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.
Myers, Christopher John

SPECIAL EDUCATOR IMPROVEMENT PROCEDURE (SEIP): AN INNOVATIVE APPROACH TO SPECIAL EDUCATION TEACHER EVALUATION FOR THE IMPROVEMENT OF INSTRUCTION

The University of Arizona

University Microfilms International

300 N. Zeeb Road, Ann Arbor, MI 48106

Ph.D. 1983
SPECIAL EDUCATOR IMPROVEMENT PROCEDURE (SEIP): AN
INNOVATIVE APPROACH TO SPECIAL EDUCATION TEACHER
EVALUATION FOR THE IMPROVEMENT OF INSTRUCTION

by

Christopher John Myers

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1983
As members of the Final Examination Committee, we certify that we have read
the dissertation prepared by Christopher John Myers
entitled Special Educator Improvement Procedure (SEIP): An
Innovative Approach to Special Education Teacher
Evaluation for the Improvement of Instruction

and recommend that it be accepted as fulfilling the dissertation requirement
for the Degree of Doctor of Philosophy.

Date

Date

Date

Date

Final approval and acceptance of this dissertation is contingent upon the
candidate's submission of the final copy of the dissertation to the Graduate
College.

I hereby certify that I have read this dissertation prepared under my
direction and recommend that it be accepted as fulfilling the dissertation
requirement.

Dissertation Director
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Christopher John Myers
ACKNOWLEDGMENTS

I would like to express my appreciation to the members of my dissertation committee, Glen Nicholson, Keith Meredith, Sarah Dinham, Bill Healey, and Jeanne McCarthy, not only for their assistance in the preparation of my dissertation but for their support throughout my doctoral program. In particular, I would like to thank Glen Nicholson for guiding me through the final phase of my doctoral career.

I would also like to thank Pat Dean, Paige Stone and Dave Conroy, who played instrumental roles in the implementation of this study. Their perseverance contributed greatly to the successful completion of the study.

Additionally, I would like to acknowledge the Special Educator Improvement Procedure (SEIP) team members, who unselfishly contributed their time to this project. I am especially appreciative of the experimental subjects' willingness to be videotaped on numerous occasions throughout the study.

Most of all, I want to thank my wife, Nancy, who not only assisted in the development of the SEIP, but provided me with the emotional support that I needed to complete the project.
# TABLE OF CONTENTS

**LIST OF TABLES** .................................................. vi

**LIST OF ILLUSTRATIONS** ........................................ viii

**ABSTRACT** .......................................................... ix

1. THE PROBLEM AND RELATED LITERATURE .......................... 1

   Introduction ..................................................... 1
   Need for Teacher Evaluation .................................... 1
   Purposes of Teacher Evaluation ................................ 4
   Reactions to Teacher Evaluation ................................ 6
   Teacher Evaluation Effectiveness ................................ 7
   Problems in Evaluating Teacher Performance .................. 9
   School Personnel Support for Improving Teacher Evaluation 
       Teacher Evaluation ........................................ 11
   Evaluation of Special Educators versus Regular Educators .... 12
   How to Evaluate Teachers ....................................... 15
       Observation .................................................. 16
       Assessing Student Outcomes ................................ 20
       Teacher Appraisal Conference ............................... 23
   Who Should Evaluate ............................................. 24
       Self Evaluation ............................................. 24
       Supervisor Evaluation ...................................... 27
       Peer Evaluation .............................................. 28
       Consumer Evaluation ........................................ 30
       Evaluation by Others ......................................... 31
       Multiple Evaluators ......................................... 32
   Evaluation Tools .................................................. 34
       Videotape Recorder .......................................... 34
       Observation Instruments ..................................... 37
   Promising Approaches to Teacher Evaluation for Improvement .. 41
   Summary of the Research on Teacher Evaluation ............... 46
   Need for Additional Research on Teacher Evaluation .......... 50
   Purpose of the Current Study ................................... 51
   Hypotheses ....................................................... 53
TABLE OF CONTENTS--Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. METHOD</td>
<td>56</td>
</tr>
<tr>
<td>Design</td>
<td>56</td>
</tr>
<tr>
<td>Subjects</td>
<td>56</td>
</tr>
<tr>
<td>Dependent Variables</td>
<td>61</td>
</tr>
<tr>
<td>Independent Variable</td>
<td>64</td>
</tr>
<tr>
<td>Independent Observers</td>
<td>67</td>
</tr>
<tr>
<td>Pretreatment Screening</td>
<td>68</td>
</tr>
<tr>
<td>Pilot Study</td>
<td>70</td>
</tr>
<tr>
<td>Analysis</td>
<td>71</td>
</tr>
<tr>
<td>3. RESULTS</td>
<td>74</td>
</tr>
<tr>
<td>Revusky's R Statistic</td>
<td>74</td>
</tr>
<tr>
<td>A Priori Contrasts</td>
<td>77</td>
</tr>
<tr>
<td>SEIP Participant Questionnaire Analysis</td>
<td>80</td>
</tr>
<tr>
<td>4. DISCUSSION</td>
<td>86</td>
</tr>
<tr>
<td>Analysis of Statistical Test Results</td>
<td>86</td>
</tr>
<tr>
<td>Analysis of SEIP Participant Questionnaire Results</td>
<td>93</td>
</tr>
<tr>
<td>Responses to Quantitative Questions</td>
<td>94</td>
</tr>
<tr>
<td>Responses to Qualitative Questions</td>
<td>95</td>
</tr>
<tr>
<td>Appropriateness of the SEIP for Regular Educators</td>
<td>98</td>
</tr>
<tr>
<td>Implications for Future Research</td>
<td>99</td>
</tr>
<tr>
<td>Implications for Teacher Evaluation</td>
<td>102</td>
</tr>
<tr>
<td>APPENDIX A: DESCRIPTION OF THE STUDY FOR PARTICIPANTS</td>
<td>104</td>
</tr>
<tr>
<td>APPENDIX B: TEACHER PERFORMANCE ASSESSMENT INSTRUMENTS--REVISED (TPAI-R)</td>
<td>105</td>
</tr>
<tr>
<td>APPENDIX C: SPECIAL EDUCATOR IMPROVEMENT PROCEDURE (SEIP) PARTICIPANT QUESTIONNAIRE</td>
<td>113</td>
</tr>
<tr>
<td>APPENDIX D: SEIP SCRIPT</td>
<td>115</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>121</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table | Page
--- | ---
1. Job Titles of SEIP Feedback Team Members | 58
2. Percentage of SEIP Participants' TPAI-R Ratings above the Minimum Competency Level | 69
3. Mean Ratings of Subjects' Teaching Performance on TPAI-R | 75
4. Transformed Mean Ratings of Subjects' Teaching Performance on TPAI-R | 75
5. Rank of Each Subject's TPAI-R Transformed Mean Rating for Week Following SEIP Feedback Session | 77
6. A Priori Contrasts Using Tukey's HSD Procedure for Week 2 | 78
7. A Priori Contrasts Using Tukey's HSD Procedure for Week 3 | 78
8. A Priori Contrasts Using Tukey's HSD Procedure for Week 4 | 79
9. A Priori Contrasts Using Tukey's HSD Procedure for Week 5 | 79
10. A Priori Contrasts Using Tukey's HSD Procedure for Week 6 | 80
11. A Priori Contrasts Using Tukey's HSD Procedure for Week 7 | 80
12. Frequency of Responses to First Quantitative Question on SEIP Participant Questionnaire: "In General, How Beneficial Do You Feel the SEIP Feedback Session is For Helping Special Educators Improve Their Instruction?" | 83
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Frequency of Responses to Second Quantitative Question on SEIP Participant Questionnaire: &quot;How Beneficial to You Was Your Participation in the SEIP Feedback Session?&quot;</td>
<td>84</td>
</tr>
<tr>
<td>14. Frequency of Responses to Third Quantitative Question on SEIP Participant Questionnaire: &quot;How Adequate is the Revised Teacher Performance Assessment Instruments (TPAI-R) for Assessing Special Educators' Teaching?&quot;</td>
<td>85</td>
</tr>
<tr>
<td>15. Mean TPAI-R Ratings by Subjects and by Other Feedback Team Members</td>
<td>91</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Multiple Baseline Design for Implementing SEIP</td>
<td>57</td>
</tr>
<tr>
<td>2.</td>
<td>Subjects' Mean TPAI-R Ratings for Each Week of Study</td>
<td>76</td>
</tr>
</tbody>
</table>

viii
ABSTRACT

The purpose of this study was to investigate the effectiveness of an evaluation procedure designed to assist special educators improve their instruction. A single subject, multiple baseline across subjects design was used to test this evaluation procedure, the Special Educator Improvement Procedure (SEIP), with six teachers of self-contained special education students over a seven week period. The SEIP was composed of the following components: (1) presenting multiple videotaped samples of the subject's teaching performance to a team of evaluators, including the subject, the subject's supervisor, a peer and another educator chosen by the subject; (2) having the evaluation team assess the subject's instruction using the revised Teacher Performance Assessment Instruments (TPAI-R); (3) allowing the subject to compare his or her self assessment with the assessments of other evaluation team members; and (4) having team members make specific suggestions as to how the evaluatee might improve his or her instruction.

Following the seventh week of the study, two independent observers used the TPAI-R to rate videotaped samples of the subjects' teaching. These ratings were
used to test the prediction that subjects who participated in the SEIP would be ranked higher on the TPAI-R than would subjects who had not yet participated in the SEIP. However, the results of statistical analyses using Revusky's $R_n$ statistic and Tukey's HSD procedure did not support this prediction.

At the conclusion of the study, evaluation team members completed an experimenter developed questionnaire. An analysis of their responses supported the predictions that (1) the participants would rate the SEIP as being very beneficial for helping special educators improve their teaching and (2) the TPAI-R was a good instrument for assessing the adequacy of special educators' teaching, but failed to support the prediction that the participants would rate their participation in the SEIP feedback session as being very beneficial.

It was concluded that the results of the study were inconclusive regarding the effectiveness of the SEIP in assisting special educators to improve their instruction. It was suggested that additional research be conducted using a revised version of the SEIP.
CHAPTER 1

THE PROBLEM AND RELATED LITERATURE

Introduction

The current status of teacher evaluation, in general, and formative special education teacher evaluation, in particular, was examined through this research project. A review of pertinent research highlights the strengths and weaknesses of teacher evaluation techniques and instruments presently being used. It also points out the need for the development and empirical validation of effective evaluation procedures that can help teachers improve their instruction. Following this review, an experimental research study, which was designed to examine the effectiveness of an innovative formative special education teacher evaluation procedure, is described.

Need for Teacher Evaluation

During this time of shrinking resources, the public is demanding that schools demonstrate educational and fiscal accountability (Sapone, 1980). Pressure is being applied to school personnel at every level to perform optimally and to show that public education has a beneficial impact upon students. Parents no longer
unquestioningly accept the education that the schools provide for their children. They want assurances that the educational needs of their children are being met. The parents of handicapped children, in particular, have become more vocal as a consequence of the rights afforded them under P.L. 94-142. Should the education that their children receive be deemed inappropriate, it is possible that they may begin initiating liability lawsuits against teachers for failure to adequately teach their children (Berdine, Cegelka & Kelly, 1977).

In response to this accountability movement, schools are putting more effort into developing and implementing accountability systems and procedures. Teacher evaluation, in particular, is receiving increased attention because it is seen as one way that schools can assess how well they are educating students and then make changes for improving instruction based upon this assessment. As Kemerer (1979) noted: "The accountability movement has special significance for education, because education is an expensive labor-intensive operation. Only by gathering and analyzing data on effectiveness of personnel can principled decision making to achieve the greatest institutional efficiency take place in the face of shifting resources" (p. 25).

This emphasis on accountability and evaluation is felt even more strongly among the ranks of special
educators than among teachers of non-handicapped students. As Moyapointed out, "At a time when the need for special teachers is at an all time high, evaluation becomes even more crucial...[T]he feeling that a live body in the classroom is better than none cannot be tolerated; excellence must be encouraged" (1980, p. 40). In addition to satisfying the general public, special education teachers have other masters to serve. Because most special education programs receive public and private funding over and above the funding received for educating non-handicapped students, special educators must also demonstrate their competence to the satisfaction of the agencies that control the allocation of these funds (Soeffing, 1975).

The accountability movement and agency requirements are not providing the only pressure for teacher evaluation. Many states, including Arizona, California, Connecticut, Florida, Idaho, Kansas, New Jersey, North Carolina, Oregon, Pennsylvania, South Dakota, Virginia, and Washington, have enacted laws requiring that state educational agencies and local school districts evaluate professional school personnel. For example, for teachers in Georgia to be fully certified, they must demonstrate to the satisfaction of a team of evaluators that they possess 14 important teaching competencies. In California, the Stull Bill requires that each school district evaluate its teachers and recommend
areas for improvement. The 1974 Arizona Legislature enacted a System of Assessment and Evaluation, which states:

There shall be a system of assessment and evaluation of the performance of certificated teachers within each school district of the state which shall involve the development and adoption by each school district of objective assessment and evaluation guidelines for the improvement of instruction (Ariz. Rev. Stat. Ann. §15-268).

**Purposes of Teacher Evaluation**

Two major reasons are given for evaluating teachers. First, teacher evaluation can be used to help teachers improve their classroom performance or instruction (formative evaluation). Secondly, teacher evaluation can be used for making administrative decisions, such as renewal of contracts, granting of tenure, distributing merit pay, making assignments and dismissals (summative evaluation). Of the two, improvement of instruction is generally considered the most important function of teacher evaluation (Allen, Barnes, Reece, & Roberson, 1970; Hall, 1980; Herman, 1973; McIntyre, 1980; McKenna, 1981).

Because the ultimate goal of education is to facilitate the growth and development of students, it is important that evaluation be used to help teachers be more effective in serving students' needs. Evaluation systems which do not provide a means for assisting teachers in improving their instruction are not living up to their full potential (Kemerer, 1979).
Evaluation can be used in a number of ways to help teachers improve their instruction. First, evaluation can provide teachers with feedback regarding how well they are fulfilling their teaching responsibilities. This information can then, in turn, be used by the teachers as a basis for making adjustments in their teaching. Secondly, the information gathered during the evaluation process can be used to develop inservice training that will assist teachers in remediating weaknesses (Herman, 1973). Third, evaluators can often make useful suggestions regarding what should be taught and how it should be taught. And finally, the evaluation process itself can motivate teachers to optimize their job performance. Knowing that their performance will be assessed can serve as an incentive for teachers to do their best.

Although there is general agreement that teacher improvement is the most important purpose of evaluation, there are conflicting opinions as to the degree to which evaluation is actually used for this purpose. On the one hand, the results of a survey conducted by the Educational Research Service (1978) showed that 96% of the responding school districts used evaluation for improvement purposes. On the other hand, Johnston and Yeakey (1979) have reported that administrators most often use evaluation for job oriented purposes, such as granting of tenure, rather than for professional development and improvement. While these
two positions may represent true differences in research findings, they may also indicate a difference between what schools say they do and what they actually do.

Reactions to Teacher Evaluation

In contrast to the support given to teacher evaluation in general, the reactions of evaluators, evaluatees and the general public to current evaluation practices have generally been negative. Evaluation often raises feelings of anxiety, distrust and resentment among teachers. They fear that they may be assessed unfairly due to the vague and ambiguous nature of many evaluation instruments and procedures and the ineptness demonstrated by evaluators. They may feel a need to defend themselves rather than seeking to improve their skills (Brown, 1976; Feldvebel, 1980; Johnston & Yeakey, 1979; McNeil, 1981; Wolf, 1973). However, Bolton (1973) reported that the acceptability of evaluation increases considerably when teachers view it as a means to improve instruction rather than as a basis for making administrative decisions.

Supervisors who are charged with the responsibility of evaluating teachers often feel uncomfortable with evaluation as well. Their administrative training rarely includes coursework or practical experience in evaluating teachers (Educational Research Service, 1978). Consequently, they are forced to stumble along using trial and error methods, which are likely to be of little help to teachers.
Teacher Evaluation Effectiveness

A considerable volume of research has been conducted on teacher evaluation. Educators have devoted thousands upon thousands of hours developing teacher evaluation systems, procedures and instruments. An even greater amount of time has been consumed in the evaluation process and followup activities. However, despite this tremendous investment of time and effort, the effectiveness of both regular education and special education teacher evaluation can be summed up simply: it is not effective (Kauffman, Hallahan, Payne & Ball, 1973; Wolf, 1973). Scriven's assessment of the state of teacher evaluation is even more ungracious. He states: "Teacher evaluation is a disaster. The practices are shoddy and the principles are unclear... [O]ne cannot yet point to a single exemplary system..." (1981, p. 244).

