17 (Praises student contributions). Greater use of the behaviors was associated with trained TAs and with high spoken English ability when analyses were carried out for all 16 TAs, suggesting that the performance of the behaviors may be a function of both the training and English ability. In order for training to be effective the TA's English ability
may need to be high. When the data for the six American TAs were removed from the analyses no significant differences in use of the individual behaviors were found among the ten foreign TAs taken as a whole or by treatment.

The spoken English level for the foreign TAs was set at greater than 235 and below 201 on the SPEAK scale respectively for "high" and "low" ability. While this decision caused two TAs in the treatment and control groups not to be included in the analyses it did serve to separate the more comprehensible TAs from the less comprehensible TAs.

When the items rated as frequency data were summed overall a significant difference was found in the use of these behaviors among the trained foreign TAs. Only one TA fell into the high ability category and the mean for the behaviors was 78. Two TAs fell into the low ability category for which the cell mean was 43. The division of SPEAK scores here represented the difference between "completely comprehensible" and "generally not comprehensible". The two remaining TAs fell into an intermediate category not addressed in the analysis. Given the small n the results must be interpreted with caution but the suggestion is that a certain level of spoken English is required for the behaviors in this section to be demonstrated more highly.
It is possible that the training these TAs received to use the behaviors was nullified by their spoken English level which effectively prevented them from using the behaviors in the classroom. Among the untrained TAs there were no TAs with SPEAK scores below 200 and so a similar conclusion could not be reached.

Supporting evidence for a relationship between spoken English ability and the use of certain behaviors was provided from the correlational analyses performed. The strong positive relationship existing only for the trained TAs suggests that training may be important in the use of the behaviors in the classroom.

In the case of undergraduate achievement and spoken English ability, the same SPEAK division (high >235, low <201) served to provide approximately the same number of undergraduate students in each cell of the analysis. No significant differences were found in undergraduate achievement for the students of foreign TAs with high or low spoken English ability.

When the American TAs were included in the analyses with an arbitrary SPEAK score of 300 (maximum) certain use of behaviors was found to be related inversely to student achievement. As explained above, the achievement of the undergraduate students of the American TAs was lower than
that of the foreign TAs. It would be unwise to conclude that high spoken English ability is associated with low undergraduate achievement. Furthermore, the American TAs were responsible for the significant and higher use of certain behaviors. As discussed previously, it may be the use or over-use of these behaviors, or the presence of a number of other variables that accounts for this inverse relationship finding. Given that what is being dealt with here, spoken English ability, is only a small part of what impacts upon student achievement it is difficult to reach any firm conclusions.

Different Measures of TA English Ability

Research Question 10 investigated the relationship between different measures of English ability. The Test of English as a Foreign Language (TOEFL) which tests listening comprehension is in general use across the U.S. as a device for assessing the potential ability of foreign TAs to communicate in the classroom. The SPEAK test which is a test of spoken English in which the examinee's pronunciation, grammar, fluency and overall spoken comprehensibility are assessed is now used by a number of universities across the United States. Correlational analyses found a weak, non-significant relationship between these two measures. Given that the two measures address different aspects of English ability, the result should not be surprising. What is more
important is the implication that one measure is no substitute for the other.

A fairly strong, significant relationship was found between the SPEAK scores and scores on the GRE-Verbal test. The latter, unlike TOEFL, is standardized on the American population and it was suggested in the review of the literature that it constituted a possible predictor of classroom overall teaching ability. The findings suggest that it is at least a better proxy for spoken English ability than TOEFL.

Summary of the Discussion of Results

Some evidence was found that training foreign TAs produced more effective TAs when compared with foreign TAs who had not taken the training program. The criteria upon which effectiveness was measured was that of the achievement of the undergraduate students of the TAs on a common end-of-term examination.

Only two pairs of foreign TAs (Pairs 3 and 6), each containing one treatment and one non-treatment TA, were constituted and assigned as recitation instructors in such a way that the design controls were not violated. In both instances the undergraduate students of the treatment TAs performed significantly better at the 0.05 level than those of the non-treatment TAs. The presence of different facul-
ty teaching the undergraduate students of the TAs in a given pair and/or an inequality between the TAs in spoken English ability would appear to have potential for influencing the outcome of instruction in terms of undergraduate achievement and so make comparisons non-productive.

