

1989

A study of reading achievement gains in classes of special education teachers using the Beginning Teacher Assistance Program indicators of competence

Cynthia Henshaw
College of William & Mary - School of Education

Follow this and additional works at: <https://scholarworks.wm.edu/etd>



Part of the [Educational Administration and Supervision Commons](#), and the [Special Education and Teaching Commons](#)

Recommended Citation

Henshaw, Cynthia, "A study of reading achievement gains in classes of special education teachers using the Beginning Teacher Assistance Program indicators of competence" (1989). *Dissertations, Theses, and Masters Projects*. William & Mary. Paper 1539618330.
<https://dx.doi.org/doi:10.25774/w4-2jav-9y42>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Dissertations, Theses, and Masters Projects by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 8923057

**A study of reading achievement gains in classes of special
education teachers using the Beginning Teacher Assistance
Program indicators of competence**

Henshaw, Cynthia, Ed.D.

The College of William and Mary, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A Study of Reading Achievement Gains in Classes
of Special Education Teachers
Using the Beginning Teacher Assistance Program
Indicators of Competence

Dissertation Presented to the Faculty
School of Education
College of William and Mary

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
Cynthia Henshaw
April, 1989

A Study of Reading Achievement Gains in Classes
of Special Education Teachers
Using the Beginning Teacher Assistance Program
Indicators of Competence

Cynthia E. Henshaw

Approved April, 1989

Lori Korinek

Lori A. Korinek, Ph. D

Robert J. Hanny

Robert J. Hanny, Ph. D

F. Douglas Prillaman

F. Douglas Prillaman, Ed. D
Chairman of Doctoral Committee

Dedication

This dissertation is dedicated to my family. To my husband for his pride and joy in my accomplishment. To my dear son who gave up a mother while I worked to complete it. To my lovely daughter who waited patiently for my attention on the many days when I worked writing and editing.

Acknowledgements

It is with heartfelt thanks that I acknowledge the help of the members of my committee. To Dr. Douglas Prillaman, I express sincere gratitude for his support, encouragement, help, and most of all for his friendship. He came to my aid always when I needed his help. To Dr. Korinek and Dr. Hanny, I express thanks for all of their suggestions for improving each draft of this research. Their expertise and assistance were invaluable. Finally, I express my very sincere gratitude to the administrators and special education teachers in Petersburg City and Chesterfield County Schools who helped me in completing my study. Without their help, this dissertation would not be a reality.

Table of Contents

Dedication	iii
Acknowledgements	iv
List of Tables	viii
List of Figures	ix
Chapter 1	2
Introduction	2
Justification for Study	4
Statement of the Problem	9
Limitations of the Study	11
Theoretical Rationale	13
Definition of Terms	21
Research Hypotheses	23
Ethical Considerations	29
Chapter II	37
Review of Research	37
Definitions of Teaching	40
Criteria of Teaching Effectiveness	44
Measuring Teaching Effectiveness	48
Conclusions and Implications	58
Chapter III	62
Methodology	62

Population and Sample	62
Procedures	64
Intervention	66
Ethical Considerations	68
Instrumentation	69
Sample and Data Gathering Procedures	77
Research Design	79
Specific Directional Hypotheses	79
Statistical Analysis Techniques	80
Summary	81
Chapter IV	83
Overview of Methodology	83
Findings	87
Hypothesis 1	89
Hypothesis 2	90
Hypothesis 3	99
Chapter V	104
Summary	104
Conclusions	107
Implications	114
Limitations of the Study	119
Recommendations for Future Research	120

References	122
Appendices	134
Appendix A	135
Appendix B	136
Appendix C	137
Appendix D	138
Abstract	139

List of Tables

Differences Between Pre and Post-tests of Reading Achievement	91
Ranges and Means of Frequency of Teaching Behaviors of Subjects	93
Correlations Between Teacher Behaviors and Reading Achievement of Handicapped Students	95
T-test for Differences in Reading Achievement with Independent Variables of Teacher Behavior During Period and Teacher Response	101
T-test for Difference in Reading Achievement for All Independent Variables	103

List of Figures

Reading Progress Record	86
Classroom Process Record	88

**A Study of Reading Achievement Gains in Classes
of Special Education Teachers
Using the Beginning Teacher Assistance Program
Indicators of Competence**

Chapter 1

Introduction

Considerable research has examined what it is that effective teachers do. Specific teacher behaviors which support student achievement have been identified and validated through extensive studies correlating teacher behavior with measures of student achievement (Medley, 1977). Knowledge of what is effective teaching is substantially greater today than it was a decade ago (Brophy & Good, 1985). From teacher effectiveness research certain teacher behaviors have been extracted which are regarded by some experts as generic competencies shared by effective teachers of all students.