Teacher evaluation is not effective in its primary functions of facilitating instructional improvement and administrative decision making. The fact that teacher evaluation has not adequately assisted school administrators in making crucial personnel decisions is primarily due to the failure of evaluation procedures and instruments to reliably discriminate between effective and ineffective teachers. These procedures and instruments often do not "...sort out or recognize obvious distinctions in performance among teachers apparent to administrators, teachers
themselves, students, parents, and anyone else familiar with the operation of a particular school" (Newton, 1980, p. 45).

Evaluation procedures directed toward helping teachers to improve their instruction have proven to be equally inadequate. As pointed out by Feldvebel: "There is research evidence that ratings and written evaluations in their present form are of limited use in offering supervisory help to teachers for improving instruction. Only about one-third of the teachers who are rated or evaluated report any changes resulting from the written evaluations" (1980, p. 416). Peterson and Ward (1980) and Moya (1980) concur with this assessment.

The failure of the professional literature to support the effectiveness of teacher evaluation practices and instruments may be partially due to the dearth of empirical research in this area. In comparison to the huge amount of resources put into evaluating teachers and into the development of evaluation instruments and procedures, the time and effort devoted to research concerned with assessing the effectiveness of teacher evaluation systems is miniscule. Few studies have investigated the reliability, validity and effect of many commonly used techniques in teacher evaluation (Levin, 1979). None of the studies herein reviewed provided empirical evidence validating the effectiveness of a particular procedure or technique for improving instruction.
Problems in Evaluating Teacher Performance

Teacher evaluation has been plagued by a number of difficulties that have limited its effectiveness and precipitated some of these negative reactions. Perhaps the greatest hurdle to effective teacher evaluation has been the lack of agreement regarding what qualities and behaviors should be demonstrated by a good teacher. The majority of educators can identify poor teachers, but consensus seems to be lacking in identifying excellent teachers. This problem is common to regular educators and special educators alike. As Semmel (1978) pointed out, "We do not know which teaching behaviors are causally related to the growth of handicapped pupils. It is doubtful that special educators could get any closer than regular educators to achieving a consensus on who is and who is not a good teacher" (p. 27). This dilemma, in turn, has left many teachers and evaluators to wonder: how can teachers be evaluated with validity if the experts cannot agree what an effective teacher should be like? McDonald has summed up this sentiment in stating that because there is not agreement as to what is effective teaching, "...our evaluations of teaching effectiveness will be imprecise, have modest validity, and be open to critical and frequently antagonistic debate" (1980a, p. 73).

A second problem that hinders the implementation of effective evaluation practices concerns the context in
which evaluation occurs. Situational variables over which a teacher has little control, such as the physical environment, the curriculum, and the previous training of the students, may undermine the teacher's performance. Such factors need to be considered and weighed by evaluators so that teachers will be assessed equitably.

A third problem area is the incompatibility of different roles the evaluator may often assume. On the one hand, the evaluator may need to make administrative decisions regarding the continuance of a particular teacher, while on the other hand, the evaluator may want to help the teacher improve his or her instruction (Educational Research Service, 1978). In trying to assume both roles, the evaluator may fail to be effective in either role.

Fourth, practical considerations may hinder effective evaluation. Sufficient effort may not be put into evaluation because of the substantial amount of time that it requires and because of the limitations sometimes created by teacher union contracts (McDonald, 1980b).

And finally, evaluation practices often are not consistent with those supported by research. Instead, teacher evaluation is often done informally, in a covert and haphazard way and primarily relying on subjective judgments of the evaluators (McCleary, 1980).

Given these problems, it is not surprising that the development of effective teacher evaluation practices
has been greatly inhibited. On the other hand, awareness of these stumbling blocks should help researchers develop evaluation procedures and instruments that are sensitive to these problems, and, consequently, are more effective.

School Personnel Support for Improving Teacher Evaluation

In reaction to the ineffectiveness of current evaluation practices, teachers and school administrators have both supported development of procedures and instruments that will improve teacher evaluation. Both teachers and administrators are searching for a better, fairer, more reasonable and accurate way both to evaluate teaching performance and to encourage professional development. Pressures from both administrators and teachers have aimed at replacing vague criteria and unstated assumptions with unambiguous statements both of areas for evaluation and of standards for application (Newton, 1980, p. 45).

The results of a recent survey of high school faculty in a large metropolitan school district also support the need for improvement. Seventy-three percent of the respondents indicated it was important or very important to make staff evaluation procedures more consistent and useful to both teachers and administrators (Crowder, 1983).

School administrators are interested in the improvement of teacher evaluation procedures for somewhat different reasons than are teachers. They recognize that teacher evaluation provides a process by which they can affect the quality of instruction. They also understand that well designed systems of evaluation utilizing clear evaluation
criteria may protect them if their personnel decisions are challenged in court (Newton, 1980).

Professional organizations of both regular and special educators have also voiced support for continued improvement in teacher evaluation. For example, the Council for Exceptional Children professional preparation guidelines express the need for effective evaluation practices, including self evaluation and evaluation by others. Through evaluation, special educators are able to fulfill their professional responsibility to continually improve their teaching skills (Moya, 1980).

Evaluation of Special Educators Versus Regular Educators

Although thousands of books and articles have been written about teacher evaluation in general, only a small fraction of this literature specifically pertains to special education teacher evaluation. This inattention to special educators may be due to the fact that until recently, special education programs were relatively unobtrusive and special education teachers comprised only a small portion of the school work force. However, now that special education is growing in size and becoming more visible, it is more important that appropriate evaluation procedures be developed and implemented among special educators. As pointed out by Kauffman et al. (1973), "The need for a consistent evaluation system that is sensitive to individual differences
is felt most keenly in special education, where exception-
alility of the child is the **raison d'etre** of the teacher and modulation of instruction is essential" (pp. 261-2).

Among researchers who have studied special education teacher evaluation, however, there is lack of agreement regarding the extent to which special educators can be evaluated using the same procedures and instruments that are used with regular educators. Some researchers believe that the competencies needed by special education teachers and regular education teachers are different, while others feel that only the setting is different. Most of the special education teacher evaluation literature seems to support the contention that special educators can be evaluated in the same manner and using the same instruments as regular educators. The results of a survey conducted by Moya (1980) in California support this position. She found that only 14% of the districts responding to her survey had separate guidelines pertaining to the evaluation of special education teachers. The remaining districts treated the special education teachers no differently than the regular education teachers in the evaluation process. In another study, Dorwood (1963) was not able to identify competencies needed by teachers of emotionally disturbed students that were different from those required by regular education teachers. The position that special education teachers require no special consideration is perhaps summed up best
by Howe, who stated: "Special educators would probably be judicious in evaluating teacher effectiveness as regular educators do unless it is demonstrated at some point that different behaviors are important for teachers of handicapped learners" (1981, p. 113).

On the other hand, some authorities believe that teacher evaluation procedures and instruments need to be sensitive to the differences between special educators and regular educators. Winborne (1981) has stated that "The evaluation tool itself is often inadequate for special education teachers. Standard forms and systems of evaluation such as the Flanders Interaction Analysis are not appropriate" (p. 24). In apparent agreement with this position, several researchers have developed specific instruments for evaluating special education teachers. For instance, Brown (1976) has developed a checklist for assessing teachers of learning disabled students; Berdine, Cegelka and Kelly (1977) have implemented a performance rating system for assessing competencies of special education student teachers; McCarthy, Lund and Bos (in press) have developed and used assessment instruments and procedures designed to monitor the quality of instruction provided by teachers serving young handicapped children; and Miller (1980) has introduced an assessment scale for evaluating competencies of special educators.
How to Evaluate Teachers

Just as researchers and educators have failed to reach consensus as to what makes an effective teacher, they also disagree as to how teachers should be evaluated. Most would agree, however, that how teachers are evaluated is largely dependent upon the purpose of evaluation. The procedures and instruments most appropriate for formative evaluation may not be those of choice for summative evaluation.

In deciding how to evaluate teachers, the evaluator also needs to consider the specific context in which instruction occurs. Factors such as student characteristics, curriculum, and the instructional materials available to the teacher, need to be considered (Tikunoff & Ward, 1978). In addition, the evaluator needs to weigh the effects on the teacher's instruction of the goals and objectives of the school, opportunities available for professional development, time available for instruction, class size and the organization of the school (McKenna, 1981). These factors serve to constrict or expand the impact of the teacher's instruction upon students and, consequently, influence the teacher's effectiveness.

Two primary ways of evaluating teacher effectiveness have been discussed in the research literature. The first is to assess the teaching process, i.e., the behavior of the teacher in the classroom that presumably results in
student learning. This is most often evaluated by observing the teacher in the classroom. Evaluation of the teaching act is particularly important when the purpose of evaluation is to help the teacher improve instruction. As pointed out by McIntyre, "If we are interested in helping teachers improve, which is widely accepted as the major purpose of teacher evaluation, then teaching process is the only handle we have as far as the teaching is concerned" (1980, p. 15).

The other major way of evaluating teacher effectiveness is through assessing outcomes. The principal outcome to be assessed is what students have learned (presumably as a result of the teachers' instruction), although the teachers' non-teaching responsibilities might also be evaluated. The accountability movement has led to a greater emphasis being placed upon what teachers should accomplish as a consequence of their instruction. "The focus has shifted from delineating how to perform the task to prescribing what result to achieve" (Newton, 1980, p. 48).

Observation

Moya (1980) has reported that observation, either alone or in combination with other techniques, is the method most often used in the evaluation of special education teachers in California. It is probable that observation is used with equally high frequency elsewhere, and with
good reason. Classroom observation enables the evaluator to assess the climate of the classroom, the teacher's rapport with the students, the interaction between the teacher and the students, and the overall functioning of the class (Evertson & Holley, 1981). Since such data cannot be obtained from any other source, observation is an indispensable technique for assessing the teaching process.

After observing the teacher in the classroom, observers are in a position to give teachers feedback about their performance. This feedback can be valuable in heightening teachers' self awareness regarding their strengths and weaknesses and can provide a basis for allowing the teachers to compare what they thought was occurring in the classroom with what the observers reported seeing. This awareness, in turn, can lead to modification of the teachers' behavior (Feldvebel, 1980; Rayder & Taylor, 1979).

Despite much to recommend observation as a teacher evaluation method, it has several limitations and weaknesses. For one, the obtrusiveness of observers in the classroom may trigger a reactivity effect, whereby the teacher and students behave differently than when observers are not in the classroom (Evertson & Holley, 1981; Hastings, 1973). This reactivity effect can be at least partially mitigated, however, through following a few simple procedures. First, the observer(s) should plan to visit the classroom several times in addition to those times when formal evaluation will occur.
This should help to desensitize the teachers and students to having an unfamiliar person in the classroom. Secondly, observations should occur on more than one occasion. Three to five observations have been suggested as being optimal (Herman, 1973; McDonald, 1980b). Multiple observations not only help the teacher to relax, since all of the evaluative "eggs" are not being "placed in one basket", but also serve to increase the reliability and validity of the observations. Third, the teacher being evaluated should be informed about the purpose and nature of the observation. This should help to allay unfounded fears that the teacher may have about the observation. Finally, the observers should be trained to be as unobtrusive as possible while they are in the classroom (Evertson & Holley, 1981).

Popham (1974) has pointed out another weakness of observation. He has stated that observation only allows the evaluator to assess the process of teaching and not the product or result of the instruction. While this may be true, it is not a weakness, as such, because observation is not intended to be used for product assessment.

A third criticism has been levied by Pottinger (1978), who noted that observers are not able to observe many important factors that are related to successful teaching performance, such as motivation, diagnostic listening abilities, empathy, and flexibility in teaching style.
This criticism appears sound and points out the need for observation to be supplemented by other techniques.

While observation is not a perfect technique, it can be a very valuable one when it is used appropriately. Its usefulness can be enhanced even more when guidelines supported by the professional literature are followed. Research has shown that the most important and perhaps most difficult task of the evaluator is to decide what behaviors should be observed. In general, the evaluator should observe behaviors that are representative of how the teacher usually behaves in the classroom (Berliner, 1976), that are observable and clearly defined, and that have been shown to be important in producing student learning gains (McDonald, 1980b; Peterson & Ward, 1980). Through carefully and consciously selecting behaviors to observe, the evaluator is more likely to obtain data that not only accurately reflect important teaching behavior but are also acceptable to the teacher being observed.

After deciding upon the behaviors to observe, the evaluator needs to determine the observational system and/or instrument to be used. There is general agreement that observational systems/instruments should be as objective as possible so as to minimize biases of the observers. Instruments that are composed of low inference descriptors of behavior tend to be more objective than are instruments requiring greater inference. As with any assessment
instrument, the observation instrument chosen should be characterized by both intrarater and interrater reliability, which can be enhanced by using a large number of indicators of the trait being measured (Evertson & Holley, 1981). In addition to being objective and reliable, the observation system/instrument selected needs to be simple enough so that it is practical to use, yet complex enough to capture the important aspects of the teacher's performance.

A final consideration in optimizing observation pertains to the observer(s). It has been suggested by several researchers (Hall, 1980; Tikunoff & Ward, 1978) that multiple observers should be used in assessing a teacher's performance. Multiple observers can provide more evaluative feedback to the teacher being observed than can a single viewer. They can also validate their findings with one another, increasing the confidence of the observers and the teacher in the data that are gathered. Regardless of whether multiple observers are used, however, it is important that the observer(s) be instructed to follow the specified observation procedure and that they be thoroughly trained in using the observation instrument (Bolton, 1973; Deneen, 1980; Evertson & Holley, 1981).

Assessing Student Outcomes

Proponents of using student outcomes for evaluating teachers believe that measuring student learning is the most valid way of assessing teacher effectiveness. They believe
that students should change through the schooling process and that the students' teacher(s) should be held responsible for the changes that occur (or do not occur). With regard to the education of handicapped students, Kauffman et al. wrote: "Empirical evidence of the exceptional child's achievement in differing educational environments is necessary if special education is to serve children well. Such evaluation entails the construction of methods for assessing teacher effectiveness as a function of objective, quantifiable changes in children's behavior" (1973, p. 262).

Despite the intuitive appeal of using student outcome measures for evaluating teacher effectiveness, there has been considerable opposition within the teaching profession to their use. This negative reaction is not without justification. There are a large number of relatively severe, unresolved problems currently associated with using measures of student learning for evaluation purposes. First, and foremost, there are a large number of variables, such as socio-economic status, ability, home environment, peer group influence and previous instruction, that confound the effect of the teacher's instruction upon what and how much students learn. Of all these variables affecting student learning, the influence of the teacher is not great, being estimated as contributing only 10-25% to pupil performance (McDonald, 1980a; Soar & Soar, 1975). Additionally, even if it were possible to reliably and validly separate out a single
teacher's contribution to student learning, it would not be cost effective to do so.

A second problem in using student outcome measures is the weakness in instruments used to measure what students learn. Currently, good measures do not exist for such outcomes as self-esteem, persistence and discipline. In addition, existing standardized measures of other outcomes, such as achievement, may not reflect what was taught (Berliner, 1976; Rosenshine & McGaw, 1973). Consequently, the students of good teachers might show little gain because of inappropriate measurement instruments rather than because of ineffective teaching. This problem is amplified with respect to special education teachers because many standardized instruments have not been normed on handicapped students and are therefore not appropriate for assessing their achievement.

Soar and Soar (1975) have attacked the use of student outcome measures for evaluating teachers on practical grounds. Because there is so much variation in student achievement among classes taught by a given teacher, it would take data from about 20 classes to make reliable decisions about the teacher, which would be impractical to do. They also contend that if teachers are primarily evaluated on the basis of student learning gains, the teachers may have a tendency to concentrate all of their efforts on those
areas that will be assessed to the detriment of other subject areas that are equally important but are not evaluated.

It would seem that the use of student outcomes for evaluating teachers is an evaluation approach that has potential but has not yet come of age. Perhaps through the development of more practical procedures and appropriate measures, this approach can provide a more viable teacher evaluation option in the future.

Teacher Appraisal Conference

Regardless of the activities selected for inclusion in the teacher evaluation process, most evaluators schedule one or more teacher appraisal conferences or interviews during the course of assessment procedures. Frequently, the evaluator(s) and teacher who is to be evaluated meet prior to the implementation of evaluation procedures. At this pre-evaluation conference, the evaluator(s) may outline evaluation procedures, explain assessment criteria, and, if the teacher is to be observed, specify what will be observed (Gilliland, 1980; Moya, 1980).