All of the trained TAs demonstrated a cognitive assimilation of the lower-inference behaviors addressed in the training program. Nevertheless, it was found that the assimilated knowledge was not uniformly translated into the demonstrated performance of these behaviors in the classroom.

The trained TAs using the behaviors significantly more often than the non-trained TAs were essentially native Americans who had taken the program with the foreign TAs. The use of certain behaviors of an interactive nature by the American TAs, both trained and untrained, was found to bear an inverse relationship to undergraduate achievement. The trained American TAs used the behaviors more frequently but no significant differences were found between the achievement of their undergraduate students and those of the non-trained American TAs.

The use of behaviors over the six-week observation period gradually declined for the trained American TAs. It would appear that some factor(s) bringing about the extin-
guishing of the behaviors was present. This factor may simply have been a total lack of reinforcement of the behaviors once training had ceased. Such reinforcement may also have prompted the trained foreign TAs to use the behaviors more frequently.

Certain behaviors were used frequently by trained and non-trained TAs and these reflected behaviors closely associated with the processing of content. Behaviors infrequently used by both groups tended to be those that would be associated with a level of teaching that had gone beyond the survival stage. That is, such as giving advance organizers, presenting summative reviews, and preparing material prior to class in the form of handouts or overhead transparencies.

Analyses of Research Question 7 revealed that the use of certain interactive behaviors was related to the interaction between training and spoken English ability. Training was most effective in bringing about high use of the behaviors when the language ability of the TA was high.

Extreme differences in spoken English ability seemed to account for the overall use of certain behaviors by foreign TAs. Those TAs who were categorized as "generally not comprehensible" used the cluster of behaviors far less frequently. As in other analyses the number of TAs involved
was small (n=3) because only two TAs fitted the two categories of spoken English. Correlated analyses using all ten foreign TAs tended to support the finding for only trained TAs. Only the trained foreign TA group had members with very low and very high scores on spoken English.

No relationship was found between the spoken English level of foreign TAs and the achievement of their undergraduate students. The Test of English as a Foreign Language (TOEFL) is frequently used by departments as a means of assessing a foreign graduate student’s level of spoken English. Analyses revealed that the GRE-Verbal examination score would be a better proxy as it had a strong and significant relationship with the TA scores on a test of spoken English (SPEAK).

Recommendations for Further Research

Design Recommendations

Given the small number of TAs involved in this study and difficulties encountered in constructing meaningful cut-off points for analyses of the data, this study must be viewed primarily as an initial investigation or pilot study of the field. The study could be replicated under its present design preferably with a TA population of at least 100. The present study attempted to control, unsuccessfully in some cases, for a number of variables. These included the level
of the course, the faculty involvement in the teaching process, the time and day of the recitation sections observed, and a number of variables peculiar to the TAs of which the equivalent pairs were constituted.

No attempt was made, nor would it have been successful, to control for disparity in the number of students in recitation sections. In many cases the disparity was great leading to large differences in variation of undergraduate examination scores and to disproportionate numbers of undergraduates in cells during certain analyses.

It is strongly recommended that, where possible, data be collected from sections that are more closely equivalent in numbers of undergraduate students and that are taught by one member of faculty in lecture. This would necessitate several hundred students in the lecture situation with multiple recitation sections. Such an occurrence is not unusual in large university settings particularly if the course is a service one.

While scores from a common examination, objectively scored, as in this study can provide some measure of undergraduate achievement, it is suggested that other measures might be adopted. In this present study, in order to perform a number of analyses, it was necessary to ignore the fact that data collected came from examinations in three
different courses. This was considered justifiable in that the courses were of similar nature, scored objectively, and with the same upper and lower limits. With a larger sample, as described above, recourse such as this may not be necessary.

An alternative way of measuring and comparing the achievement of undergraduate students of trained and non-trained TAs would be to collect data on a random sample of undergraduates in courses for which the recitation sections taught by the TAs were prerequisite. Unfortunately there are a number of intervening variables that may make any causal connections between student achievement and TA training and teaching difficult. As suggested earlier, undergraduate students can take compensatory action when confronted by a less effective teaching-learning situation. Such factors are difficult to either measure or control.