With documentation of specific teaching competencies have emerged recommendations for how best to evaluate and document competence among preservice and inservice teachers. These recommendations are based largely on teacher effectiveness research, which suggests that teachers do have differential effects on student learning (Veldman & Brophy, 1974) and that effective teachers can be identified and their classroom behavior investigated (Brophy, 1973).

Many current recommendations for valid evaluation of competence are a direct outcome of teacher effectiveness research (Medley, Coker, & Soar, 1984; Darling-Hammond, Wise, & Pease, 1983). Advocated are classroom or process observation systems for teacher evaluation similar to those used by the teacher effectiveness researchers. Indeed, it was suggested over a decade ago that "ultimately, such research should yield...data that would provide a rational and valid basis for constructing and using process observations as teacher accountability criteria" (Brophy, 1973, p. 251).

Public demands for teacher accountability and a body of research suggesting that classroom observations of teacher behavior are a valid indicator of competence have led a number of states, among them Virginia, to mandate classroom observation as part of the requirements for teacher certification. The classroom observation is designed to assess the teacher's professional knowledge on the job. From the observation, a determination is made about the professional competence of the teacher (Beginning Teacher Assistance Program, Phase II Final Report, 1984). Professional competence must be demonstrated by

the teacher in order to qualify for certification (Morsink, Dykes, Algozzine, & Fardig, 1985; Beginning Teacher Assistance Program, Phase II Final Report, 1984).

In Virginia, the State Board of Education has specified fourteen areas in which every teacher who is granted a Collegiate Professional Certificate must be competent and has established the Virginia Beginning Teacher Assistance Program (BTAP) for the purpose of assessing the competence of beginning teachers. The basis of BTAP is a "set of measurable or observable indicators through which beginning teachers can demonstrate their competence in each competency area specified by the Board of Education" (Beginning Teacher Assistance Program, Phase II Final Report, 1984, p. 43). All indicators were generated from a review of teacher effectiveness research.

Justification for Study

Recently, the Florida Department of Education funded the "Personnel Competencies Research Project," under the direction of the Department of Special Education, University of Florida, Gainesville. The overall goal of the project was to identify and document competencies of exceptional education teachers. Of major concern

was "the appropriateness of the Florida Performance Measurement System (FPMS) for evaluation of Exceptional Student Education (ESE) personnel" (Morsink et al., 1985, not paginated). The Handbook of the Florida Performance Measurement System was one of five major sources of information for a "meta" review of the literature used to construct the assessment component of BTAP (Beginning Teacher Assistance Program, Phase II Final Report, 1984). Concerns about the appropriateness of FPMS for evaluating exceptional education personnel in Florida raise concerns about the appropriateness of BTAP for evaluating special education teachers in Virginia.

It is the purpose of this study to examine the appropriateness of one component of BTAP, the Classroom Process Observation, for evaluating certain special education teachers in categorical programs for mildly handicapped students in Virginia. In addition to examining BTAP, this study may serve to validate preliminary findings of the Florida Personnel Competencies Research. Those findings suggest that effective teachers of mildly and moderately handicapped students demonstrate many of the same behaviors which are indicators of effective teaching in regular

education, as well as some behaviors which differ significantly.

Background

Prominent researchers (Algozzine, Morsink, & Algozzine, 1986; Morsink, Soar, Soar, & Thomas, 1986; Ysseldyke, Thurlow, Mecklenburg, & Graden, 1984; Skiba, Sevcik, Wesson, King, & Deno, 1983) have cautioned against generalizing the findings of teacher effectiveness research in regular education to special education without the empirical validation of process-product studies. Yet, very few process-product studies, that is studies which correlate teacher behavior with measures of student outcomes, have been done in special education. This study, which is a process-product study, can serve to validate the indicators being used in Virginia to assess the competence of special education teachers. It can add to current knowledge of effective teaching in special education classrooms by validating certain competencies. It may also add to knowledge about research methodology in process-product studies in special education.