Used even more frequently, perhaps, is the conference to convey evaluative information. During this conference, the evaluator(s) and evaluatee can discuss the teacher's strengths and weaknesses and other evaluation findings (Gilliland, 1980; Pine & Boy, 1975). This conference serves a number of important functions, including: warning the teacher that improvement is needed; motivating
the teacher; assessing the degree to which performance standards have been met; recognizing superior performance; and determining what can be done to sustain or improve teaching performance (Haefele, 1981).

Who Should Evaluate

Who should participate in the teacher evaluation process? This is another question that is not easily answered. The best answer to this question is that it depends upon several factors, including the purpose of the evaluation, the context within which the evaluation occurs and the resources available for evaluation. Participants in teacher evaluation may include: the teacher being evaluated; the teacher's supervisor; peers; consumers, including students, parents and other residents; outside specialists; subordinates; or a combination of two or more of the above.

Self Evaluation

Self evaluation or self assessment is the process by which a teacher examines his or her teaching for the purpose of instructional self-improvement. It often involves the use of self ratings, in which the teacher indicates the degree to which he or she exhibits certain competencies or qualities, or self reports, in which the teacher describes his or her teaching strengths and weaknesses in a narrative form. Regardless of the particular form that teacher self
assessment takes, there are multiple advantages to having teachers evaluate their own instruction. First, self evaluation reduces the threat posed by having other persons involved in the evaluation process (Bolton, 1973). It allows the teachers to assess their strengths and weaknesses and then make changes without being judged by external evaluators. Second, self assessment allows the teachers the opportunity to engage in self confrontation, whereby they must honestly examine their teaching. This self examination can often serve as the impetus for the teachers to make changes that will result in improvement.

Perhaps the most important advantage of using self assessment is that it involves the teachers directly in the evaluation process (Carroll, 1981). Through so doing, the teachers are more likely to feel that they are a part of an evaluation team and have some control over the outcome rather than feeling that evaluation is something that is being done to them. This feeling of personal involvement is particularly important when evaluation is for improvement purposes because the interest and motivation to improve must come from within the teachers. It cannot be forced upon the teachers by external evaluators.

Although many educators strongly support the participation of teachers in the evaluation process, teachers seldom do participate. The results of a survey conducted in 1977 by Educational Research Service (1978) showed that
only about one-third of the nation's school districts require teacher self appraisal. In a more recent survey, Moya (1980) reported that only 20% of California's school districts utilized self evaluation by itself or in combination with other data to evaluate special education teachers.

This reluctance to have teachers evaluate themselves may be due in part to research findings that teachers often underrate or overrate themselves compared to others' evaluations of them, and that teacher self assessments have been found to be negligibly related to student achievement gains and to ratings by students (McNeil & Popham, 1973), colleagues and administrators (Carroll, 1981). While at first glance these criticisms appear to undermine the use of teacher self evaluation, they may actually provide support for its use. For instance, one would hope that a teacher's self assessment would not be identical to those of other evaluators, such as supervisors and students, or there would be no need to get the input of more than one evaluator. Additionally, since the use of student achievement measures are fraught with the difficulties outlined earlier, the fact that teacher self assessments are not highly related to these measures may serve to validate rather than invalidate their use. Finally, the findings that teachers tend to underrate or overrate themselves may only point out the need to use teacher self assessment in combination with other evaluation components.
Supervisor Evaluation

Supervisors are involved in teacher evaluation more often than any other personnel. Moya (1980) reported that the building principal was primarily responsible for evaluating special education teachers in approximately 88% of the California school districts surveyed.

While some of the reasons for including the teacher's supervisor in the evaluation process are certainly justified, others appear to be lacking merit. In support of using supervisory evaluation, it cannot be denied that supervisors usually have the responsibility of assuring that teachers carry out their duties satisfactorily and for providing resources to help teachers improve. There is probably also truth in the assertion that supervisors are able to make comparative judgments of teachers' performances because they have regular contact with many different teachers.

Herman (1973) contends that supervisors' training make them best qualified to evaluate teachers. Based upon her research, however, Moya (1980) believes this assertion to be untrue because supervisors are rarely trained in teacher evaluation. She reported that many supervisors complained that they did not feel comfortable in evaluating their subordinates because of insufficient training in evaluation. "Rarely did principals...express feelings of competence in making suggestions or recommendations concerning ways a program or teaching skills could be improved" (p. 105).
Equally invalid is the belief that the supervisors' evaluations of teachers are unbiased. To the contrary, supervisors' judgments may be tainted by political considerations. Supervisors may be overly lenient in evaluating teachers because they are either uncomfortable in communicating negative information or they use leniency as a way of establishing authority (Educational Research Service, 1978).

Another limitation of supervisor evaluation is that supervisors may not have expertise in the content area taught by the teacher being evaluated or have been away from the content area for some time. This criticism appears to be particularly pertinent to special education teachers' supervisors, who often have had little if any teaching experience with handicapped students or in administering special education programs (Winborne, 1981).

Peer Evaluation

There is considerable support for involving peers (fellow teachers) in the evaluation of their colleagues. Peer teachers are able to offer a unique contribution to the evaluation process, particularly when the evaluation is being conducted for instructional improvement. Peers are perceived by the teachers being evaluated as being sympathetic, and, consequently, they are able to help reduce the threat of evaluation. Teachers consider their peers to be the most competent to judge what goes on in
their classrooms because peers have experience in the grade level or subject matter taught by the evaluatees (Educational Research Service, 1978; Herman, 1973; Williams, 1980). Because of their common experiences, peers are also able to assist the teachers being evaluated in discovering solutions to problems and are able to offer specific suggestions for improvement (Meierdiercks, 1981). An additional benefit to using peer evaluation is that research has shown performance assessments provided by peers to be valid and reliable (Cederblom & Lounsbury, 1980).

With such advantages to recommend peer evaluation, it is surprising that only 3% of the nation's school districts used peer evaluation in 1977 (Educational Research Service, 1978). This finding is perhaps less surprising, however, when the cost of peer evaluation is considered. School districts which are already hard pressed financially may be reluctant to pay for the released time required for peer evaluators to be trained and to participate in the evaluation process. Should this explanation be correct, a possible solution could lie in the reallocation of time within the current school day to evaluation. Of course, this would mean that schools would have to be willing to make a greater commitment to teacher evaluation and improvement of instruction than currently exists.

Another reason that peers are not used more in teacher evaluation may be due to the fear that intragroup
conflict could result as a consequence of negative evaluations (Herman, 1973). However, this possibility would be less likely to become a reality if peers were only involved when the purpose of evaluation was to provide feedback for instructional improvement.

An additional limitation of peer evaluation discussed by Herman (1973) is that objectivity may be reduced because the peer is a member of the same employee group as the teacher being evaluated. It is feared that the "good ol' boy" system may prevent peer teachers from being honest in evaluating their colleagues. However, this possible limitation can be countered to some extent by specifying performance criteria by which peers are to assess their fellow teachers and by using peers unknown to the evaluatee.

Consumer Evaluation

There is a growing trend toward using the input of students, parents and other residents of the school district in the evaluation of teachers, perhaps in reaction to the accountability movement (Educational Research Services, 1978; Herman, 1973). Since the primary function of educators is to teach students, it seems only fitting that these student "consumers" should provide feedback to teachers as part of the evaluation process. In addition to students, the input of parents and other members of the community, whose interactions with teachers are also important, would be sought.
Consumer evaluation has not yet become widespread. In 1977, only 2% of the nation's schools solicited input from students in the evaluation process (Educational Research Service, 1978). When assessment of instruction by secondary school students is used, however, it has been shown to offer a number of benefits. First, a large number of assessments of a single teacher can be obtained at little cost. Second, because students have daily contact with several teachers, they have a basis for making comparative judgments (Herman, 1973). Third, student assessments have been shown to be reliable. And fourth, because their perspective is unique, students are able to provide information that cannot be provided by others (Levin, 1979).

On the other hand, the benefits of student evaluation need to be weighed against the limitations. Students may be biased in assessing their teachers by such factors as the teachers' strictness and the amount of work required. Their judgments may also be influenced by their peers (Herman, 1973). Consequently, what may appear to be the opinions of 30 or 40 different students may only be those of a handful.

Evaluation by Others

Professionals from outside the school district are sometimes brought in to assist in teacher evaluations. Because these "external" evaluators, who are often specialists in evaluation, are independent of the local district politics, they are able to provide a fresh perspective and
objectivity to the evaluation process. In Georgia, for example, an external evaluator serves as one member of a three member team that assesses the competencies of teachers seeking certification (Ellett, Capie & Johnson, 1980).

Subordinates, such as teacher aides, are sometimes also asked to provide input to the teacher evaluation process (Bolton, 1980). Subordinates can offer information about a teacher's supervisory skills that cannot be obtained elsewhere, in much the same way that students provide information about an instructor's teaching.

Multiple Evaluators

A great number of school districts are beginning to use more than one evaluator in the teacher evaluation process. They seem to have taken to heart the maxim "two [or more] heads are better than one". Many evaluation authorities agree that teacher evaluation should involve both self evaluation and evaluation by one or more others. Tikunoff and Ward (1978) have noted that it is important to involve both the teacher, as participant observer, and external persons because "...assessment of teaching must build from an inside/outside perspective. This becomes even more critical when working with handicapped students. Neither the teacher nor a single non-participant (outside) observer may have the knowledge and insights to judge a particular teaching event as appropriate" (p. 35).
The use of multiple evaluators would appear to have another benefit as well. The assessment provided by any one evaluator could be used to validate the assessments of other evaluators. In this way, the limitations of using only one type of evaluator (e.g., supervisor, student) can be overcome to a great extent.

A number of school districts throughout the country utilize teams of evaluators to assess teachers. For example, the Atlanta Public schools use a three member team to assist teachers in instructional improvement (Educational Research Service, 1978). The Dallas Independent School District has initiated the Teacher Evaluation Pilot Program, which utilizes a team consisting of the teacher being evaluated, the principal, and either the department chairperson (secondary grades) or the instructional lead teacher (elementary grades) to assess the performance of teachers (Taite, 1980). The Salt Lake City schools use a remediation team consisting of a district learning specialist, building principal, a colleague of the teacher being assessed, and a teacher union representative, to assist teachers who have been identified as failing to progress satisfactorily (McNeil, 1981). And in the teacher evaluation system that is used in Kalamazoo, Michigan, teachers are rated by their principal, students, and peers, as well as rating themselves (Feldvebel, 1980).

Teacher organizations, such as the American Federation of Teachers, also are on record as favoring the use of
multiple evaluators. When the purpose is to help the teacher improve, they support the use of multiple resources, such as peers, supervisors, college personnel, self analysis, and assessments by students, in the evaluation process (McNeil, 1981).

**Evaluation Tools**

Once it is decided who will evaluate the teachers and how they will be evaluated, the evaluators' next task is to select the tools that will be used in conducting the evaluation. The tools most often used are evaluation instruments, such as check lists, frequency counts, narrative summaries, and rating scales. These instruments are often used in conjunction with observation. An additional evaluation tool, which has become more widely used as a result of technological advancements, is the audio/video tape recorder. The versatility of this tool allows it to be used for a variety of purposes in the evaluation process.

**Videotape Recorder**

It has been mentioned that it is important for the teachers being evaluated as well as others to participate in the evaluation process. However, when observation is used as an evaluation procedure, the teachers being observed have a different perspective than do the observers. Because the teachers are unable to see and hear the same things as the other evaluators, their perceptions of what occurred may
be distorted. Videotape recordings (VTRs) help to eliminate this difficulty through allowing the evaluatees to be exposed to the same stimuli as are the other observers.

There are a large number of advantages to using VTRs as tools in evaluating teachers. VTRs are particularly useful in facilitating the teachers' self assessments because they allow the teachers to see themselves as others see them. VTRs can reduce the subjectivity of the teachers' self assessments because they eliminate the necessity for the teachers to rely upon their recollection of what occurred in the classroom. In addition, they provide a way for the evaluatees to verify feedback provided by others (Allender & Yanoff, 1977).

VTRs have been shown to be excellent tools for helping teachers improve their instruction. Through viewing their own teaching, teachers are able to reach new levels of self-awareness and to become more self critical. Fuller and Manning (1973) contend that this self confrontation can be intensified and enhanced through having the teachers who were taped view and discuss the tapes with others. This allows the teachers to identify discrepancies between their view of reality and that of the observers. Other observers can also help the teachers focus on specific teaching behaviors. Fuller and Manning further posit that introspection resulting from self confrontation will subsequently lead to the teachers making needed changes in
their behavior. Research has shown that four to twenty minute videotaped segments are needed to make a difference in the awareness levels of the videotaped teachers (Hosford & Neuenfeldt, 1979; Moritz & Martin-Reynolds, 1980).

VTRs are also able to play an important communication function. They can convey negative information that others do not want to communicate or cannot communicate in a totally acceptable manner to the videotaped teachers (Fuller & Manning, 1973). In this way the evaluatee is able to receive important information without the evaluators having to be the bearers of bad news.

VTRs have other practical advantages as well. They facilitate the observational process for both the observers and the teachers being observed (Bailey, 1979; Sandoval, 1975). While in vivo observation is restricted as to when and where the observation can occur, VTRs allow the observer(s) to view the tapes at a convenient time and location. This is especially important when busy supervisors or peers who teach during the school day are to be part of the evaluation team. Second, while it may not only be inconvenient but disruptive to have several persons observing in a classroom at the same time, VTRs can enable many people to view the evaluatees' instruction without inconvenience. Finally, and perhaps most importantly, VTRs allow observers to view the tapes more than one time. While live observation only allows the observers one chance
to view and assess a specific behavior, VTRs enable observers to view the behavior as many times as desired. This is a particular advantage when there is disagreement among the observers regarding what was seen.

Despite their numerous advantages, VTRs are not without limitations. For one, videotape recording equipment may precipitate a reactivity effect in the classroom just as would in vivo observation (Sandoval, 1975). Fortunately, this reactivity effect can be minimized if the VTR operator will: explain the mechanical operation of the VTR equipment and demonstrate its use to the students who are to be videotaped; allow the students to view themselves on tape; explain why the taping is being done; and use multiple tapings so the teacher will not feel that he or she has to give his or her best performance during a single taping.

VTRs may also need to be restricted from use with those few teachers who could be hurt or damaged by viewing themselves because they are not able to change their behavior (Fuller & Manning, 1973). Logically, the best way to counter this limitation would be for evaluators to be sensitive to the evaluation needs of the teachers being observed, and to modify evaluation techniques accordingly.

Observation Instruments

Observation instruments are probably the most frequently used tools in teacher evaluation, and for good reason. "Evidence indicates that observation instruments
can be very useful for giving teachers feedback on aspects of their teaching because they are reliable and focus on discrete aspects of the teaching process" (Levin, 1979, p. 242). As with any evaluation tool, however, some observation instruments are better than others. The best instruments are those that are reliable, valid and utilize objective statements of behavior (Feldvebel, 1980).

**Rating Scales.** Rating scales usually require the observer to specify the degree to which the teacher being observed demonstrates each of a group of specified behaviors. They are used by observers to assess teacher competence more often than any other observation instrument, probably because of the many advantages that they offer (Shearron, 1978). The greatest advantage of rating scales is that they are simple to understand and convenient to use (Herman, 1973). Because teacher evaluation activities are only a small part of the responsibilities that those who evaluate teachers must perform, they naturally choose evaluation instruments that require a minimum of time to use. Rating scales meet this need because their use requires little training and they can be completed quickly.

Of course, convenience and simplicity would be of little importance if rating scales did not also have utility. On this count, rating scales have been shown to provide useful information about instruction (Rosenshine,
1970). They also provide a basis for comparison among several raters.

Another advantage of rating scales is that they offer versatility. Since they can be weighted different ways for different purposes, they can be used for both formative and summative evaluation (Shearron, 1978).

Despite their popularity, rating scales have a number of drawbacks and limitations, most of which pertain to the raters and not the rating instruments. Because evaluators are often lenient, ratings are plagued by a ceiling effect: most teachers receive the top or next to the top ratings (McIntyre, 1980; Rosenshine, 1970). Another related problem is the halo effect, wherein each item reflects the rater's overall assessment of the teacher being observed rather than the rater's assessment of the specific behavior being measured by the item. Both of these problems can be attenuated to a great extent through adequate training of the raters in the use of the rating instrument. Raters should be trained as to what behaviors are to be observed and how a teacher's behavior should be rated.