Questions Raised and Recommendations

If the training program did indeed cause the undergraduate students of the two foreign TAs to achieve better, the question arises as to exactly what the training did for the TAs in question. If the TAs did not use the behaviors promoted in the training program significantly differently to the non-trained TAs, what other aspects of their effectiveness were changed?
It is possible that the instrument measuring the use of behaviors did not measure the behaviors that were important in the given context. This suggests that there is a need to investigate what behaviors are demonstrated by TAs who are considered effective in recitation sections. This may vary with course level, for example, service courses at the 100 level and courses for subject majors at the 400 level. The needs of the students in these different levels may be such that the behaviors demonstrated by the TA will be different.

The behaviors in this present study are those directed toward the understanding of the content taught. This suggests the development in the students of problem solving skills and an in-depth understanding of the relationship of different parts of the content. Such a suggestion may be quite inappropriate for 100 level courses. Where such a concept is appropriate it is recommended that a tracer technique be applied to the assessment of TA effectiveness. This concept focuses on a specific condition that represents the activities of the subjects under study. If the condition is unlikely to occur it should not be selected and the prevalence rate should be expected to be high enough to permit collection of adequate data from a limited population. Assessment of undergraduate skill in problem solving and application of knowledge is a possible tracer
activity for future research in TA effectiveness when the above criteria are met.

The training program may have had a number of hidden effects on the TAs which subsequently impacted on the effectiveness of their teaching. The simplest effect may have been a Hawthorne effect in which the actual training, the content, was not important but the fact that training was received was important. This effect may manifest itself in a number of ways. The TA may feel more confident, less anxious and so transmit the feeling to the students. It is recommended that attitudinal measures be applied to TAs prior to and following training to examine the changes brought about by training.

It is also recommended that reinforcement of the behaviors learned in training take place during the subsequent teaching. This would provide a means of preventing the behaviors becoming extinguished over time. It would also enable the researcher to detect those instances where the TAs were unable or unwilling to demonstrate learned behaviors. Such an inability, even with reinforcement, may be found to be limited to certain TA variables such as attitudes or level of spoken English.

Further research is required concerning foreign TAs whose spoken English level is very high or very low. The
effect of low spoken English level on student attitudes, learning, and compensatory behaviors and upon the TAs ability to demonstrate learned behaviors needs further research.

Finally, the concepts of recitation teaching and foreign TA training need to be investigated very carefully. It is important to establish exactly what kind of teacher behaviors are important in the recitation classroom and whether or not these are subject to context differences such as content matter and course level. It is equally important to establish the impact of the cultural background of foreign TAs in both their preparation for and performance in classrooms in the United States.

Concluding Remarks

This study resulted from a perceived need to address concerns expressed in the literature regarding the preparation of teaching assistants in general and foreign teaching assistants in particular. A review of the literature identified a number of areas of concern. For TAs in general these included the ability to deliver content in such a way as to promote student understanding in the classroom. Certain lower-inference teacher behaviors purporting to further this end were identified and incorporated into a training program. For foreign TAs in particular, language
and culturally-based factors were identified as possible inhibitors to communication in the classroom. Consequently, a cultural component was incorporated into the training program in order to expose the foreign TAs to the teaching style extant in the United States. The TAs participating in the training were tested for their spoken English ability and were matched as closely as possible on this variable with TAs who did not receive the training. These TA "equivalent pairs" were also matched as closely as possible in terms of ethnic origin.

A number of limitations were voluntarily imposed upon the study. These included those resulting from a quasi-experimental design as TAs were not randomly assigned to the treatment or control group. Additionally, it was not possible to assign all equivalent pairs either to the same professor as TAs in a course or to the same course. Temporal and financial constraints prohibited the observation of TAs in the classroom other than on three occasions during a six-week period. An unanticipated limitation was the small number of participants available for study at the time it was carried out. This rendered some of the proposed analyses extremely difficult to carry out and others open to extremely careful interpretation given the small n. It was necessary to treat the undergraduate achievement scores in the three different courses as from situations sufficiently
similar as to warrant not performing separate analyses by course.

Nevertheless, the study provided some partial answers to questions concerning the training of TAs and their subsequent performance in the classroom. The evidence of success is inconclusive in terms of undergraduate achievement or by the degree of use of the lower-inference behaviors included in the program and later observed in the classroom. Even though cognitive assimilation of the behaviors was demonstrated, the increased use of the behaviors was not uniformly apparent when comparing trained and untrained TAs except by those trained TAs whose spoken English level was high.