One of the greatest obstacles to validating competencies of special education teachers has been the documented inadequacy of most commonly

used tests in assessing outcomes with handicapped children. "Questions about the appropriateness of extant instruments have been raised in regard to diagnosis and assessment for clinical purposes, as well as about research and evaluation efforts where achievement, self-concept, sociometric methods, and even IQ serve as dependent variables or program outcomes to be tapped" (MacMillan, Keogh, & Jones, 1986, p. 693). The competencies measured in both BTAP and FPMS were derived from research correlating teacher behavior with scores on standardized tests of achievement, primarily in reading and mathematics. Standardized tests of achievement are generally designed to provide information about the achievement of children in the middle range. The validity of such tests for children outside this range is questionable, "since test scores become unreliable at the extreme ends of the score distribution" (MacMillan et al., 1986, p. 694). In the norming of most standardized tests, handicapped children were not included in the standardization sample. Thus, attempts to determine the impact of teacher behavior on the achievement of handicapped children using standardized tests as the outcome measure are fraught with difficulty. Without

adequate norms, it is impossible to determine whether growth is fast, slow, or at the expected rate (MacMillan et al., 1986). Therefore, comparisons among pupils and groups of pupils cannot be made with any degree of reliability. Research purporting to validate competencies of special education teachers using standardized tests as the outcome measure may certainly be challenged.

It has been argued by special educators that norm-referenced, standardized achievement tests do not effectively measure learning of handicapped students (Marston, Deno, & Tindal, 1983). In response to concerns about the use of standardized tests with handicapped children, investigators in special education have developed "direct measurement techniques" which correlate highly with standardized achievement tests, yet are sensitive to short-term gains in student learning (Tindal, Marston, & Deno, 1983; Marston et al., 1983). Direct measures of student achievement were used in one process-product study which investigated effective direct instruction practices in special education (Englert, 1984). Further investigation of direct, curriculum based measures of student achievement is of research

interest, however. In fact, the use of direct, curriculum based measures of student achievement in process-product studies of special education teacher effectiveness has been cited as "a challenging objective for future research" (Wolking, 1985, not paginated).

Virginia has established certain indicators of competence which it considers critical for all teachers. It appears that these indicators of competence can be effectively validated for special education teachers using direct, curriculum based measures of student achievement. This is the intent of this study.

Statement of the Problem

Is Virginia's Beginning Teacher Assistance Program valid for evaluating the competence of special education teachers? Is there a relationship between the BTAP indicators of competence and desired outcomes with handicapped children? To date the research which has been done suggests that the answers to these questions are far from definitive.

Much of the research on effective teaching has been done at the elementary school level, often in classrooms of teachers with low SES students (Brophy & Good, 1985). Some has been

done at the secondary school level (Brophy & Good, 1985; Evertson, Anderson, Anderson, & Brophy, 1980; Stallings, Needels, & Stayrook, 1979). Very little research, however, has examined effective teaching in special education. One of the problems with this lack of research is that it calls into question the validity of systems for evaluating the competence of special education teachers like Virginia's Beginning Teacher Assistance Program, which are based on the teacher effectiveness research.

It should be noted that the research which has been done appears to support the effectiveness of certain patterns of teaching in special education similar to those identified in regular education, particularly for mildly handicapped students (Englert, 1984; Morsink et al., 1985). Findings, however, must be considered preliminary. To date, studies of teacher effectiveness in special education have been largely descriptive (Algozzine et al., 1986; Morsink et al., 1985). Some few (Frick, Polsgrove, & Reith, 1986; Englert, 1984) have examined the efficacy of direct instruction, a pattern of instruction synthesized from the research on teaching (Rosenshine, 1986). None has sought to correlate

indicators of competence like those in BTAP with measures of student achievement. It clearly seems necessary to document the appropriateness of the BTAP indicators of competence for assessing special education teachers if questions about validity are to be definitively answered.