Another problem that affects many rating scales is that they lack operational definitions of the behaviors being assessed. The obvious solution to this problem is to use rating instruments that have clearly defined assessment dimensions and that provide statements clearly describing the observable behaviors that are to be assessed. In
addition, it has been suggested that evaluators use rating instruments that minimize what the observer must do during the observation; use items requiring little judgment and inference (McNeil & Popham, 1973); use criteria that are important to the learning environment of any classroom; and weight the assessment criteria (Deneen, 1980).

As mentioned earlier, it is also suggested that multiple raters be used. "The general rule is that the number of judges should be no less than three since that number gives a minimal notion of whether consensus exists in a specific evaluation" (French-Lazovik, 1981, p. 83).

**Other Observation Instruments.** When using frequency or count systems, the observer records the number, frequency or presence of particular events or behaviors. These systems have limited utility in the observation of teaching, however, because they do not allow the observer to assess the qualitative dimension of teaching and the appropriateness of the teaching skills that are demonstrated (Berliner, 1976).

The check list, another observation instrument enjoying limited use, is composed of a list of characteristics or behaviors that are considered essential for competent teachers to demonstrate. It requires that the observer indicate whether the teacher being observed demonstrates each listed behavior.
Narrative reporting requires that the observer merely record what occurred in the classroom using non-technical language. Given proper training, observers have been shown to provide objective, useful information while using this tool. Williams (1980) believes that the narrative report is the best observation instrument because "...it gives a more complete view of the observation and can provide the reflective statements required for commendations and recommendations for improvement" (p. 11).

Promising Approaches to Teacher Evaluation for Improvement

As previously noted, many of the teacher evaluation procedures and instruments used in the past, as well as those currently being used, are useless at best. Fortunately, there are a number of developing evaluation systems and accompanying instruments that show promise in helping teachers improve their instruction.

The majority of these promising approaches to formative evaluation may be classified as competency-based, which means that the evaluator assesses the degree to which the teacher being evaluated demonstrates pre-specified competencies judged to be important for teaching. In most cases these competencies are described in behavioral terms and are observable. Often, too, minimal competency levels are specified. Many of these competency-based assessment
systems have developed out of competency-based teacher training programs, particularly in special education.

The first step in designing competency-based assessment procedures and instruments is to "...define essential teacher competencies using behavioral performance criteria that can be readily observed and measured in the classroom" (Miller, 1980, p. 5). Several states, including California, Colorado, Georgia, and Ohio, have undertaken studies to develop and validate competencies needed by special educators. In California, for example, generic and specific competencies were developed for special educators in the following areas:

1. acquaintance with factual material about exceptional children;
2. counseling students;
3. managing the classroom and program administration;
4. communicating with parents;
5. communicating with other professionals;
6. developing and planning instructional programs;
7. evaluating the instructional process;
8. describing and assessing student behavior;
9. implementing instructional programs; and
10. defining instructional goals and objectives (Black, 1973).

After important competencies are identified, the next step is to incorporate these competencies into an evaluation system. Tawney (1980) developed such a system,
which examined teacher-student interaction, student on task behavior, and changes in student academic performance. Following assessment, teachers are offered remedial assistance in those areas in which criterion standards are not met.

At the heart of competency-based evaluation systems are the assessment instruments, such as rating scales, that are used by the evaluators. Because of their importance in the evaluation process, care in selection of appropriate instruments is essential. Although many rating scales exhibit the weaknesses described previously, there are a growing number of competency-based rating scales or behaviorally anchored rating scales (BARS) being developed that provide more reliable and valid assessment of teaching behaviors. Each item on these scales depicts a specific teaching competency and each point on the rating scale is tied to a specific, observable behavior or performance value. These instruments generally demonstrate greater interrater reliability because raters interpret the evaluation criteria in a more uniform way (Jacobs, Kadry & Zedeck, 1980).

Although competency-based assessment has much to recommend it for use in special education teacher evaluation, it is not without its critics. Pottinger (1978) and Tikunoff and Ward (1978) pointed out that proponents assume that the whole (general teaching competency) is equal to the sum of
its parts (specific competencies). They argued that this is a faulty assumption because a teacher who demonstrates particular microskills is not necessarily competent overall. Another criticism of competency-based assessment has been that statements of teacher competencies are not empirically validated through demonstrating their effect upon student performance. Instead, they tend to be grounded on expert opinion (Berliner, 1976; Shores, Cegelka & Nelson, 1973; Travers, 1981). These criticisms generally appear well grounded and point out the need for additional research on competency-based assessment.

An example of one of the better competency-based instruments is the Teaching Behavior Inventory (TBI), which consists of items depicting specific, observable behaviors that are rated on a five point scale, of which points one, three and five are anchored to behavioral descriptors (McCarthy, Lund & Bos, in press).

Another exemplary instrument is the Teacher Performance Assessment Instruments (TPAI), which consists of three separate parts that jointly assess 14 important teaching competencies. Items on each part are assessed with five point scales, and each scale point is tied to a behavioral descriptor. The TPAI is used widely throughout Georgia (Ellett, Capie & Johnson, 1980).

A third instrument appearing to have some merit is the Competency Development Scale for Teachers of Exceptional
Children (CDS). The CDS was designed to assess competencies of special educators for the primary purpose of providing data that could be used in planning appropriate inservice programs. It uses a behaviorally anchored five point rating scale to assess competencies of special educators in the following five major skill areas: general competence; assessment, objectives and curriculum planning; the teacher/learning situation; evaluation and records; and parent involvement (Miller, 1980).

The Teacher Intern Performance Rating System (TIPRS), which is used to assess the competencies of student teachers enrolled in the University of Kentucky's competency-based special education training program, also shows promise. The student teacher's university supervisor and supervising teacher each assess the student nine times using the Daily Rating Form, which is the main component of the TIPRS. This instrument allows the evaluators to examine the trainees' interactions with students and to assess lesson plans and pupil work using a five point rating scale (Berdine, Cegelka & Kelly, 1977).

Despite the promise that the previously described competency-based assessment instruments show, it is difficult to evaluate their psychometric adequacy and their effectiveness in measuring teacher behavior due to the unavailability of sufficient data. With the exception of
the Teacher Performance Assessment Instruments, reliability and validity data are not readily available for these instruments.

**Summary of the Research on Teacher Evaluation**

Teacher evaluation research shows that while teachers, evaluators and the public are placing greater emphasis upon reliable, valid and efficient evaluation of teachers, the development of evaluation procedures and instruments to meet this need has not kept pace with the demand. This dearth of procedures and instruments is even greater in the more specific category of special education teacher evaluation.

Reactions to current teacher evaluation practices have generally been negative. Teachers who are evaluated feel uncomfortable at best, and hostile at worst. Supervisors are also not happy with present evaluation procedures, primarily due to their lack of training in teacher evaluation. In addition, evaluation has not generally been found to be effective for helping teachers improve instruction or in supporting administrative personnel decisions. The consensus among evaluators and evaluatees, alike, is that improvement is needed in teacher evaluation procedures and instruments if they are to serve a useful purpose.

However, despite agreement that teacher evaluation practices should be improved, researchers' efforts to make
improvements have been hampered by a number of problems, including: lack of consensus as to what constitutes a good teacher; lack of control over situational and contextual variables that may affect teaching; the conflict between the summative and formative roles of the evaluator; and practical considerations, such as teacher union contracts.

Teachers are generally evaluated for two reasons: 1) to assist them in improving their instruction; and 2) to assist their supervisors in making administrative decisions. Although researchers agree that instructional improvement is the most important of the two reasons for evaluating teachers, there is disagreement regarding the extent to which it is emphasized in practice.

There is lack of agreement among practitioners and researchers regarding the extent to which special education teachers need to be evaluated with procedures and instruments that differ from those used to evaluate regular education teachers. Some feel that teaching handicapped students requires somewhat unique competencies, which would require the development of special evaluation instruments. Others believe that all teachers can be assessed with a common set of evaluation procedures and instruments. In practice, at least, the latter position appears to have the most support.

Although decisions concerning how teachers are evaluated and who participates in evaluation are largely dependent upon the purpose(s) of the evaluation, as well as
other factors, there is growing support for the position that multiple evaluators, including the teacher being evaluated, the evaluatee's supervisor, peers, and students, should participate, and that measures of multiple variables from multiple sources should be used.

Teacher evaluation most often involves assessing the teaching process (classroom instruction) or the product of instruction (student learning). Although it is certainly logical that teachers should be held accountable for what their students learn (or do not learn), there is a multitude of problems associated with measuring student outcomes for the purpose of evaluating teachers, including the following: other variables, such as home environment and previous instruction, may confound the effect of a teacher's instruction on student learning; adequate measures of such outcomes as self esteem and persistence do not exist; and teachers may under-emphasize those instructional areas that are not assessed. Evaluation of the teaching process is likewise not without its drawbacks. However, process evaluation is most often recommended, particularly when the purpose of evaluation is instructional improvement.

Teachers are most often evaluated through observation of their instruction, either alone or in combination with other techniques, such as a teacher appraisal conference. Observation instruments that are most frequently used include rating scales, frequency counts, check lists, and
narrative reports. The best of these instruments are those that are research based, utilize objective statements of behavior, and have been demonstrated to be reliable and valid. Competency-based, behaviorally anchored rating scales, in particular, show considerable promise, especially when used in conjunction with a competency-based evaluation system.

In addition to good evaluation instruments, using videotape recordings (VTRs) in the teacher evaluation process has proven to be valuable, particularly when the purpose of evaluation is to help the teacher improve instruction. VTRs can serve as a foundation for more reliable and valid evaluation, and can facilitate instructional improvement.

Research has shown that some conditions are more likely to facilitate improvement in teaching than others. Teachers are most likely to demonstrate improvement when the sample of their teaching performance that is evaluated is typical rather than atypical; they judge evaluative feedback to be trustworthy, informative, unambiguous and reflecting their "true" performance; their self assessments are more positive than the assessment by others; and they have favorable attitudes toward evaluation. Additionally, improvement is facilitated when models of alternative behavior are made available; provisions are made for teachers to establish and practice new behaviors; teachers are
actively involved in the improvement process; and there is a support system available to help the teachers improve.

In addition to identifying conditions that will facilitate instructional improvement, researchers have also made suggestions for optimizing overall evaluation effectiveness. It has been recommended that evaluators be thoroughly trained; that evaluators not rely on a single evaluation technique or kind of data, but use several approaches and types of data; that evaluation criteria be flexible enough to encompass different teaching styles; and that a common understanding exist between the evaluator and evaluatee regarding the goals of the evaluation, what data will be collected and how data will be collected.

**Need for Additional Research**

**On Teacher Evaluation**

It has been pointed out that teacher evaluation, particularly in regard to instructional improvement, is an important area of concern both inside and outside of the educational community. As a result, it has been the subject of a large number of studies in the professional literature. Although some of these studies have been based upon empirical research, the majority are not data based (Levin, 1979). This is particularly true of studies pertaining to special education teacher evaluation. Within the teacher evaluation literature, a multitude of procedures, techniques and instruments for assisting teachers
in improving their instruction have been described. Despite the fact that many of these procedures and instruments are intuitively appealing and have good face validity, few of them are supported by adequate empirical research. Researchers have not demonstrated that these evaluation procedures and instruments are effective in helping teachers improve their instruction. In many cases, reliability and validity data are not even reported. Considering the importance of teacher evaluation in terms of its direct impact upon multitudes of teachers and administrators and its indirect consequences for students, it is imperative that teacher evaluation procedures and instruments not only be developed but that their effectiveness be demonstrated through empirical research.

**Purpose of the Current Study**

The scarcity of empirical research demonstrating the effectiveness of teacher evaluation, in general, and special education teacher evaluation, in particular, served as a catalyst for the researcher in designing the current study. Although several promising approaches for evaluating special education teachers have been developed in recent years, many of these procedures are not used for instructional improvement purposes and none of the approaches have empirically demonstrated their effectiveness in helping teachers improve. Consequently, it became apparent that there was need to not only develop a procedure that would
help special education teachers improve their instruction but to empirically determine the adequacy of the procedure in effecting positive change. With this in mind, an experimental study was conducted to field test an evaluation procedure designed to meet this need.

A formative evaluation procedure, called the Special Educator Improvement Procedure (SEIP), was designed to utilize those techniques and methods supported by research and to be practical enough for implementation in the schools. The SEIP was also designed to be generic enough for use with all special education teachers, regardless of the handicapping condition, severity level or age of the handicapped students being taught.

The SEIP assists special education teachers in improving their teaching through providing them with feedback from multiple assessments of their instruction. It integrates several evaluation components, each of which has been supported by the previously cited research studies for use in formative teacher evaluation. The components of the SEIP include: presenting multiple videotaped samples of the evaluatees' teaching performance to a team of evaluators, including the special educator being evaluated, the evaluatee's supervisor, a peer, and another educator chosen by the evaluatee; having the evaluation team assess the teacher's instruction using a reliable and valid competency-based assessment instrument; allowing the evaluatee to
compare his or her self assessment with the assessments of other evaluation team members; and having team members make specific suggestions as to how the evaluatee might improve his or her instruction. Although each of these components has been employed independently with some success, no one has previously combined them into a single comprehensive package. The strength of the SEIP lies in the fact that it capitalizes upon each technique's individual contribution to instructional improvement, as well as those resulting from the interaction of the components. Thus, the composite effect of the SEIP should be greater than the sum of each component's independent effect.

**Hypotheses**

It was hypothesized that special educators participating in the Special Educator Improvement Procedure would improve their instruction. Specifically, it was predicted that:

1. Subjects who participated in the Special Educator Improvement Procedure (SEIP) would be ranked higher on the revised Teacher Performance Assessment Instruments (TPAI-R) than would subjects who had not yet participated in the SEIP for each week of the study.

2. Assuming that Subject One participated in the SEIP following the first week, that Subject
Two participated in the SEIP following the second week, and so forth:

a. For the second week, the transformed TPAI-R mean of Subject One would be greater than the transformed TPAI-R means of Subjects Two through Six, and the transformed TPAI-R means of Subjects Two through Six would not differ.

b. For the third week, the transformed TPAI-R means of Subjects One and Two would not differ but would be greater than the transformed TPAI-R means of Subjects Three through Six, and the transformed TPAI-R means of Subjects Three through Six would not differ.

c. For the fourth week, the transformed TPAI-R means of Subjects One through Three would not differ but would be greater than the transformed TPAI-R means of Subjects Four through Six, and the transformed TPAI-R means of Subjects Four through Six would not differ.

d. For the fifth week, the transformed TPAI-R means of Subjects One through Four would not differ but would be greater than the transformed TPAI-R means of Subjects Five
and Six, and the transformed TPAI-R means of Subjects Five and Six would not differ.

e. For the sixth week, the transformed TPAI-R means of Subjects One through Five would not differ but would be greater than the transformed TPAI-R mean of Subject Six.

f. For the seventh week, the transformed TPAI-R means of Subjects One through Six would not differ.

It was also hypothesized that the participants in the SEIP would perceive the SEIP as being beneficial. Specifically, it was predicted that the SEIP participants, including the experimental subjects and the other members of the feedback teams, would rate:

1. the SEIP feedback sessions as being very beneficial to extremely beneficial for helping special educators improve their teaching;

2. their participation in the SEIP feedback session as being very beneficial to extremely beneficial; and

3. the TPAI-R as being good to excellent for assessing special educators' teaching.
CHAPTER 2

METHOD

Design

A single subject, multiple baseline across subjects design (Hersen & Barlow, 1976) was used in this study. This design allows for the detection of changes in an individual's behavior over time. In addition, because subjects who have not yet received treatment serve as controls for the subject(s) receiving treatment, the multiple baseline design controls possible historical and maturation threats to internal validity. Refer to Figure 1 for a graphic representation of the design.

Subjects

Six special education teachers\(^1\) from six schools in Tucson Unified School District, a large metropolitan school district with approximately 54,000 students, served as subjects for this study. Potential subjects were recruited from among teachers who teach handicapped students in self-contained classes.\(^2\) The purpose and requirements of the

\(^1\)Refer to Table 1 for specific information about the subjects and other members of the SEIP feedback team.