It would appear that cultural background and/or spoken English ability may inhibit the practice of learned behaviors in the classroom situation. In the given recitation situation, the use of certain behaviors may be inappropriate. The use of behaviors by the TAs tended to be extinguished over time suggesting that some form of reinforcement is essential if learned behaviors are to be practiced. The study raised a number of questions for further research. These included: the impact of training on the TAs in other than the intended areas of impact; the type of behaviors appropriate for a given level of recitation situation; the impact of TA background culture and spoken Eng-
lish ability upon TA potential for training, and ability to make use of the training in the classroom; and the effect on undergraduate students in the classroom in terms of compensatory behaviors when confronted by a TA who experiences difficulty in communicating.
Appendix A

REPORT OF OBSERVED TEACHER BEHAVIORS

Instructor: ........................................

Course: ............................................

Date: ..............................................

Observer: ........................................

Instructions

You are to use this instrument to report your observations of the classroom teaching behaviors of the instructor named above. You are to report specific, concrete things that the instructor does (or does not do) in his or her classroom teaching. It is important that your responses be "descriptive" rather than "evaluative". Also, your responses should be based solely upon your own personal observations during the course.
Each section of the instrument begins with a heading describing the dimension under which specific teacher behaviors fall. Please estimate the frequency (never, almost never, sometimes, often, very often) with which each of these behaviors is exhibited by the instructor in question by circling only one of the numbers beside each behavior.

**Example**

Here is an example of how to respond. Suppose Item 3 reads:

The Instructor:

3. Repeats things when students do not understand.

Think! How often does (did) your instructor do this during the course? Choose one of the answers below from the answer scale:

1 = never

2 = almost never

3 = sometimes

4 = often

5 = very often
If your best answer happens to be "often", then for Item 3 on your sheet place a circle around the number 4 following the statement, as shown below.

3. Repeats things when students do not understand.............. 1 2 3 4 5

The data you provide using this instrument will be shared with the instructor in summary form only. Individual responses will not be shared with the instructor and will remain entirely confidential. Data will be used for the purposes of research on university teaching conducted by Brian K. Davis of the College of Education of The Ohio State University. The names of observers will never appear in any kind of report on this research.

THE INSTRUCTOR

I. Provides for Student Understanding and Assimilation of Instructional Content:

1. Answers students' questions............. 1 2 3 4 5
2. Asks questions to find out
   if students understand................. 1 2 3 4 5
3. Repeats things when students do not understand.................. 1 2 3 4 5
4. Explains something and then stops
so that students can ask questions......1 2 3 4 5
5. Explains things simply ......................1 2 3 4 5
6. Teaches at a pace appropriate to
   the topic and to the students..............1 2 3 4 5
7. Gives students enough time for
   practice........................................1 2 3 4 5
8. Gives students a chance to think
    about what has been taught................1 2 3 4 5
9. Stays with the topic until
    students understand...........................1 2 3 4 5
10. Shows similarities and differences
    between things................................1 2 3 4 5

II.
Explanation/Demonstrates How to Do the Work
by Use of Examples

11. Uses examples when explaining ...........1 2 3 4 5
12. Works examples (e.g., on the
    chalkboard) and explains them............1 2 3 4 5
13. Shows students how to do the
    work by use of examples....................1 2 3 4 5
14. Teaches step-by-step ......................1 2 3 4 5

III.
Structure Instruction and Instructional Content/Presents
Content/Presents Content in a Logical Sequence:
15. Presents content in a logical manner 1 2 3 4 5
16. Points out what is important
   for students to learn 1 2 3 4 5
17. Informs students of course/lesson
   objectives 1 2 3 4 5
18. Tells students what they are expected
   to know or should be able to do on
   on completion of instruction 1 2 3 4 5
19. Summarizes the material presented
   in class 1 2 3 4 5
20. Distributes time adequately over
    the topics covered 1 2 3 4 5

IV. Establishes Rapport with the Students:

21. Addresses students by name 1 2 3 4 5
22. Praises student contributions 1 2 3 4 5
23. Encourages student participation 1 2 3 4 5
24. Looks at the class when teaching 1 2 3 4 5

V. Uses Communication Skills:

25. Pronounces words clearly 1 2 3 4 5
26. Can be clearly heard in all
    parts of the classroom 1 2 3 4 5
27. Makes use of a variety of teaching
aids (e.g., blackboard, overhead
projector, slides, handouts)...........1 2 3 4 5

Appendix B

REPORT OF OBSERVED TEACHER BEHAVIORS:

Behavioral definitions for items

I. Provides for Student Understanding and Assimilation of Instructional Content:

1. Answers students' questions.

The teacher answers content-related questions asked by student(s).