Limitations of the Study

Generally, studies of teacher effectiveness in regular and special education have been done using the methods and tools of observational research. Observational research overcomes many of the limitations of survey research and yields more accurate quantitative data than that obtained by self-report (Borg & Gall, 1983). Because of these advantages, it is the method for the current study.

While observational methods overcome certain limitations of survey research, these methods have limitations of their own. One of the limitations of observational research is that the presence of the observer often changes the behavior of those being observed (Borg & Gall, 1983). While this limitation can be potentially overcome by having the observer visit a classroom a number of times before recording any observational data, access to classrooms and time constraints preclude this

researcher from making nonobservational visits to classrooms. Other serious limitations of this study are small sample size and the researcher as sole observer. This researcher recognizes that the larger the sample of teachers observed, the more reliable the data obtained. She also recognizes that at least two independent observers are required to determine the reliability of data that are collected. She must, however, rely entirely on her own resources for obtaining observational data, necessitating a small sample size and the researcher as sole observer. It must be noted that these factors have the potential to reduce the reliability and validity of any findings from the study.

While there are limitations to this observational study, certain precautions have been taken to minimize these. First, the researcher was trained to a high degree of reliability and objectivity as a BTAP observer. Second, subjects were entirely unaware of the teaching behaviors that were being observed and recorded. Third, a standard observational schedule developed and validated by the Beginning Teacher Assistance Program was used to collect observational data.

observational data. Finally, the researcher did not previously know or work with any subject in the study.

Theoretical Rationale

Teacher certification is a state function. By issuing a license to an individual to teach in Virginia, the state is certifying that that individual is competent to teach. The state's primary concern in certifying teachers is whether candidates for certification possess the competencies necessary to safely practice their profession (Beginning Teacher Assistance Program, Phase II Final Report, 1984).

Measuring the Competence of Teachers

The focus of most state efforts aimed at evaluating the competence of teachers is on some form of paper-pencil test. Such tests seek to measure the competence of teachers by testing knowledge of subject-matter or professional knowledge and/or by identifying attitudes or personality characteristics which purportedly predict teaching success. The National Teacher Examinations is such a test which is fairly widely used by states as one basis for teacher certification (Medley et al., 1984).

In addition to paper-pencil tests, at least three states, Georgia, Florida, and Virginia, have mandated the use of low-inference measures of teacher performance in the classroom as part of the criteria for teacher certification. In Virginia, the Beginning Teacher Assistance Program was developed with the purpose of ensuring "that every teacher who receives the Collegiate Professional Certificate has demonstrated the possession of selected competencies" (Beginning Teacher Assistance Program, Phase II Final Report, 1984, p. 5). One of the requirements of BTAP is that every beginning teacher in the state of Virginia demonstrate functional knowledge of fourteen generic indicators of competence in actual performance in the classroom (Beginning Teacher Assistance Program, Phase II Final Report, 1984). A beginning teacher's competence is assessed by an observer using a structured observation scale developed from research on teacher effectiveness. On the observation scale, specific teacher behaviors or categories of behavior are defined. The observer looks for and records a behavior when it is demonstrated by the teacher. The record of teaching behaviors is scored using a scoring key which is applied after

the record is completed. Judgments about what is effective teaching are made beforehand and incorporated into the scoring key.

Advocates of low-inference evaluation of teachers argue that it overcomes inadequacies of other currently used methods, including paper-pencil tests, achievement test scores of students, and ratings of teacher performance in the classroom. Research suggests that paper-pencil tests of competence may measure basic literacy or subject matter knowledge, but there is no evidence to suggest that scores on such tests predict teaching success (Medley et al., 1984). Systems in which teacher performance is judged on the basis of student achievement gains at year's end fail to recognize that the raw materials a teacher works with vary widely; some students simply know more than others, some have greater ability, some are more motivated. Although statistical procedures can adjust for student differences, "class effects" are observed "despite statistical controls, even in highly consistent teachers" (Brophy, 1973, p. 251). Such statistical limitations argue against the use of student gain scores for validly evaluating teachers. Rating scales, which are the most often

used method of evaluating teachers, depend almost entirely on the opinions about effective teaching which the rater holds. Research suggests the experts who devise the rating scales as well as those who do the ratings are often misinformed or ignorant of effective teaching behaviors.