\(^2\)A self-contained class is defined as a class composed entirely of handicapped students who spend more than 50% of each school day with the same teacher.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Week</th>
<th>Week</th>
<th>Week</th>
<th>Week</th>
<th>Week</th>
<th>Week</th>
<th>Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>00I</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>B</td>
<td>00</td>
<td>00I</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>C</td>
<td>00</td>
<td>00</td>
<td>00I</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>D</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00I</td>
<td>00</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>E</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00I</td>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>F</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00</td>
<td>00I</td>
<td>00</td>
</tr>
</tbody>
</table>

0 = repeated measure
I = introduction of treatment

Figure 1. Multiple Baseline Design for Implementing SEIP
<table>
<thead>
<tr>
<th>Team</th>
<th>Subject</th>
<th>Position on Feedback Team</th>
<th>Other Educator</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Teacher of learning disabled</td>
<td>----(^a)</td>
<td>Teacher of emotionally mildly handicapped</td>
</tr>
<tr>
<td>B</td>
<td>Teacher of educable mentally handicapped</td>
<td>Building principal</td>
<td>Teacher of trainable mentally handicapped</td>
</tr>
<tr>
<td>C</td>
<td>Teacher of primary grade mildly handicapped</td>
<td>Building Principal</td>
<td>School Counselor</td>
</tr>
<tr>
<td>D</td>
<td>Teacher of emotionally handicapped</td>
<td>Building principal</td>
<td>Teacher of emotionally handicapped</td>
</tr>
<tr>
<td>E</td>
<td>Teacher of multiply handicapped</td>
<td>Building principal</td>
<td>Psychologist</td>
</tr>
<tr>
<td>F</td>
<td>Teacher of severe language impaired</td>
<td>Building principal</td>
<td>Speech pathologist</td>
</tr>
</tbody>
</table>

\(^a\) This team consisted of three members because the principal was unable to participate due to illness.

\(^b\) This teacher had formerly taught learning disabled students.

\(^c\) This teacher had formerly taught educable mentally handicapped students.

\(^d\) This teacher had formerly taught multiply handicapped students.
study were described to the faculty in general terms by the principal at each of the six participating schools. Those teachers expressing an interest in participating in the study were asked to attend another meeting where details about the study were explained and where they were given a brief written description of the study (Appendix A). Teachers attending this meeting were told that individuals selected to participate in the study would be videotaped in their classrooms two times each week for seven weeks. At one point during the seven weeks, each participant would be asked to observe several of the videotape recordings (VTRs) with three other professionals and to rate the quality of his or her teaching on these tapes. They were told that the purpose of the study was to help them evaluate and improve their performance in the classroom. The teachers were also told that participants would benefit from the study through receiving feedback that could be useful for instructional improvement.

From among those teachers who volunteered, one to four participants were selected from each school to conditionally participate in the study. The number of potential participants varied from school to school as a result of differences in the rate of voluntarism and the experimenter's desire to select teachers of students having different handicaps. The volunteers' participation as subjects was contingent upon receiving an initial, pretreatment overall
rating above the minimum competency level established by the Teacher Assessment Project at the University of Georgia on fewer than 75% of the items on the revised Teacher Performance Assessment Instruments (TPAI-R). However, due to conditions that will be explained subsequently, two of the subjects did not meet this criterion. Only one subject was ultimately chosen to participate from each school to reduce the possibility of contamination of the treatment effect resulting from communication among participants who had received the treatment with those who had not yet received treatment. Only teachers of self contained special education classes were recruited for the study to eliminate class type (i.e., self contained, resource, etc.) as a possible confounding variable.

Because of practical and design considerations, neither the school district nor individual subjects were randomly selected to participate in the study. Thus, the results could not be generalized beyond those teachers who volunteered to participate. It could not be assumed that the teachers who volunteered are similar to the teachers who did not volunteer. In addition, it could not be assumed that special education resource teachers or teachers of nonhandicapped students would respond to the SEIP similarly to the experimental subjects.
Dependent Variables

The primary dependent variable for this study was classroom teaching performance as measured by a modified version of the Classroom Procedures and Interpersonal Skills portions of the Teacher Performance Assessment Instruments (TPAI). Classroom teaching performance was defined as those aspects of a teacher's observable classroom behavior that impact upon the students' learning and development, and that include:

1. using instructional techniques, methods, and media related to communicating with students;
2. demonstrating a repertoire of teaching methods;
3. encouraging student involvement in instruction;
4. demonstrating an understanding of the subject being taught;
5. organizing time, space, materials, and equipment for instruction;
6. demonstrating enthusiasm for the subject being taught; and
7. managing classroom interaction.

It was assumed that these behaviors adequately represent the construct of teacher performance.

The TPAI is a competency based assessment instrument that assesses 14 teaching competencies. Each of the competencies is rated on a five point scale with two to
five performance indicators. Each of the scale points is defined by a one sentence behavioral descriptor, which serves to increase interrater agreement through the reduction of ambiguity.

Capie, Ellett and Johnson (1979) and Ellett (1980) have reported the results of a number of reliability studies conducted with the TPAI. Interrater reliability was investigated for both in vivo and videotaped instruction. When three member observation teams rated classroom teachers after observing their live teaching performance, interrater agreement coefficients ranged from .81 to .98 with a mean of .89. However, when the observation teams rated videotaped samples of teachers' instruction, coefficients of interrater agreement ranged from .43 to .91 with a mean of .61. In studies of internal consistency, the authors have reported Cronbach's alpha coefficients ranging from .75 to .95.

Capie (1980) demonstrated the predictive validity of the TPAI by correlating teachers' competency ratings on the TPAI with an effectiveness index, which was based upon pre-post achievement test gains. They obtained validity coefficients ranging from .50 to .72. The authors also correlated the TPAI with the Purdue Observational Rating Scales to demonstrate concurrent validity and obtained significant correlations ranging from .32 to .75.
For the current study, the TPAI was slightly modified by the experimenter so that it would be more appropriate for use with special education teachers. This revised version (TPAI-R) consists of 20 items from the Classroom Procedures and Interpersonal Skills portions of the TPAI that were applicable to special educators. A "Not Applicable" response category was added to each item to provide a response option for those situations when a specific item was inappropriate. The TPAI-R may be found in Appendix B.

During each of the seven weeks of the study, each of the six subjects was videotaped on two randomly selected occasions while engaged in normal teaching activities in the classroom for approximately one-half hour on each occasion. Unusual and non-teaching activities, such as recess and parties, were not sampled. An additional two sessions were videotaped at the outset of the study for each subject, but were not included in the analysis; these sessions were used to determine participant eligibility and also allowed the subject and his or her students to become accustomed to having the video equipment in the classroom, thus reducing possible reactivity effects. A 10 minute segment from each taped session was randomly selected for analysis.

A secondary dependent variable for this study was the perceived benefit of the Special Educator Improvement
Procedure (SEIP). This was measured by means of an experimenter developed questionnaire (see Appendix C), which was administered to the subjects and other SEIP team members at the close of the study. Participants were asked to rate the benefit of their participation in the SEIP on a five point scale ranging from not beneficial to extremely beneficial. In addition, they were asked to assess how beneficial the SEIP feedback session would be for helping special educators improve their teaching and how adequate the TPAI-R was for assessing special educators' teaching using five point scales ranging from not beneficial to extremely beneficial and poor to excellent, respectively.

In addition to rating the SEIP, participants were asked to assess the specific strengths of the SEIP and the TPAI-R and to suggest how this procedure and instrument might be improved. These data were gathered in order that future modifications could be made to the SEIP that would make it more effective in helping special educators improve their instruction.

For the study, it was assumed that improvement in teaching performance is desirable; that special educators are able to demonstrate improvement; and that special education teachers must want to improve for change to occur.

**Independent Variable**

The Special Educator Improvement Procedure (SEIP), which will be described subsequently, served as the
independent variable for this study. During each week of the study, a subject was randomly chosen (without replacement) to participate in the SEIP, which helped to control for a possible selection threat to internal validity. Subjects were not told which videotaped sessions would be used for the SEIP so as to reduce the possibility that this knowledge would alter their classroom behavior.

Before data were collected, a feedback team was established for each subject. Each team was composed of the subject, the subject's immediate supervisor, a peer, and another educator, who was selected by the subject. Each of the feedback team peer members was a teacher who currently taught or had previously taught a class that was similar to that taught by the subject but was located at a different school. This precaution prevented the possibility that the peers might have been biased (either positively or negatively) as a result of previous interactions with the subject. Peers were selected from those teachers who volunteered for the study but who were not selected as subjects. Four of the six peers scored above the minimum competency level on at least 75% of the TPAI-R items. The other two peer feedback team members did not meet this criterion due to the limited number of volunteers and the need to match subjects with peers who taught similar students.
The SEIP was developed by the experimenter as a procedure which would integrate a number of evaluation techniques that had shown promise when used independently. It is composed of the following components:

1. Feedback team members jointly view two videotaped samples of the subject's classroom performance that are gathered within a week prior to assessment. By using VTRs, subjects are able to view their own behavior, and feedback team members can view these behavior samples at a time that is convenient for them. Two samples are used to better represent the teacher's performance. The VTRs are shown more than once when requested by feedback team members.

2. The subject is encouraged to answer questions and clarify issues raised by the other feedback team members about the VTRs.

3. The feedback team members rate the subject's performance on the VTRs using the revised Teacher Performance Assessment Instruments (TPAI-R).

4. The subject is asked to leave the room. During this time, the experimenter assists the other feedback team members in reaching consensus regarding the rating for each item. For a given item, agreement by two of the three feedback team members constitutes consensus.

5. The subject is invited back into the room and the agreed upon set of ratings is presented to the subject for comparison with his or her self ratings.
6. Based upon the strengths and weaknesses identified on the TPAI-R, the feedback team discusses how the subject might improve his or her instruction.

Feedback team members were trained to use the TPAI-R by rating a specially prepared training VTR. Interrater agreement was assessed through group consensus, which was defined as agreement in ratings among three of the four feedback team members with the rating of the rater not in agreement being no more than two points apart from the other raters. Consensus on 70% of the items (14 of 20 items) was used as a criterion for the successful completion of the training.

A standardized script for implementing the SEIP was used for each subject's feedback session (see Appendix D). This was done to insure uniformity across the sessions for all six subjects.

Independent Observers

To ascertain changes in the subjects' teaching over the course of the study, two independent observers were employed to rate all of the subjects' videotaped performances using the TPAI-R at the conclusion of the seven week data gathering period. The observers rated a total of 16 videotape recordings (VTRs) for each subject (two VTRs for each of seven weeks plus two VTRs for determining eligibility for the study). The VTRs were presented to the observers in random order so as to have reduced a possible
instrumentation threat to internal validity. Graduate students in education who were not aware of the purpose of the study served as the observers. They were thoroughly trained in using the TPAI-R with videotapes prepared for this purpose. Interrater agreement was assessed using intraclass correlation (Blalock, 1979). Interrater reliability coefficients improved from .78 to .92 during training. The observers did not begin rating the subjects' videotapes until a correlation of .90 had been obtained with the training tapes. The two observers' ratings were later combined into a single mean for each week's two segments for analysis purposes. This was done in order that the index used in the analysis would be more representative of the subject's performance.

Pretreatment Screening

Prior to the first week of the study, the independent observers rated two videotaped samples of the potential participant's teaching performances. The percentage of items on the revised Teacher Performance Assessment Instruments (TPAI-R) on which these volunteers were rated as being competent (performing above the minimum competency level established by the Teacher Assessment Project at the University of Georgia) are shown in Table 2. Originally, subjects were to be selected from among those volunteers who were rated above the minimum competency level on fewer than 75% of the TPAI-R items, and peers were to be selected
Table 2. Percentage of SEIP Participants' TPAI-R Ratings above the Minimum Competency Level

<table>
<thead>
<tr>
<th>School</th>
<th>Participant Position</th>
<th>SEIP Role</th>
<th>Percentage of TPAI-R Ratings above Minimum Competency Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Teacher of learning disabled</td>
<td>Subject</td>
<td>72</td>
</tr>
<tr>
<td>A</td>
<td>Teacher of emotionally handicapped</td>
<td>Peer</td>
<td>72</td>
</tr>
<tr>
<td>B</td>
<td>Teacher of educable mentally handicapped</td>
<td>Subject</td>
<td>87</td>
</tr>
<tr>
<td>B</td>
<td>Teacher of primary grade mildly handicapped</td>
<td>Peer</td>
<td>93</td>
</tr>
<tr>
<td>C</td>
<td>Teacher of primary grade mildly handicapped</td>
<td>Subject</td>
<td>89</td>
</tr>
<tr>
<td>D</td>
<td>Teacher of emotionally handicapped</td>
<td>Subject</td>
<td>51</td>
</tr>
<tr>
<td>D</td>
<td>Teacher of emotionally handicapped</td>
<td>Peer</td>
<td>90</td>
</tr>
<tr>
<td>D</td>
<td>Teacher of emotionally handicapped</td>
<td>Peer</td>
<td>67</td>
</tr>
<tr>
<td>E</td>
<td>Teacher of multiply handicapped</td>
<td>Subject</td>
<td>62</td>
</tr>
<tr>
<td>F</td>
<td>Teacher of severe language impaired</td>
<td>Subject</td>
<td>66</td>
</tr>
<tr>
<td>F</td>
<td>Teacher of emotionally handicapped</td>
<td>Peer</td>
<td>75</td>
</tr>
<tr>
<td>G</td>
<td>Speech pathologist</td>
<td>Peer</td>
<td>87</td>
</tr>
</tbody>
</table>

from among those volunteers who scored above the minimum competency level on 75% or more of the TPAI-R items.

However, to meet the requirement that for each Special Educator Improvement Procedure (SEIP) feedback team, the subject and peer teach at different schools and that the
peer teach or have previously taught students having the same handicap as the students taught by the subject, the 75% criterion could not be used in every case. As can be seen in Table 2, two of the subjects were rated above the minimum competency level on more than 75% of the TPAI-R items and two of the peers were rated above the minimum competency level on fewer than 75% of the TPAI-R items.

Pilot Study

Prior to the implementation of this study, the Special Educator Improvement Procedure (SEIP) was piloted with a teacher of a class of self contained mentally handicapped students in a small Tucson school district. Following videotaping of the teacher in her classroom on two different days, the SEIP was implemented with her feedback team, which consisted of herself, her immediate supervisor (building principal), the Director of Special Education, and a teacher of the mentally handicapped from another school. After the feedback session was completed, the experimenter solicited comments from the feedback team concerning the strengths and weaknesses of the SEIP. Based upon this feedback, the SEIP was modified slightly. This modified version was subsequently used in conducting the larger study.
Analysis

Revusky's $R_n$ statistic (Hersen & Barlow, 1976) was used to test the subhypothesis that for each week of the study, subjects who participated in the SEIP would be ranked higher on the TPAI-R than would subjects who had not yet participated in the SEIP. $R_n$ compared the ranking of the subjects' mean ratings each week with the probability of obtaining those ranks by chance. Should the treatment be effective, the subject who received the treatment in a given week would be ranked above those subjects who had not yet received the treatment. The sum of the ranks across all weeks constituted the statistic $R_n$. Statistical significance was determined using a .05 alpha level. $R_n$ was chosen because it would enable the researcher to statistically demonstrate the effectiveness of the SEIP for each of the subjects individually, as well as the overall effectiveness of the treatment.

For analysis purposes, the data were transformed to eliminate the effects of the differences in absolute means among subjects. That is, the transformation adjusted for the fact that a more competent teacher would be more highly rated on a given week than a less competent teacher. The following formula was used for this purpose:

$$\bar{X}_T = \frac{B_i - \bar{A}_i}{\bar{A}_i}$$

where $\bar{X}_T =$ transformed mean
\[ B_i = \text{performance level for subject } i \]
\[ \text{when the experimental intervention} \]
\[ \text{was introduced} \]
\[ \bar{A}_i = \text{mean performance across all baseline} \]
\[ \text{days for subject } i \]

The a priori contrasts specified in the second subhypothesis (i.e., for each week of the study, the transformed TPAI-R mean of the subject who had participated in the SEIP during the previous week would be significantly greater than the mean rating of the subjects who had not yet participated in the SEIP, but would not differ from the mean ratings of those subjects who had already participated in the SEIP), was tested using Tukey's HSD procedure (Winer, 1971), which allows one to specify the probability of making a Type I error for all pairwise comparisons. Statistical significance was determined using a .05 experimentwise alpha level.