2. Asks questions to find out if students understand.

The teacher asks questions about the content taught in order to find out if students understand what has been said (taught). The question must not be rhetorical. The student does not have to respond although it should be clear that the teacher expects an answer.

The teacher may initiate this behavior or it may occur in response to a student's question or comment or some non-verbal cue from students indicating that they do not understand.
3. **Repeats things when students do not understand.**

The teacher repeats aspects of the content of instruction (previously addressed) which students directly communicate to the teacher that they do not understand, e.g. "Would you repeat that?"; "I don't understand that"; etc.

4. **Explains something and then stops so that students can ask questions.**

The teacher, after explaining, repeating, or reviewing some aspects of content or responding to a student's question or comment deliberately stops and provides time for students to ask questions about the content of instruction. The teacher may say "Okay?" or "Do you have questions?", prior to pausing.

5. **Explains things simply**

The teacher makes a visible effort to reduce material to a level at which the student will grasp the point rather than present the student with a statement without explicating its meaning. The teacher may say, "This means..."; "This comes down to..."; "So what we have is simply..."; "This is what counts...".

6. **Teaches at a pace appropriate to the topic and to the students.**
The teacher presents material such that his or her speed of delivery provides time for the average student to take notes (including from the blackboard or overhead projector) and ask questions about the material.

7. **Provides time for students to practice** (e.g., work problems).

   The teacher, during the class period, provides specific time for students to do written or practical assignments related to the content of instruction. This may take the form of individual or group work. The teacher plays an active (leading) role in the case of group work (i.e., the teacher works examples with student contribution).

8. **Gives students a chance to think about what has been taught.**

   The teacher explains some aspects of the content of instruction and then deliberately pauses to provide time for students to think about what has been said. The teacher, after explanation, explicitly tells the students that he/she is providing time to think about what was said.

9. **Shows similarities and differences between things.**

   The teacher describes, explains or shows how two or more things (e.g., ideas, concepts, objects, ways of doing things, etc.) are alike and/or how they differ.
II. Explains/Demonstrates How To Do The Work by Use of Examples

10. Shows students how to do the work by use of examples.

The teacher presents students with written examples (in the form of a handout, transparency, or blackboard work) and shows the students how to work them with or without student involvement.


The teacher has the content matter of the lesson sequenced in such a way that steps within a problem, between parts of a problem, between parts of a lesson, or the development of course content build only upon previously covered material; a smooth transition from one part of the lesson to another; no fumbling or hesitation.

III. Structures Instruction and Instructional Content/Presents Content in a Logical Sequence:

12. Points out what is important for students to learn.

The teacher deliberately draws students' attention to those aspects of the content of instruction it is important for them to learn. The teacher may say, for example, "It is important for you to know this..."; "You must understand this..."; "The rule to learn is..."; "Remember, the important point is...".

13. Informs students of the course/lesson objectives.
The teacher, at the beginning of the course or start of the lesson indicates the content to be covered. The teacher may say, "What I will go over..."; "The topic today is..."; "The purpose of our work today is..."; etc.

14. **Tells students what they are expected to know or should be able to do on completion of instruction.**

The teacher may say, "Now you will be able to work on the examples following chapter six"; "This will enable you to work out...".

15. **Summarizes the material presented in class.**

The teacher, on completion of the lesson presentation, gives a summary of the instructional content presented in the lesson.

IV. **Establishes Rapport with the Students:**

16. **Addresses students by name.**

The teacher, during the class, refers to the student by name when addressing the student or responding to a comment or question from the student.

17. **Praises student contributions.**

The teacher acknowledges that he or she values student contributions by making some agreeable verbal response to the contributions. The teacher may say, "That's a good point"; "Yes. Thank you."; "Thanks for raising that."; "Right. Well done."; etc.
18. **Encourages student participation.**

The teacher specifically makes statements designed to promote student involvement in the lesson. The teacher may say, "Feel free to interrupt with questions or comments." or "I'd like you to provide me with some of the answers as we work this problem." etc.