In the late 1970's, Coker, Medley, and Soar (1980) conducted research in one hundred classrooms in a school system in Georgia to determine the validity of expert opinions regarding effective teaching behavior. The research design involved the following: (a) development by teachers in the system of a list of teacher competencies; (b) selection by research staff of a set of appropriate measurement instruments to record behaviors relevant to these competencies; (c) collection of classroom behavior and pupil gain data by teachers recruited from the school system and trained to use the instruments; (d) reduction of these data by the research staff to measures of the listed competencies; and (e) an analysis of the relationships between the competence measures and the measures of pupil growth. By definition, each of the competencies in the list developed by the teachers in the Georgia school system should have been positively

related to student gains. Five of 13 significant relationships were, in fact, negative, calling into question the use of expert opinion as a basis for evaluating the competence of teachers.

Low-Inference Evaluation

There is a body of research which suggests that low-inference evaluation of teacher competence may be more valid and reliable than other methods. Richard Manatt and Shirley Stow point out in the introduction to their Clinical Manual for Teacher Performance Evaluation (1984) that scholars such as Rosenshine; Dunkin and Biddle; Good, Biddle and Brophy; Medley; and Peterson and Walberg have thoroughly reviewed the teacher effectiveness research and have legitimized studies linking certain teacher behaviors to measures of student achievement in basic reading and mathematics. Some researchers suggest, and the state of Virginia subscribes to the belief, that it is possible to determine specific teacher evaluation criteria on the basis of this process-product research and to organize these into a sound model for evaluating teachers (Beach and Reinhartz, 1984).

Validation of Indicators of Competence

For low-inference evaluation to be valid, teacher behaviors which are thought to be indicators of competence must be empirically tested to verify relationships to desired student outcomes (Medley et al., 1984; Soar, Medley, & Coker, 1983). Most process-product studies have correlated teacher behavior with measures of student achievement on standardized tests. The use of standardized test scores as the measure of student learning has been criticized. Typically, standardized tests of achievement are designed to measure acquisition of functional academic skills. They may not measure acquisition of higher-level thinking and problem-solving skills, for example. They also may not be sensitive measures even of student gains in functional academic skills. Standardized tests are designed to measure differences among individuals. The most efficient norm-referenced test is one which maximizes population variance. As a result, norm-referenced tests may not be sensitive measures of pupil progress (Tindal, German, Marston & Deno, 1983).

Curriculum Based Measurement

In process-product studies of teacher effectiveness in special education, the lack of

sensitivity of standardized tests to student improvement is a threat to validity. Conclusions about what is effective teaching behavior in special education are suspect if the criterion measure is not a sensitive measure of the dependent variable, just as conclusions about effective teaching in regular education may be suspect if the desired outcome is something other than functional academic skill.

Some special educators have advocated the use of direct measures of student achievement using curriculum based assessment procedures as a more reliable and valid measure of student outcomes than standardized tests. Researchers at the Institute for Research on Learning Disabilities, University of Minnesota have conducted extensive studies of the reliability and validity of direct measurement techniques for measuring pupil progress (Shinn & Marston, 1985; Marston et al. 1983; Tindal et al., 1983; Marston & Deno, 1982). In a series of studies, these researchers have shown that "a student's oral reading rate on a passage from his or her basal reader or a list of words from the reader correlated highly with standardized achievement tests of decoding ($r=.90$) and reading comprehension ($r=.80$)" (Marston et

al., 1983, p. 3). Further, in studies of the reliability of these measures of pupil progress, high coefficients for test-retest reliability, parallel-form reliability, and interjudge reliability have been obtained for correct performance scores (Tindal, Marston & Deno, 1983).

Curriculum based measures of pupil progress have been shown to be technically adequate. They have also been shown to be more sensitive to pupil progress than standardized measures. In a study of the effectiveness of direct measurement techniques and standardized achievement tests for measuring within-individual change over a ten-week period, Marston, Deno, & Tindal (1983) found that greater student gains were evident on the direct measures. Significant differences were obtained for 16 of 20 comparisons of direct measures and standardized test measures ($p=.001$).