The hypothesis concerning the participants' assessments of the SEIP and TPAI-R was tested by examining the median ratings of the three quantitative questions on the SEIP Participant Questionnaire. Based upon the most negative response being assigned a rating of one and the most positive response being assigned a rating of five, median ratings greater than 3.5 were considered to support the prediction that the SEIP participants would assess the
feedback sessions as being beneficial and the TPAI-R as being a good instrument for evaluating special educators.
CHAPTER 3

RESULTS

Revusky's $R_n$ Statistic

The results of the analysis of the subjects' rankings on the TPAI-R using Revusky's $R_n$ statistic were not statistically significant. This means that for any given week of the study, the transformed mean rating of the subject who had just received feedback through the SEIP was not consistently greater than the ratings of the subjects who had not yet received SEIP feedback. This finding failed to support the subhypothesis that subjects who participated in the SEIP would be ranked higher on the TPAI-R than subjects who had not yet participated in the SEIP.

Table 3 shows the subjects' mean TPAI-R ratings for each week of the study. These ratings are also graphed in Figure 2. Table 4 shows the subjects' transformed mean ratings for each week through the week following their feedback session. Table 5 depicts the relative rank of each subject's transformed mean rating during the week following his or her participation in the SEIP. Table 5 also shows the value of the $R_n$ statistic.
### Table 3. Mean Ratings of Subjects' Teaching Performances on TPAI-R

<table>
<thead>
<tr>
<th>Subject</th>
<th>Week</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>A</td>
<td>4.23</td>
<td>3.98a</td>
<td>4.20</td>
<td>4.35</td>
<td>4.23</td>
<td>4.05</td>
<td>4.00</td>
</tr>
<tr>
<td>B</td>
<td>4.43</td>
<td>4.15</td>
<td>4.25</td>
<td>4.33</td>
<td>4.30</td>
<td>4.40</td>
<td>4.18</td>
</tr>
<tr>
<td>C</td>
<td>4.20</td>
<td>4.15</td>
<td>4.28</td>
<td>4.50</td>
<td>4.40</td>
<td>4.33</td>
<td>4.18</td>
</tr>
<tr>
<td>D</td>
<td>3.44</td>
<td>3.80</td>
<td>4.05</td>
<td>3.80</td>
<td>3.53</td>
<td>3.73</td>
<td>3.73</td>
</tr>
<tr>
<td>E</td>
<td>4.05</td>
<td>4.15</td>
<td>3.98</td>
<td>4.20</td>
<td>4.05</td>
<td>4.00</td>
<td>3.75</td>
</tr>
<tr>
<td>F</td>
<td>4.10</td>
<td>4.20</td>
<td>4.48</td>
<td>4.18</td>
<td>4.13</td>
<td>3.93</td>
<td>3.90</td>
</tr>
</tbody>
</table>

*aUnderline indicates that the subject participated in the SEIP during the previous week.

### Table 4. Transformed Mean Ratings of Subjects' Teaching Performances on TPAI-R

<table>
<thead>
<tr>
<th>Subject</th>
<th>Week</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>A</td>
<td>---</td>
<td>-.059a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>---</td>
<td>-.062</td>
<td>-.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>---</td>
<td>-.012</td>
<td>.024</td>
<td>.070</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>---</td>
<td>.105</td>
<td>.119</td>
<td>.010</td>
<td>-.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>---</td>
<td>.025</td>
<td>-.031</td>
<td>.035</td>
<td>-.011</td>
<td>-.021</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>---</td>
<td>.024</td>
<td>.078</td>
<td>-.019</td>
<td>-.027</td>
<td>-.069</td>
<td>-.064</td>
</tr>
</tbody>
</table>

*aUnderline indicates that the subject participated in the SEIP during the previous week.
Figure 2. Subjects' Mean TPAI-R Ratings for Each Week of Study. aIndicates treatment introduced.
Table 5. Rank of Each Subject's TPAI-R Transformed Mean Rating for Week Following SEIP Feedback Session

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td>15^a</td>
</tr>
</tbody>
</table>

^a The sum of the ranks constitutes the Rn statistic, which is not significant at the .05 level.

A Priori Contrasts

Tukey's HSD procedure was used to statistically compare the transformed TPAI-R mean ratings of pairs of subjects during each week of the study. A .05 experiment-wise alpha level was used to determine statistical significance. It had been predicted that for each week of the study, the transformed mean rating of the subject who had participated in the SEIP during the previous week would be significantly greater than the ratings of the subjects who had not yet participated in the SEIP but would not differ from the ratings of those subjects who had already participated in the SEIP.

As can be seen in Tables 6-11, none of these a priori contrasts were found to be statistically significant. This means that the subjects' mean ratings on the TPAI-R did not differ from one another irrespective of whether or not they had participated in the SEIP. Consequently, the prediction
concerning expected differences in ratings between those
who had participated in the SEIP and those who had not yet
participated was not supported.

Table 6. A Priori Contrasts Using Tukey's HSD Procedure
for Week 2

<table>
<thead>
<tr>
<th>Subject</th>
<th>A</th>
<th>C</th>
<th>F</th>
<th>E</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>.004</td>
<td>.051</td>
<td>.088</td>
<td>.092</td>
<td>.170</td>
</tr>
<tr>
<td>A&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>.047</td>
<td>.084</td>
<td>.088</td>
<td>.166</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>.037</td>
<td>.041</td>
<td>.119</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td>.004</td>
<td>.082</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td>.078</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Participated in SEIP during previous week.

Table 7. A Priori Contrasts Using Tukey's HSD Procedure for
Week 3

<table>
<thead>
<tr>
<th>Subject</th>
<th>E</th>
<th>A</th>
<th>C</th>
<th>F</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>B&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.024</td>
<td>.055</td>
<td>.055</td>
<td>.110</td>
<td>.151</td>
</tr>
<tr>
<td>A</td>
<td>.031</td>
<td>.031</td>
<td>.086</td>
<td>.127</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.000</td>
<td>.055</td>
<td>.097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>.055</td>
<td>.097</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Participated in SEIP during previous week.
### Table 8. A Priori Contrasts Using Tukey's HSD Procedure for Week 4

<table>
<thead>
<tr>
<th>Subject</th>
<th>D</th>
<th>B</th>
<th>E</th>
<th>A</th>
<th>C&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>.029</td>
<td>.031</td>
<td>.054</td>
<td>.072</td>
<td>.089</td>
</tr>
<tr>
<td>D</td>
<td>.002</td>
<td>.025</td>
<td>.043</td>
<td>.060</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>.023</td>
<td>.041</td>
<td>.058</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td>.018</td>
<td>.035</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td>.017</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Participated in SEIP during previous week

### Table 9. A Priori Contrasts Using Tukey's HSD Procedure for Week 5

<table>
<thead>
<tr>
<th>Subject</th>
<th>F</th>
<th>E</th>
<th>B</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>D&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.039</td>
<td>.058</td>
<td>.073</td>
<td>.076</td>
<td>.094</td>
</tr>
<tr>
<td>F</td>
<td>.019</td>
<td>.034</td>
<td>.037</td>
<td>.055</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>.000</td>
<td>.018</td>
<td>.036</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td>.003</td>
<td>.021</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td>.018</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Participated in SEIP during previous week
Table 10. A Priori Contrasts Using Tukey's HSD Procedure for Week 6

<table>
<thead>
<tr>
<th>Subject</th>
<th>A</th>
<th>E&lt;sup&gt;a&lt;/sup&gt;</th>
<th>D</th>
<th>C</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>.034</td>
<td>.047</td>
<td>.070</td>
<td>.073</td>
<td>.094</td>
</tr>
<tr>
<td>A</td>
<td>.013</td>
<td>.036</td>
<td>.039</td>
<td>.060</td>
<td></td>
</tr>
<tr>
<td>E&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>.023</td>
<td>.026</td>
<td>.047</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>.003</td>
<td>.024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td>.021</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Participated in SEIP during previous week

Table 11. A Priori Contrasts Using Tukey's HSD Procedure for Week 7

<table>
<thead>
<tr>
<th>Subject</th>
<th>F&lt;sup&gt;a&lt;/sup&gt;</th>
<th>A</th>
<th>C</th>
<th>B</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>.014</td>
<td>.040</td>
<td>.048</td>
<td>.048</td>
<td>.080</td>
</tr>
<tr>
<td>F&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.026</td>
<td>.034</td>
<td>.034</td>
<td>.066</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>.008</td>
<td>.008</td>
<td>.040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.000</td>
<td>.032</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>.032</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Participated in SEIP during previous week

SEIP Participant Questionnaire Analysis

The three quantitative questions on the SEIP Participant Questionnaire were analyzed by examining the participants' median ratings. The least favorable response was assigned a rating of one and the most favorable response
was assigned a rating of five. Tables 12-14 depict the responses of the SEIP participants to each question as a whole as well as broken down by SEIP role (subjects, peers, supervisors, and other educators). Questionnaires were completed by every participant in the study except for one supervisor.

Table 12 shows the responses to the first question, which asked: "In general, how beneficial do you feel the SEIP feedback session is for helping special educators improve their teaching?" As a group, the participants indicated that the SEIP feedback was very beneficial (median rating of 4.03) for helping special education teachers to improve their instruction. Among the four subgroups, the supervisors' ratings were most positive (very beneficial to extremely beneficial). These results support the subhypothesis that the participants in this study would assess the SEIP feedback session as being very beneficial or extremely beneficial for helping special educators improve their instruction.

Table 13 depicts the responses to the second question, which asked: "How beneficial to you was your participation in the SEIP feedback session?" Overall, the participants indicated that their participation in the SEIP was between moderately beneficial and very beneficial to them (median rating of 3.5). The supervisors' ratings (very beneficial) and the subjects' ratings (very
beneficial) were most positive while the peers' ratings
(somewhat beneficial to moderately beneficial) were least
positive. These results do not support the prediction
that the participants would assess their participation in
the SEIP feedback session as being very beneficial or
extremely beneficial.

Table 14 shows the responses to the third quanti-
tative question, which asked: "How adequate is the TPAI-R
for assessing special educators' teaching?" As a whole,
the SEIP participants rated the TPAI-R as being a good
instrument (median rating of 3.80) for assessing special
educators' instruction.

The subjects', peers' and supervisors' ratings
(good) were most favorable, while the other educators'
ratings (satisfactory) were least positive. These results
support the prediction that the participants would rate the
TPAI-R as being good or excellent for assessing the instruc-
tion of special education teachers.
Table 12. Frequency of Responses to First Quantitative Question on SEIP Participant Questionnaire: "In general, how beneficial do you feel the SEIP feedback session is for helping special educators improve their instruction?"

<table>
<thead>
<tr>
<th>Participant Role</th>
<th>Rating</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
<th>Median Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Beneficial</td>
<td>Somewhat Beneficial</td>
<td>Moderately Beneficial</td>
<td>Very Beneficial</td>
<td>Extremely Beneficial</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>Peer</td>
<td>2</td>
<td>4</td>
<td></td>
<td>6</td>
<td>3.75</td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td></td>
<td>4.50</td>
<td></td>
</tr>
<tr>
<td>Other Educator</td>
<td>5</td>
<td>1</td>
<td></td>
<td>6</td>
<td>4.10</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>15</td>
<td>4</td>
<td>22</td>
<td>4.03</td>
<td></td>
</tr>
</tbody>
</table>
Table 13. Frequency of Responses to Second Quantitative Question on SEIP Participant Questionnaire: "How beneficial to you was your participation in the SEIP feedback session?"

<table>
<thead>
<tr>
<th>Participant Role</th>
<th>Rating</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
<th>Median Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Beneficial</td>
<td>Somewhat Beneficial</td>
<td>Moderately Beneficial</td>
<td>Very Beneficial</td>
<td>Extremely Beneficial</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>3.83</td>
</tr>
<tr>
<td>Peer</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>2.50</td>
</tr>
<tr>
<td>Supervisor</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4.00</td>
</tr>
<tr>
<td>Other Educator</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>3.17</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>22</td>
<td>3.50</td>
</tr>
</tbody>
</table>
Table 14. Frequency of Responses to Third Quantitative Question on SEIP Participant Questionnaire: "How adequate is the revised Teacher Performance Assessment Instruments (TPAI-R) for assessing special educators' teaching?"

<table>
<thead>
<tr>
<th>Participant Role</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poor</td>
</tr>
<tr>
<td>Subject</td>
<td>1</td>
</tr>
<tr>
<td>Peer</td>
<td>1</td>
</tr>
<tr>
<td>Supervisor</td>
<td>1</td>
</tr>
<tr>
<td>Other Educator</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
</tr>
</tbody>
</table>
CHAPTER 4

DISCUSSION

Overall, the results of this study are inconclusive regarding the effectiveness of the Special Educator Improvement Procedure (SEIP) in assisting special education teachers improve their instruction. On the one hand, subjects who participated in the SEIP were not rated significantly higher on the revised Teacher Performance Assessment Instruments (TPAI-R) than were subjects who had not yet participated in the SEIP. On the other hand, study participants assessed the SEIP as being very beneficial for helping special educators improve their teaching.

Analysis of Statistical Test Results

Despite the expectation that the effectiveness of the SEIP would be clearly demonstrated, the inconclusive findings should perhaps not have been entirely unexpected. McDonald (1980a) has pointed out that because our understanding of teaching is limited and imperfect, our expectations as to how much improvement in instruction will result from teacher evaluation procedures should be modest. In addition, the difficulty of demonstrating the
causal relationship between evaluation and instructional improvement is evidenced by the absence of empirical studies in the professional literature showing that evaluation results in improved instruction.

In the current study there are many possible reasons why the subjects did not demonstrate statistically significant changes on the TPAI-R ratings after participating in the SEIP. First, the SEIP may not have been powerful enough to have effected measurable change in the subjects. This explanation appears to be highly feasible. The single, relatively short feedback session may have been insufficient in frequency and/or duration to have evoked change in the subjects' teaching. It may have been a case of right medicine but wrong dosage. It is reasonable to believe that just as teachers have acquired their current skills over an extended period of time, it will also take time for them to learn and demonstrate new skills. The experimenter had considered using longer feedback sessions on more than one occasion but did not implement this idea due to the unlikelihood of recruiting volunteers who would agree to participate for longer periods of time. Should the study be replicated, it would seem advisable to use multiple feedback sessions (perhaps three) lasting at least two hours each as a way of increasing the power of the procedure.
A second possible explanation as to why this study failed to demonstrate statistically significant results is also related to time. Because the study was of such short duration (7 weeks), there may not have been enough time for the SEIP to have had its full impact. This explanation appears less feasible because there are no trends in the data to suggest that the subjects' instruction would have improved significantly had the study been extended. However, were multiple feedback sessions to be used in subsequent studies, it might be advisable to extend them over one or two semesters to give the teachers time to absorb the feedback provided and to implement suggested changes in their instruction.

Third, the instrument used in the study (TPAI-R) may not have been sensitive enough to have detected changes in the subjects' teaching. This explanation is also feasible. The Teacher Performance Assessment Instruments (TPAI) was not specifically developed to measure instructional change among special education teachers. Instead, the TPAI was designed to measure generic competencies applicable to all teachers. Thus, small subtle changes in the subjects' behavior resulting from participation in the SEIP may not have been detected by the TPAI-R.

Fourth, because the instructional periods that were videotaped each week varied, the TPAI-R ratings may have reflected differences in lessons or instructional
periods (e.g., small group reading session versus large group health lesson) rather than changes in the quality of instruction. Although a variety of lesson types was purposely videotaped to obtain a more representative sample of each subject's instruction, this procedure may have reduced the possibility of demonstrating significant changes in instruction. The feasibility of this explanation is further supported by the fact that the subjects received feedback concerning only one or two specific instructional periods during the SELP feedback session. Thus, even if they were to change their behavior in respect to those instructional periods, this change may not have been detected unless they were videotaped during those periods in subsequent recording sessions. For purposes of demonstrating change, then, it might have been better to have videotaped the subjects during the same instructional period for each week of the study.

A fifth, highly feasible explanation for the failure to obtain statistically significant findings in this study relates to the subjects. Although the experimenter made a considerable effort to recruit less able teachers (i.e., teachers who were inexperienced or reported to be in need of improvement), few of these teachers were willing to volunteer for the study. To the contrary, most of the volunteers for the study were experienced teachers who already possessed moderately good skills. As a result,
two of the six subjects selected for the study did not meet the selection criterion of receiving pretreatment ratings above the minimum competency level on fewer than 75% of the items on the TPAI-R. Thus, even though all of the subjects were rated on the TPAI-R as needing to improve in some areas, there were fewer competencies on which the better teachers could demonstrate significant gains over the course of the study. Were this study to be replicated with less skilled teachers, it is possible that they would demonstrate greater gains following participation in the SEIP.