19. **Looks at the class when teaching.**

The teacher generally has his or her face turned toward the students taking in all parts of the room, except when writing on the blackboard or overhead projector, at which times the teacher periodically looks toward the students.

V. **Spoken Word and Aids**

20. **Pronounces words clearly.**

The teacher pronounces words in such a way that the student is unlikely to be in doubt as to what the word is; pronunciation difficulties are infrequent and minimal.

21. **Can be clearly heard in the classroom.**

The teacher modulates his or her voice level such that, irrespective of the room size, all students have an opportunity to hear clearly what is said by the teacher.

22. **Makes use of a variety of teaching aids.**

The teacher, during the lesson, makes use of more than one means of presenting content visually to the students.
teacher may use the blackboard and then hand out material for the students to work on; hand out material and then discuss the material using the blackboard or illustrate some part of the content through use of a slide projector or transparency as well as using the blackboard.

Appendix C

REPORT OF OBSERVED TEACHER BEHAVIORS

Instructor............................................

Course..................................................

Date...................................................

Observer.............................................

Instructions

You are to use this instrument to report your observations of the classroom teaching behaviors of the instructor named above. You are to report specific, concrete things that the instructor does (or does not do) in his or her classroom teaching. It is important that your responses be "descriptive" rather than "evaluative". Also, your responses should be based solely upon your own personal observations during the course.
Each section of the instrument begins with a heading describing the dimension under which specific teacher behaviors fall. Please estimate the frequency (never, almost never, sometimes, often, very often) with which each of these behaviors is exhibited by the instructor in question by circling only one of the numbers beside each behavior.

Example

Here is an example of how to respond. Suppose Item 3 reads:

The Instructor:

3. Repeats things when students do not understand.

Think! How often does (did) your instructor do this during the course? Choose one of the answers below from the answer scale:

1 = never
2 = almost never
3 = sometimes
4 = often
5 = very often
If your best answer happens to be "often", then for Item 3 on your sheet place a circle around the number 4 following the statement, as shown below.

3. Repeats things when students when students do not understand

The data you provide using this instrument will be shared with the instructor in summary form only. Individual responses will not be shared with the instructor and will remain entirely confidential. Data will be used for the purposes of research on university teaching conducted by Brian K. Davis of the College of Education of The Ohio State University. The names of observers will never appear in any kind of report on this research.

I.

THE INSTRUCTOR

I.

Provides for Student Understanding and Assimilation of Instructional Content:

1. Answers students' content-related questions. 1 2 3 4 5

2. Asks questions to find out if students understand content. ............... 1 2 3 4 5

3. Repeats things on request when students do not understand. ............... 1 2 3 4 5
4. Explains something and then stops
   so that students can ask questions...........1 2 3 4 5

5. Explains things simply........................1 2 3 4 5

6. Teaches at a pace appropriate to
   the topic and to the students...............1 2 3 4 5

7. Gives students enough time for practice
   (involves them in examples being worked)....1 2 3 4 5

8. Gives students a chance to think about
   what has been taught (stops and says,
   "Now think about this")......................1 2 3 4 5

9. Shows similarities and differences
   between things...............................1 2 3 4 5

II.
   Explains/Demonstrates How to do the Work by
   Use of Examples

10. Shows students how to do the work by
    use of examples..............................1 2 3 4 5

11. Teaches step-by-step..........................1 2 3 4 5

III.
   Structures Instruction and Instructional
   Content/Presents Content in a Logical Sequence:

12. Points out what is important for students
to learn.................................1 2 3 4 5

13. Informs students of course/lesson objectives
   (i.e., material to be covered).............1 2 3 4 5

14. Tells students what they are expected to know or
   able to do on completion of instruction....1 2 3 4 5

15. Summarizes the material presented in class.1 2 3 4 5

IV.
Establishes Rapport with the Students:

16. Addresses students by name................1 2 3 4 5

17. Praises student contributions
    (e.g., says, "Good. Right.").............1 2 3 4 5

18. Encourages student participation (e.g. says, "I want
    you to get involved in this solution")....1 2 3 4 5

19. Looks at the class when teaching...........1 2 3 4 5

V.
Voice and Aids

20. Pronounces words clearly...................1 2 3 4 5

21. Can be clearly heard in all
    parts of the classroom...................1 2 3 4 5
22. Makes use of a variety of (more than one) teaching aids (e.g. blackboard, overhead projector, handouts).........................1 2 3 4 5

Appendix D

REPORT OF OBSERVED TEACHER BEHAVIORS: OBSERVERS' INSTRUMENT

Date:..... Time:.... Place:.... TA:.... Observer:........