Validation of BTAP for Special Education

There is a strong rationale for investigating the appropriateness of BTAP competencies for assessing the competence of teachers of special education. A major conclusion drawn from the research is that there simply is no one set of effective teaching behaviors (Medley, 1977). Effective teaching has been shown to vary by SES,

IQ, age, and anxiety level of students (Brophy, 1979; Soar & Soar, 1972; Rosenshine, 1983; 1986). It is likely to vary by handicapped and nonhandicapped. Advocates of low-inference measures for evaluating teachers are emphatic that teaching behaviors must be subjected to empirical testing to verify their relationship to student outcomes (Soar et al., 1983). There exists a feasible method by which BTAP competencies can be empirically validated as appropriate for assessing the competence of special education teachers. The results could have significance for the Virginia Department of Education, for researchers in special education, and certainly for beginning special education teachers in the state of Virginia.

Definition of Terms

This study contains a number of specialized terms. These are defined below.

1. BTAP--acronym for the Beginning Teacher Assistance Program in Virginia (Beginning Teacher Assistance Program, Phase II Final Report, 1984).

2. competency--any single knowledge, skill, or professional value which is believed to be

relevant to the successful practice of teaching (Medley et al., 1984).

3. competence--the repertoire of competencies a teacher has (Medley et al., 1984).

4. curriculum based measurement (CBM)--a type of evaluation system, in which the curriculum is the source of items for testing (Skiba et al., 1983).

5. direct measurement--criterion-referenced measures which test the same skills as those that have been taught; they often use the same response mode as that employed initially in teaching the skills (Mirkin et al., 1982).

6. low-inference measurement--a structured system for observing teacher behavior in which operationally defined behaviors are coded by trained observers as they occur in the classroom (Evertson & Brophy, 1974).

7. mildly handicapped--a term used to refer to students who have been classified as educable mentally handicapped, learning disabled, or emotionally disturbed (Hallahan & Kauffman, 1977).

8. process-product studies--studies correlating specific teacher behavior with measures of student outcomes (Morsink et al., 1985).

9. teacher effectiveness research--body of research demonstrating a relationship between the behavior of teachers and learning outcomes of students (Brophy & Good, 1985).

Research Hypotheses

The purpose of this study is to validate BTAP competencies as indicators of competence of a group of special education teachers by analyzing relationships between measures of teachers' behavior and measures of student learning. It is hypothesized that:

1. There is a significant relationship between the behavior of special education teachers and learning outcomes of handicapped students.

2. Learning outcomes of handicapped students can be sensitively measured by curriculum based measurement.

3. BTAP indicators of competence are related to curriculum based measures of reading achievement of mildly handicapped students.

Linking Teacher Behavior and Student Achievement

There is a significant body of research in regular education demonstrating a relationship between the behavior of teachers and learning outcomes of students (Brophy & Good, 1985). This research is known as process-product research,

teacher effectiveness research, and more recently teacher effects research. Process-product research has established relationships between teaching behaviors such as: (a) instructing in groups, (b) reviewing and discussing assignments, (c) praising student successes, and (d) providing support and feedback, and achievement in basic reading and mathematics among secondary students (Stallings et al., 1979; Evertson et al., 1980). Teacher behaviors, including time spent on academic tasks; teacher direction; teacher presentation of information; drill; teacher supervision of activities; direct and narrowly focused academic questions; control over the classroom and immediate reinforcement of student responses, have been correlated with reading and mathematics achievement of elementary students (Soar & Soar, 1972; Stallings, 1974; Medley, 1977; Anderson, Evertson & Brophy, 1979). From the research, there has emerged a consistent pattern of teaching which is effective for teaching a body of content or well-defined skills (Rosenshine, 1986). This pattern, referred to as direct instruction, has been divided into six teaching functions which are: "review, presentation of new material, guided practice, feedback and

corrections, independent practice, and weekly and monthly reviews" (Rosenshine, 1986, p. 64). In a number of studies, direct instruction has been correlated with measures of student achievement in language arts and mathematics.