One can only speculate as to why the better teachers were more willing to volunteer for this study than were the less able teachers. It may be that better teachers, who are often more experienced, feel more comfortable with evaluation, while less able teachers, who are often less experienced, feel uncomfortable and perhaps even threatened by any type of evaluation. It is also possible that good teachers continually seek ways to improve, such as participation in the SEIP, while poor teachers avoid such opportunities, particularly when there is risk involved or when there is extra time required.

A sixth tenable reason as to why the predicted outcomes were not supported relates to the statistical tests. Whenever statistically insignificant results are obtained, it is always possible that the predicted differences exist but were not discovered because the statistical
tests lacked sufficient power to detect the differences. This possibility is unlikely in this case because there is no trend in the data to indicate that the predicted differences may exist.

Seventh, researchers (Brock, 1981; Fuller & Manning, 1973) have found that teachers tend to demonstrate greater improvement in instruction when their self assessments are more positive than the assessments of them made by others. As can be seen in Table 15, the subjects' TPAI-R ratings of themselves in the current study were not greater than the ratings made by the other feedback team members in five of six cases. Thus, because the subjects saw themselves as others saw them, there was less probability that they would make changes in their teaching for the purpose of reducing the discrepancy between their self ratings and the ratings of them made by others.

Table 15. Mean TPAI-R Ratings by Subjects and by Other Feedback Team Members

<table>
<thead>
<tr>
<th>Team</th>
<th>Subjects' Self Assessment</th>
<th>Consensus Assessment by Other Feedback Team Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.6</td>
<td>4.6</td>
</tr>
<tr>
<td>B</td>
<td>4.6</td>
<td>4.5</td>
</tr>
<tr>
<td>C</td>
<td>4.0</td>
<td>4.7</td>
</tr>
<tr>
<td>D</td>
<td>3.8</td>
<td>4.0</td>
</tr>
<tr>
<td>E</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>F</td>
<td>4.2</td>
<td>4.4</td>
</tr>
</tbody>
</table>
Eighth, there appears to be an environmental or historical factor that may have mitigated against finding the predicted results. Due to the fact that it took longer to recruit the necessary volunteers for the study than anticipated, the study was conducted during the seven week period just prior to Christmas vacation. As a result, the quality of the subjects' instruction toward the end of the study may have been compromised by the subjects' anticipation of the approaching holidays. This assertion is supported by the general decrease in TPAI-R ratings among all subjects during the last two weeks of the study, as can be seen on Table 3 and Figure 2 in Chapter 3 (pp. 75-76). During the sixth week of the study, the TPAI-R ratings for four of the six subjects were lower than the previous week and during the seventh week of the study, ratings for all six of the subjects were lower or equal to their TPAI-R ratings for the sixth week. Thus, any positive effect resulting from the subjects' participation in the SEIP may have been reduced or counteracted by the effect of "vacationitis". To reduce the possible influence of this factor, it would probably be advantageous for subsequent studies to be conducted during a time period that is not contiguous with a vacation.

Finally, it is possible that statistically significant results were not obtained because the SEIP is an ineffective procedure for helping special educators improve
their teaching. That is, even if the previously discussed threats to the internal validity of the study were eliminated, it is still possible that the SEIP would not produce the expected change in the subjects' behavior. However, while this explanation cannot be completely discounted, neither can it be unequivocally concluded that the SEIP is ineffective without first dismissing the other feasible explanations.

**Analysis of SEIP Participant Questionnaire Results**

Although statistical test results did not support predicted changes in the subjects' behavior resulting from participation in the SEIP, the participants' responses on the Special Educator Improvement Procedure (SEIP) Participant Questionnaire were generally supportive of the SEIP as being beneficial to special educators seeking to become better teachers. The participants' positive assessment of the SEIP is important because it provides evidence that the SEIP may be useful for its intended purpose of helping special education teachers improve their instruction. In addition, this finding provides support for the assertion that insignificant statistical results may have been found because of one or more of the previously discussed explanations. Thus, the SEIP may hold promise as an effective formative evaluation procedure despite the failure of this study to unequivocally support this position.
Responses to Quantitative Questions

A more careful analysis of the SEIP Participant Questionnaire responses produced some interesting findings. As shown in Table 12 in Chapter 3 (p. 83), the SEIP was rated as being a very beneficial to extremely beneficial procedure for helping special educators improve their instruction by both the participants as a whole as well as each of the subgroups (subjects, peers, supervisors, and other educators). It is interesting that the supervisors were more positive toward the SEIP than any other subgroup. Perhaps this is because they see the SEIP as something that can help them perform their role as teacher evaluator better or because the SEIP is a better procedure than others that they have used.

The participants' responses to the question that asked: "How beneficial to you was your participation in the SEIP feedback session?" (Table 13, p. 83) varied considerably. While the subjects and supervisors indicated that the SEIP feedback session was very beneficial to them, the peers and the other educators rated the feedback session as being only somewhat beneficial to moderately beneficial. This finding is not particularly surprising. Obviously, the subjects had the most to gain from the feedback sessions since the feedback was designed to directly affect their instruction. The supervisors benefitted because they were exposed to a potentially useful evaluation procedure as well
as given an opportunity to fulfill one of their responsibilities of giving guidance to one of their employees. On the other hand, the peers and other educators would not have benefitted as much from the feedback session since they had less at stake.

The participants' responses to the question regarding the adequacy of the TPAI-R for assessing special educators' teaching (Table 14, p. 85) were also varied, although the respondents as a group rated the TPAI-R as being a good instrument for this purpose. It is interesting that the subjects, all of whom were special education teachers, assessed the TPAI-R as being a better instrument for this purpose than did the other educators. Perhaps the subjects were better able to see the relevance of the competencies assessed by the TPAI-R to special educators than were the other educators. This finding helps to validate the revised Teacher Performance Assessment Instruments as a tool that can be useful in the evaluation of special education teachers.

Responses to Qualitative Questions

Although the narrative comments made by the respondents on the SEIP Participant Questionnaire were not used to support or refute the effectiveness of the Special Educator Improvement Procedure (SEIP), they were useful for identifying those SEIP components that were most useful as
well as those that were in need of revision. The most often cited benefit of the SEIP was that it allowed the subjects to see and hear themselves on videotape. The videotape playback component appears to provide an important source of feedback not obtainable elsewhere. This finding is congruent with previous research that found teachers benefitted from being able to view their own instruction (Allender & Yanoff, 1977).

The second most often cited benefit of the SEIP was that subjects had the opportunity to receive feedback about their instruction from their peers. This finding also supports previous research that has demonstrated the benefit of using peers to help their colleagues improve their instruction (Educational Research Service, 1978; Herman, 1973; Williams, 1980).

Third, the study participants indicated that the SEIP was beneficial because it provided a vehicle for teachers to receive reinforcement for the things that they did well. This was an important benefit because positive feedback provided the teacher with information regarding those activities that represent good teaching practices and should be continued. Although it is worthwhile to identify weaknesses needing improvement, it is equally worthwhile to recognize areas of strength. Additionally, reinforcement may help to fuel teachers' motivation to continue to improve their instruction.
The SEIP study participants also indicated that the SEIP was beneficial because it was non-threatening; was objective; was comprehensive; and allowed the feedback team to evaluate the videotapes conveniently. As a whole, these comments appear to support the importance of integrating individual evaluation components into a comprehensive package, as was done with the SEIP.

In addition to citing benefits of the SEIP, study participants made suggestions as to how the SEIP might be improved. They suggested that: additional videotaped samples be recorded and played back during the feedback session; entire lessons be videotaped rather than randomly selected parts of lessons; more than one feedback session be used; and more time for discussion be allowed during the feedback session. Most of these suggestions were considered by the experimenter while planning the study but were dismissed as being too time consuming to be used in a study relying upon volunteer time. However, were time not a limiting factor, these suggestions would certainly be worth exploring further.

It was also suggested that the SEIP feedback team members examine lesson plans of the videotaped instruction prior to the feedback session to give them a better sense of what was being done and why it was being done. This too is a good suggestion and is one that was employed by Georgia's Teacher Assessment Project (Capie, Anderson,
Johnson & Ellett, 1979). For the current study, however, this suggestion has less value because complete lessons were not videotaped. Thus, lesson plans may not have been as useful in providing guidance to the viewers of the videotape recordings as they were in the Teacher Assessment Project.

Another suggestion made by the SEIP members was to videotape the students as well as the teacher so that the students' reactions could be observed. This suggestion would be worthwhile to explore in subsequent research. Only the special education teachers were videotaped in the current study in following with a suggestion made by Bailey (1979).

**Appropriateness of the SEIP for Regular Educators**

Although the question as to how appropriate the Special Educator Improvement Procedure (SEIP) would be for use with regular educators was not specifically addressed by this study, it is one that would be interesting to investigate in future research. In keeping with the majority of evaluation authorities' contention that special educators and regular educators may be evaluated with common procedures and techniques (Dorwood, 1963; Howe, 1981; Moya, 1980), one might hypothesize that the SEIP would be equally applicable to regular educators and
special educators. To test this hypothesis, it is suggested that the SEIP be implemented with both special education teachers and regular education teachers in subsequent studies.

**Implications for Future Research**

The results of the current study in terms of demonstrating statistically significant behavior change do not support the Special Educator Improvement Procedure (SEIP) as being an effective procedure for helping special educators improve their instruction. However, the participants' positive assessment of the SEIP indicates that it holds promise for evolving into a useful evaluation procedure. With some modification, it is feasible that the SEIP's effectiveness could be demonstrated in subsequent research studies. To test this assertion, additional research needs to be conducted with a revised version of the SEIP. Based upon comments made by study participants as well as observations by the experimenter, a number of modifications might be made to the SEIP. First, the number of SEIP feedback sessions could be increased from one to three so that the evaluatee will have more opportunity to receive information that could be useful in making instructional changes. Concurrently, the length of the study could be extended to at least an entire academic semester to give the teacher being evaluated more time to implement
changes in his or her instruction. The additional time needed to effect this change could possibly be found through incorporating the use of the SEIP into a school's inservice schedule.

Second, the SEIP could be improved by playing back an entire videotaped lesson and providing the feedback team members with the corresponding lesson plan at the feedback session. In this way, the feedback team members would have a better understanding of the lesson rationale and objectives, as well as a more complete sample of the evaluatee's teaching performance. In addition, the same instructional period or subject area could be videotaped during subsequent weeks to better assess change resulting from the SEIP.

Third, it seems advisable to videotape both the teacher being evaluated and his or her students. Feedback team members would thus be able to see what the teacher does as well as the students' reactions to what is done. The split screen videotaping approach recommended by Moritz and Martin-Reynolds (1980) would appear to be appropriate for this purpose. Two cameras would be used to simultaneously record the behavior of the teacher and students. The resulting videotape would show the teacher on one half of the monitor screen and the students on the other half of the screen, similar to what is done for some sports broadcasts.
Fourth, the effectiveness of the SEIP in demonstrating instructional improvement could possibly be enhanced through using an evaluation instrument that is more sensitive to change than the TPAI-R. Although the experimenter is not aware of a better instrument for this purpose, it is possible that a more appropriate evaluation tool may be developed in the future.

Finally, evaluatees may need to be given specific instruction as to how they should change unsatisfactory behaviors subsequent to learning what behaviors need to be improved. It is recognized that some behaviors, such as demonstrating empathy and understanding, are difficult to learn and may require additional guidance for change to occur.

In addition to making changes in the SEIP itself, it is recommended that future research with the SEIP be conducted with both regular and special educators to determine the appropriateness of the procedure for both groups. It is also suggested that less able teachers be recruited as subjects for the SEIP so as to maximize the amount of improvement that could be made. The probability of recruiting less experienced and less skilled teachers could be enhanced if the SEIP was used in a school district where teachers are required to participate in evaluation for instructional improvement.
Besides assessing the effectiveness of a modified version of the SEIP, future research should also address the effectiveness and cost efficiency of the SEIP in relation to other current and newly developed evaluation procedures. Such research would provide educators with valuable information as to which evaluation procedures are most cost effective in assisting educators, in general, and special educators, in particular, to improve their instruction and, thus, become better teachers.

**Implications for Teacher Evaluation**

Although it is the experimenter's belief that the effectiveness of the SEIP can be demonstrated after making the previously discussed modifications, the possibility that the SEIP is an ineffective procedure cannot be ruled out at this time. If this is true, then we must begin to question what we "know" about how to evaluate and help teachers improve their teaching. This conclusion would cast doubt upon previous research supporting the use of videotaped feedback, self assessment, peer evaluation, and competency-based assessment as effective formative evaluation procedures. It would also challenge the blind, unquestioning manner in which these and other teacher evaluation procedures and instruments are implemented matter of factly without considering whether they are effective or not.
Another important implication of this study for teacher evaluation is related to the allocation of resources, such as money and personnel time, for evaluation purposes. Currently, most school systems devote little money or time to teacher evaluation, resulting in the evaluation procedures' effectiveness paralleling the magnitude of the resources that are allocated. However, even through expending more resources than are normally devoted to teacher evaluation, the SEIP's effectiveness was not established. This may indicate that (1) even more resources need to be devoted to teacher evaluation for the effectiveness of current procedures to be demonstrated or (2) more effort needs to be devoted to finding alternative evaluation practices to replace those that are ineffective. In either case, the educational community will need to make a greater commitment to teacher evaluation, both in terms of policy and resources, if it is to serve its intended purpose of helping teachers become better educators.
APPENDIX A

DESCRIPTION OF THE STUDY FOR PARTICIPANTS

Purpose This is a doctoral dissertation research study to test a procedure (SEIP) designed to assist special education teachers in improving their classroom instruction.

Participants Needed Special education teachers of self contained classes and their supervisors.

Procedure Each participating teacher will be videotaped in his/her classroom twice a week (30-60 minutes each time) for eight weeks. Once during this period, a team consisting of the teacher, his/her supervisor, a peer teacher, and another educator chosen by the teacher will view portions of the 2 videotapes produced during 1 week of the study and assess them using a competency based assessment instrument. Based upon this assessment, team members will give the teacher feedback about strengths and weaknesses and make suggestions as to how the teacher might improve his/her instruction. Following the eight week period, the teacher and his/her supervisor will view and assess portions of videotapes produced during the last week of the study. Team members will also be asked to complete a short questionnaire about the SEIP following completion of the study.

Benefits of Participating
1. Opportunity for teacher to: a) view own teaching; b) receive feedback from others about his/her teaching; and c) receive suggestions for improvement.
2. Opportunity for other team members to acquire new ideas through observing teacher.

Time Commitment 3-4 hours for teacher and supervisor. 2-3 hours for other team members.

Investigator Chris Myers, a doctoral candidate in the departments of educational psychology and special education at the University of Arizona. He can be contacted at 622-8474 or 791-6138.
APPENDIX B

REVISED TEACHER PERFORMANCE ASSESSMENT INSTRUMENTS (TPAI-R)

Name of Rater

Name of Teacher Rated

Please rate the teacher's videotaped performance using the following 20 items or indicators. Each indicator is to be rated 1, 2, 3, 4, 5 or NA. NA should be chosen only when an indicator does not apply to the videotaped segments observed or when the videotaped segments do not allow you to determine if a particular behavior did or did not occur.

There are two types of items on the TPAI-R. The first type of item (e.g., item 1) has five statements pertaining to the stem and a NA (not applicable) statement. Circle the appropriate number (or NA) for each indicator. On the second type of item (e.g., item 9), circle the rating in the left hand column (or NA) that corresponds to the number of descriptors in the right hand column that were observed.
Competency I: Communicates with learners.

1. Gives directions and explanations related to lesson content.

   NA Not applicable/no opportunity to observe.

   1. Fails to give any directions or explanations (either written or oral) when there is an obvious need to do so.

       OR

       1. Directions and explanations are difficult to understand and no attempt is made to remedy the confusion.

       2. Directions or explanations are difficult to understand. Attempts to clarify confusion are largely ineffective.

       3. Although most learners appear to understand, the teacher works with the entire group to clarify misunderstandings.

       4. Only a few learners misunderstand. The teacher identifies specific learners who have difficulty with directions and explanations and helps them individually.

       5. No evidence of learner confusion about directions or explanations is evident.

2. Clarifies directions and explanations when learners misunderstand lesson content.

   NA Not applicable/no opportunity to observe.

   1. Discourages learners when they seek clarification on directions or explanations.

   2. Ignores learners when they seek clarification of directions or explanations.

   3. Restates original communication in nearly the same words if learners do not understand.

   4. Gives directions or explanations using different words and ideas when learners do not understand.

   5. In addition to the items in 4, the teacher attempts to identify areas of misunderstanding and to restate communication before learners ask.

       OR

       5. No misunderstanding by learners was evident during the lesson.

3. Uses responses and questions from learners in teaching.

   NA Not applicable/no opportunity to observe.