13. Informs students of the lesson/course objectives
(explicit statement introducing what will take place)------------------------0 1

14. Tells student what they are expected to know or should be able to do on completion of lesson
(explicit statement from teacher).........................0 1

1. Answers students' questions
(content related).........................0 1 2 3 4 5 6 7 8 9 10

2. Asks questions to find out if students understand
(not rhetorical questions; student answers expected).........................0 1 2 3 4 5 6 7 8 9 10

7. Provides time for students to practice
(teaching involves students in doing the work.........................0 1 2 3 4 5 6 7 8 9 10

16. Addresses students by name. . . 0 1 2 3 4 5 6 7 8 9 10
17. Praises student contributions
   (verbal praise, more than a
   flat "yes" or "okay")
   0 1 2 3 4 5 6 7 8 9 10

3. Repeats things when students
do not understand.
   0 1 2 3 4 5 6 7 8 9 10

4. Explains something and then stops so that
   students can ask questions (asks for student
   feedback and waits)
   0 1 2 3 4 5 6 7 8 9 10

8. Gives students a chance to think about what is
taught (explicitly states giving
   time to think)
   0 1 2 3 4 5 6 7 8 9 10

9. Shows similarities and differences between things
   (compares/contrasts, points out similarities/
   differences)
   0 1 2 3 4 5 6 7 8 9 10

10. Shows how to do the work by use of examples
   (works examples with or without student
   involvement)
   0 1

18. Encourages student participation
   (explicit statement indicating student
   involvement expected)
   0 1

5. Explains things simply (teacher statements that
   show attempt to simplify)
   0 1

6. Teaches at a pace appropriate to the topic and
students (consider the average student taking
notes/asking questions) ............................................ 0 1

11. Teaches step by step (smooth transition from one
part of lesson to another; no fumbling or
hesitation) .......................................................... 0 1

12. Points out what is important for students to
learn (explicit statements drawing students'
attention to certain
parts of the material) ............... 0 1 2 3 4 5 6 7 8 9 10

15. Summarizes the material presented in class
(gives some concluding remarks at the end
of the lesson that summarize the main points) ....... 0 1

19. Looks at the class when teaching (teacher pays
attention generally to all parts of the room) .......... 0 1

20. Pronounces words clearly (minimal, infrequent
pronunciation difficulties) ................................. 0 1

21. Can be clearly heard in all parts of the
classroom (can be heard wherever the students
are sitting) .......................................................... 0 1

22. Makes use of a variety of teaching aids (uses
more than one visual aid such as board,
handout, transparency) ........................................ 0 1
BIBLIOGRAPHY


Buckenmeyer, J.A. Preparing graduate assistants to teach. *Improving College and University Teaching*, 1972, 20, 142-144.


Flanigan, M.C. Variety, the key to training programs for faculty and teaching assistants. ADE Bulletin, 1978, 27, 44-46.


Jossef, E.L. Personal communication. The Ohio State University, 1983.


McKnight, P.C. and Bush, R.N. *Microteaching: A selected bibliography,* 1977, 6. (ERIC Document Reproduction Services No. ED 141 266)

Medley, D.N. *Teacher competency testing and the teacher educator.* Charlottesville: Association of Teacher Educators and the Bureau of Educational Research, University of Virginia, 1982.


The Graduate Voice. TA Recommendations UPDATE. The Ohio State University, Spring 1983, 5, 1, 4 & 10.

The Lantern. TA Forum: Students suggest instruction changes. The Ohio State University, March 1, 1982 (a), page 1.

The Lantern. Students ask for more TA training. The Ohio State University, April 26, 1982 (b) pages 1 & 3.

The Lantern. Guidelines for OSU TAs. The Ohio State University, April 30, 1982 (c) page 4.

The Lantern. Stronger program for training TAs receives support of Jennings, staff. The Ohio State University, May 12, 1982, (d) page 1.


Williams, J.E. The Stability of Teacher Clarity in Relation to Student Achievement and Satisfaction. Doctoral dissertation, (unpublished), The Ohio State University, 1983.