Though there are many fewer process-product studies in special education than in regular education, there is evidence of a relationship between teacher behavior and student learning in special education. Larrivee and Vacca (1982), in a special project report on the development of teacher competencies necessary for successful mainstreaming of mildly handicapped students, identify a profile of teaching behaviors effective with special needs students in the regular classroom. During the project, data on over seventy teaching variables, previously shown in process-product research to relate to student outcomes, were collected. Data in the areas of questioning style, classroom climate, individualization, classroom management, academic learning time, teaching style, and opinion and attitudinal variables were collected on 33 regular classroom teachers in grades one through six. Pre-post data were obtained for all students in the classrooms on general aptitude, academic

achievement in reading, language, and mathematics, classroom behavior, peer acceptance, school attitude, and self-concept. Based on the student data, a profile was established from which a target group of special needs students with the most discrepant profiles was selected for a given classroom. A pool of effective teachers was selected from the sample of 33 teachers by considering gains made by targeted students as well as those made by the class. These effective teachers were extensively observed, and characteristic teaching behaviors were isolated. Forty-two of the seventy teaching behaviors observed were found to be characteristic of teachers identified as effective with special needs students.

In a similar process-product study of special education teacher interns, Englert identified effective and less effective teachers of mildly handicapped students, "as differentiated by their direct instruction practices" (1984, p. 38). Twenty-eight teacher interns and fifty-two special education students participated in the study. Low-inference measures of teacher behavior on specific teaching behaviors associated with direct instruction were collected through observations of

interns. Interns were trained to keep records of student progress. Records included documentation of when skills were introduced to students and the date they were mastered, as well as graphs of each student's daily scores on criterion-referenced tests administered by teacher interns.

"Typically, criterion-referenced tests involved reading, naming, or, in the case of math, writing answers on a math probe" (Englert, 1984, p. 42). Effective teacher interns were identified by calculating a "learning score" for each pupil and a median learning score for all pupils taught by an intern. Learning scores were derived by dividing the larger of two pupil scores, pre-test and post-test, by the smaller. This quotient was the proportional growth over several weeks of instruction and was transformed to a weekly score by dividing the quotient by the number of weeks devoted to a skill. Teacher interns were ranked from high to low on effectiveness based on median learning scores of pupils. Teachers in the top half of the ranking were designated more effective, and those in the lower half were designated less effective. The teaching behaviors of the two groups were then compared using analysis of variance. Findings suggested that the

more effective group maintained a brisker pace through the lesson, elicited more correct responses from students per minute, stated the objective, presented many examples, provided error drill, and implemented precuing to maintain high levels of successful student practice (Englert, 1984, p. 46).

Curriculum Based Measurement of Student Outcomes

Englert's study documents the use of direct measures to measure student outcomes in special education. Tindal et al. (1983) used similar direct measures to analyze the effectiveness of special education placement for 96 students in grades one to six. Students were assessed three times in one school year using brief samples of student performance in reading, spelling, and mathematics, with the curriculum as the source of items for testing. A random sample of 20 regular education students from each grade level, one to six, was similarly tested. Two measures of performance were calculated from the data, an absolute score for each academic area and a discrepancy index which was derived for each grade level by dividing the lower score, from the special education students, into the higher score, from the regular education students. In general,

comparisons of the discrepancy index across time showed that the performance of the special education students increased, thus, the authors concluded, demonstrating the effectiveness of special education placement.

Summary

There appears to be evidence to support the hypotheses that (a) there is a relationship between teacher behavior and student learning in special education, (b) student learning in special education can be reliably and validly measured using direct measurement, and (c) there is a significant and positive relationship between teaching behaviors derived from process-product research and direct measures of student learning in special education. There exists, however, the need for further validation.

Ethical Considerations

According to researchers (Morsink et al., 1985; Medley et al., 1984, Coker, et al., 1980), competency based teacher evaluation is a powerful concept. In a competency based program, successful teaching behaviors are operationally defined, and the teacher can be held accountable, can in fact be required to demonstrate in the classroom a specified set of competencies.

Successful implementation of competency based evaluation, however, requires the development of an operational definition of competencies and some validation demonstrating that each competence increases the teacher's effectiveness in the classroom. Evidence must be presented to show that teachers who possess the competencies are more effective in helping pupils learn than teachers who do not. It is the purpose of this study to investigate the relationship between certain indicators of competence used to evaluate teachers in the Beginning Teacher Assistance Program in Virginia and reading achievement of handicapped children in special education classes. A further purpose is to document the validity of using curriculum based measurement to measure outcomes in a process-product study of special education teacher effectiveness. The study raises certain ethical concerns which need to be addressed.