   1. Uses negative words or actions to discourage learners from giving responses or asking questions.

   2. Ignores learners who wish to be recognized or learner contributions are accepted without disagreement or further comment.
3. Acknowledges learners who wish to be recognized and occasionally asks for learner responses or questions. Responses by the teacher are adequate.

4. Asks for responses or questions frequently throughout the lesson and provides feedback to learners.

5. In addition to the items in 4, the teacher incorporates learner responses and questions into activities.

4. Provides feedback to learners throughout the lesson.

   NA Not applicable/no opportunity to observe.
   1. Accepts learner comments or performance without feedback about their adequacy.
   2. Responds to negative aspects of student work, but few comments are made about positive aspects.
   3. Informs students of the adequacy of their performance. Few errors pass by without being addressed.
   4. Helps learners evaluate the adequacy of their own performance.
   5. In addition to 4, the teacher probes for the source of misunderstandings which arise.

Competency II: Demonstrates a repertoire of teaching methods.

5. Implements learning activities in a logical sequence.

   NA Not applicable/no opportunity to observe.
   1. Activities used in the classroom are unrelated to one another.
   2. Many ideas, skills or activities seem out of sequence.
   3. The lesson is arranged to present most ideas, skills, etc., in a logical sequence. Only occasionally is there a problem of sequence.
   4. No instances of problems in sequencing are noted.
   5. In addition to the items in 4, provision is made to acquire prerequisites if learners have not already done so.

6. Demonstrates ability to conduct lessons using a variety of teaching methods.

   Ratings               Descriptors
   NA Not applicable/no opportunity to observe.
   1. No teaching method is used acceptably.
   2. One teaching method is used acceptably.

   Teaching methods such as the following may be observed: drill, inquiry, discussion, role playing, demonstration, explanation, and problem solving, etc.
3. Two teaching methods are used acceptably.
4. Three teaching methods are used acceptably.
5. Four teaching methods are used acceptably.

Competency III: Reinforces and encourages learner involvement in instruction.

7. Provides learners with opportunities for participating.

NA Not applicable/no opportunity to observe.
1. Class activities require passive commitment.
2. The class is organized so that only a few learners participate actively; most appear to be bystanders.
3. Most learners have opportunity for active participation at some time in the class (e.g., small group discussion, physical manipulation of materials, physical movement, individual library work, etc.).
4. All learners have opportunity for active participation in some type of activity.
5. All learners have opportunity for active participation in two or more activities.

8. Maintains learner involvement in lessons.

NA Not applicable/no opportunity to observe.
1. Few learners (less than 30%) are on task, i.e., learners are not attending to the teacher, materials, or other appropriate focus for an activity.
2. Some learners (about 31-50%) are on task.
3. Many learners (about 51-70%) are on task.
4. Most learners (about 71-90%) are on task.
5. Nearly all learners (about 91-100%) are on task.

9. Reinforces and encourages the efforts of learners to maintain involvement.

**Ratings**

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>NA</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses activities which are appropriate for learners.</td>
<td>Not applicable/no opportunity to observe.</td>
<td>Uses activities which are appropriate for learners.</td>
<td>Varies pace and nature of activity.</td>
<td>Responds positively to learners who participate.</td>
<td>Identifies and responds to learners who are off task.</td>
<td>Four of the descriptors are evident.</td>
</tr>
<tr>
<td>Varies pace and nature of activity.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responds positively to learners who participate.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifies and responds to learners who are off task.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four of the descriptors are evident.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Competency IV: Organizes time, space, materials, and equipment for instruction.

10. Attends to routine tasks.

<table>
<thead>
<tr>
<th>NA</th>
<th>Not applicable/no opportunity to observe.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The teacher does not attend to routine tasks.</td>
</tr>
<tr>
<td>2.</td>
<td>The teacher attends to routine tasks in a disruptive or inefficient manner (e.g., learners need special permission for many routine tasks.)</td>
</tr>
<tr>
<td>3.</td>
<td>The teacher anticipates routine tasks and attends to them efficiently.</td>
</tr>
<tr>
<td>4.</td>
<td>Routine tasks are handled smoothly. Teacher delegates many tasks to the students.</td>
</tr>
<tr>
<td>5.</td>
<td>In addition to 4, learners are responsible for various dimensions of the task (e.g., distributing materials, picking up work area, etc.).</td>
</tr>
</tbody>
</table>

11. Uses instructional time efficiently.

<table>
<thead>
<tr>
<th>Ratings</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors is evident.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
<tr>
<td>5.</td>
<td>Four of the descriptors are evident.</td>
</tr>
</tbody>
</table>

| a. | Begins activities promptly. |
| b. | Continues activities until end of allocated time period. |
| c. | Avoids unnecessary delays during the lesson. |
| d. | Avoids undesirable digressions from the topic during the lesson. |

Competency V: Demonstrates enthusiasm for teaching and learning and the subject being taught.

12. Communicates personal enthusiasm.

<table>
<thead>
<tr>
<th>Ratings</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors is evident.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
</tbody>
</table>

| a. | Communicates enthusiasm with eye contact or facial expressions indicating pleasure, concern, interest, etc. |
| b. | Communicates enthusiasm with voice inflections stressing points of interest and importance. |
5. Four of the descriptors are evident.

Communicates enthusiasm through posture when moving about the room or sitting among students.

Communicates enthusiasm with gestures to accentuate points.

13. Stimulates learner interest.

<table>
<thead>
<tr>
<th>Ratings</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA Not applicable/no opportunity to observe.</td>
<td>a. Appears eager to begin lesson.</td>
</tr>
<tr>
<td>1. None of the descriptors is evident.</td>
<td>b. Uses interesting, unusual or important dimensions or application of the topic or activity.</td>
</tr>
<tr>
<td>2. One of the descriptors is evident.</td>
<td>c. Attempts in a manner that stimulates interest to involve all learners in activity.</td>
</tr>
<tr>
<td>3. Two of the descriptors are evident.</td>
<td>d. Personalizes lesson for students.</td>
</tr>
<tr>
<td>4. Three of the descriptors are evident.</td>
<td></td>
</tr>
<tr>
<td>5. Four of the descriptors are evident.</td>
<td></td>
</tr>
</tbody>
</table>

Competency VI: Helps learners develop positive self-concepts.


<table>
<thead>
<tr>
<th>Ratings</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA Not applicable/no opportunity to observe.</td>
<td>a. Seeks information about the interests or opinions of learners.</td>
</tr>
<tr>
<td>1. None of the descriptors is evident.</td>
<td>b. Smiles at learners or laughs or jokes with them.</td>
</tr>
<tr>
<td>2. One of the descriptors is evident.</td>
<td></td>
</tr>
<tr>
<td>3. Two of the descriptors are evident.</td>
<td>c. Maintains close contact with learners by sitting or standing near them.</td>
</tr>
<tr>
<td>4. Three of the descriptors are evident.</td>
<td></td>
</tr>
<tr>
<td>5. Four of the descriptors are evident.</td>
<td>d. Uses names of learners in a warm and friendly way when addressing them.</td>
</tr>
</tbody>
</table>

15. Demonstrates sensitivity to the needs and feelings of learners.
### Ratings

<table>
<thead>
<tr>
<th></th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors b. Encourages learners when they have difficulty.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident. c. Listens to or accepts ideas from learners. d. Is courteous when dealing with learners.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
<tr>
<td>5.</td>
<td>Four of the descriptors are evident.</td>
</tr>
</tbody>
</table>

16. Demonstrates patience, empathy and understanding.

<table>
<thead>
<tr>
<th></th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors is evident.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
<tr>
<td>5.</td>
<td>Four of the descriptors are evident.</td>
</tr>
</tbody>
</table>

#### Descriptors

- a. Is patient with wrong answers or poor performance.
- b. Is patient with learners who need additional time or explanation.
- c. Avoids sarcasm or ridicule.
- d. Shows students through words or actions that their problems or comments are understood.

**Competency VII: Manages classroom interactions.**

17. Provides feedback to learners about their behavior.

<table>
<thead>
<tr>
<th></th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors is evident.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
<tr>
<td>5.</td>
<td>Four of the descriptors are evident.</td>
</tr>
</tbody>
</table>

#### Descriptors

- a. Makes expectations about behavior clear to learners.
- b. Provides verbal feedback for acceptable or unacceptable behavior.
- c. Provides nonverbal feedback (smiles, frowns, nods, moves closer to student, etc.) for acceptable or unacceptable behavior.
- d. Uses language free of derogatory references when talking to learners.
18. Promotes comfortable interpersonal relationships.

<table>
<thead>
<tr>
<th>Ratings</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors is evident.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
<tr>
<td>5.</td>
<td>Four of the descriptors are evident.</td>
</tr>
</tbody>
</table>

- a. Speaks politely with learners.
- b. Courteous interchanges among learners exist or are encouraged.
- c. Expectations are consistent throughout the lesson.
- d. Is fair and impartial when dealing with learners.

19. Maintains appropriate classroom behavior.

<table>
<thead>
<tr>
<th>Ratings</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors is evident.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
<tr>
<td>5.</td>
<td>Four of the descriptors are evident.</td>
</tr>
</tbody>
</table>

- a. Uses techniques (e.g., social approval, contingent activities, punishment, etc.) to maintain appropriate behavior.
- b. Overlooks inconsequential behavior problems.
- c. Reinforces appropriate behavior.
- d. Maintains learner behavior that enhances the possibilities for learning for the group.

20. Manages disruptive behavior among learners.

<table>
<thead>
<tr>
<th>Ratings</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Not applicable/no opportunity to observe.</td>
</tr>
<tr>
<td>1.</td>
<td>None of the descriptors is evident.</td>
</tr>
<tr>
<td>2.</td>
<td>One of the descriptors is evident.</td>
</tr>
<tr>
<td>3.</td>
<td>Two of the descriptors are evident.</td>
</tr>
<tr>
<td>4.</td>
<td>Three of the descriptors are evident.</td>
</tr>
<tr>
<td>5.</td>
<td>Four of the descriptors are evident.</td>
</tr>
<tr>
<td></td>
<td>OR</td>
</tr>
<tr>
<td>5.</td>
<td>There is no major disruptive behavior.</td>
</tr>
</tbody>
</table>

- a. Deals with learners who have caused disruptions, rather than with entire class.
- b. Attends to major disruptions quickly and firmly.
- c. The consequences of causing disruptions are based on the severity of the disruptions.
- d. Rule violations carry consequences appropriate for learners.
APPENDIX C

SPECIAL EDUCATOR IMPROVEMENT PROCEDURE (SEIP) PARTICIPANT QUESTIONNAIRE

Please take a few moments to respond to the following questions about the SEIP and your participation in the feedback session.

Your role in the SEIP feedback session (check one):

_____ videotaped participant
_____ peer from another school
_____ supervisor
_____ other educator

In general, how beneficial do you feel that the SEIP feedback session is for helping special educators improve their teaching?

not somewhat moderately very extremely
beneficial beneficial beneficial beneficial beneficial

What are the greatest strengths/benefits of the SEIP feedback session?

How might the SEIP feedback session be improved?

113
How beneficial to you was your participation in the SEIP feedback session?

- not beneficial
- somewhat beneficial
- moderately beneficial
- very beneficial
- extremely beneficial

What was most beneficial for you about participating in the SEIP feedback session?

How adequate is the revised Teacher Performance Assessment Instruments (TPAI-R) for assessing special educators' teaching?

- poor
- fair
- satisfactory
- good
- excellent

What are the TPAI-R's strengths?

How might the TPAI-R be improved?

Other comments:
APPENDIX D

SEIP SCRIPT

The following script was followed by the experimenter during implementation of the Special Educator Improvement Procedure (SEIP) with each subject. The feedback session was begun with the experimenter addressing the four members of the feedback team: "I'd like to outline the procedures that we'll be following today. My role will be to explain these procedures and to see that they are implemented. As a preface, I'd like to mention that this study is based on the premise that every teacher has strengths and that every teacher can improve.

The first thing that you'll do today is to have an opportunity to become familiar with the assessment instrument that we'll be using--the TPAI-R. You'll view a videotape of another special educator, and then rate her performance using the TPAI-R. Then, we'll discuss your ratings and determine the degree to which they are similar to one another. If your ratings are in general agreement with one another, we'll move on to the next step in the procedure. If not, you'll have additional opportunity to practice using the TPAI-R.
Secondly, you'll view two videotaped samples of (subject)'s teaching, which were obtained during the past week. Following this showing, you may ask (subject) questions to further clarify what you saw. You may also ask to view portions of these tapes again.

Third, you'll rate (subject)'s teaching using the TPAI-R.

Fourth, (subject) will be asked to leave the room. During this time, the remaining three of you (supervisor, peer, and other) will have an opportunity to discuss and reach consensus regarding your ratings. You'll also be able to discuss (subject)'s strengths and weaknesses and agree upon specific things that he (she) can do to improve his (her) teaching.

Fifth, (subject), you'll be invited back into the room, and will have an opportunity to compare your self ratings with the consensus ratings of the other team members.

Sixth, the three of you (supervisor, peer, and other) will be asked to specify those things that were done well by (subject) and make specific suggestions as to how areas of weakness might be improved.

Do you have any questions about what we're going to be doing today?" The experimenter will answer any questions.

"Next, let's discuss the rating scale that you'll be using today. Did everyone bring their copy of the revised Teacher Performance Assessment Instrument--the TPAI-R?
Does anyone have any questions about any of the items or how to use the instrument?" The experimenter will answer any questions.

"Now I'd like you to watch this videotape of a special education teacher. While you're watching the tape, think about how you will rate her performance on the TPAI-R." The feedback team will watch the training tape.

"Now use the TPAI-R to rate this teacher's performance. It's important that you match your ratings with what you saw on the videotape as closely as possible. This will increase the probability that your ratings will agree with one another. The ratings of three out of four of you must agree on 14 of the 20 items before we move on to the next step in the procedure." The feedback team will be given time to complete the TPAI-R.

"Now, after I call out the number of an item on the TPAI-R, I'd like each of you to tell me how you rated the teacher on that item." The experimenter will record the feedback team members' ratings on each item and tally these ratings. If the criterion for consensus is not met, the experimenter will say: "For each item on which there is not consensus, I'd like each of you to explain why you rated that item as you did. Through such discussion, I'd like you to try to reach consensus on each item. Following this discussion, you'll have an opportunity to view and rate
another videotape." This procedure will continue until the consensus criterion is reached.

"Next, I'm going to show you two videotaped samples of (subject)'s teaching that were recorded during the past week. I'd like you to watch both tapes and then to use the TPAI-R to rate them in the same way that you did the training tape(s). Even though these tapes were made at different times, consider them as a whole. If there are differences in how you would rate the two segments on a particular item, circle the higher of the two ratings. A particular behavior only has to be demonstrated on one of the two segments to be counted. Remember to base your ratings only on what you see on the videotapes. Do you have any questions?" The experimenter will answer questions, show the videotapes, and give the feedback team time to complete the TPAI-R.

The subject will be asked to leave the room and the experimenter will address the other three feedback team members: "The first thing that you need to do is to reach consensus about your ratings. I'll ask each of you how you rated each item on the TPAI-R. If two of you agree how an item should be rated, that will be considered consensus. If the three of you disagree, we'll discuss the item further until there is agreement by at least two of you. When all items are done, I'd like you to decide among yourselves where (subject)'s greatest strengths and weaknesses are,
and what specific things he (she) can do to improve his (her) teaching. Only those things that at least two of you agree on will be discussed with (subject).

The subject will be asked to return to the room.

"(Subject), while you were gone, the other team members discussed their ratings of the videotapes and reached consensus on them. They also agreed upon some things that you do well and some ways that you might be able to do what you do even better. The purpose of this portion of the procedure is to share that information with you.

(Subject), here are the ratings that were agreed upon by the other feedback team members. Take a few moments to compare them with your self ratings. Please feel free to ask the others any questions that you may have about their ratings." The subject will be given time to review and discuss the ratings with the other team members.

"Now, (supervisor, peer, and other), please briefly describe (subject)'s areas of strength." After the strengths are presented, the experimenter will say:

"Please specify what things are in need of improvement and how (subject) might go about improving his (her) teaching in these areas. I'll record your suggestions and give a copy to (subject)." The experimenter will verify the written suggestions with the group.

After the subject has been given a copy of these suggestions, the experimenter will say: "At the end of


the study, I'd like each of you to complete a short questionnaire to give me feedback about this procedure".
REFERENCES