The literature does not specifically address ethics and research related to teacher evaluation. There is, however, a body of literature concerned with policy and program evaluation. It seems reasonable to review this literature as a preliminary step in developing norms by which the

professional behavior of this researcher may be guided, since this study seeks to evaluate the validity of a portion of the BTAP program.

Ethics and Policy and Program Evaluation

Policy and program evaluators may be external to the organization (Nagel, 1983), or they may be internal evaluators who are employed by the organization whose programs they evaluate (Adams, 1983). The ethical dilemmas which the external policy evaluator faces are in many respects similar to those faced by the internal program evaluator, though there are some differences. For example, the external policy evaluator may face the dilemma of whether to focus his evaluation on all consequences of a particular policy or to focus on intended consequences only (Nagel, 1983). Typically, there is pressure for the evaluator to focus only on intended consequences of a policy, since these are the consequences in which an agency is interested. A similar dilemma faced by the internal evaluator is organizational pressure to downplay negative and emphasize positive findings (Adams, 1983). In his discussion of whether the external evaluator should focus on all consequences versus intended consequences only, Stuart Nagel states that "the

failure to foresee important consequences may sometimes constitute a form of negligence that would amount to evaluation malpractice" (1983, p. 5). Likewise, the internal evaluator who downplays negative findings and emphasizes positive findings may be engaging in evaluation malpractice. The professional evaluator, and in this case, researcher, has a duty to provide objective and frank information. An evaluation which focuses on some consequences and not others may not be objective. An evaluation which emphasizes positive findings and downplays negatives is certainly not frank.

This commitment to be objective and report findings fully and frankly is emphasized in the "Proprietary Standards" for evaluation developed by the Joint Committee on Standards for Educational Evaluation, published in 1981 (Adams, 1983, p. 2). Nagel's statement regarding evaluation malpractice is germane to this study and suggests that findings must be reported objectively, fully, and frankly.

Dilemmas related to the objectivity of evaluations and to reporting of findings are not the only ones encountered by those doing evaluative studies. Other ethical dilemmas faced

by either external or internal evaluators and researchers include (a) the efficiency-equity dilemma where an organization is more concerned with benefits/costs than with equal treatment across individuals and/or groups; (b) the reporting of questionable findings dilemma in which findings that may lack validity are not sufficiently scrutinized; (c) the partisan dilemma in which the evaluator is called upon to aid in prescribing a policy which will benefit a certain group, and (d) the reinforcement for nonthreatening evaluation activities dilemma where the organization is more interested in making the evaluation unit visible than in using results of an evaluation. At root, each of these dilemmas arises from subtle or not so subtle pressure for compromise in reporting findings. Yet, the purpose of an evaluative study is to get at and fully report the truth in the interest of the public good rather than in the interest of a particular group, agency, or organization.

Ethical Responsibilities of the Researcher

Understanding the delimmas faced by those who seek to do evaluative studies of programs, this researcher accepts certain responsibilities, or ethical norms, in seeking to carry out this study.

These are: (a) to know, in so far as research has been able to establish, what competencies contribute to effective teaching, (b) to evaluate validly certain BTAP competencies as indicators of competence of special education teachers, in so far as it is possible to do so, and (c) to report findings fully, frankly, and sensitively. In the context of this study, researcher is one who has designed and seeks to carry out a study for empirically validating competencies which are the basis for evaluating special education teachers seeking certification in the state of Virginia. Competencies mean teaching behaviors which relate to student learning. Valid evaluation means that the researcher seeks to measure that which she says she is measuring, and further, that validity of instrumentation is fully documented. Fully documented validity of instrumentation is accepted as a major responsibility, given that an ancillary purpose of the research is to document the validity of curriculum based measurement in an empirical study.

These norms require that this researcher read the research, attend conferences, participate in inservice training, and engage in other learning activities which may potentially increase her

knowledge about effective teaching and valid research. Further, the researcher accepts the responsibility to use her knowledge to make the study a valid one. In addition to increasing her own knowledge, this researcher accepts the obligation to support efforts to increase knowledge in the field of teacher evaluation. She advocates open discussion of problems related to validity and seeks to increase the validity of her study. Finally, this researcher seeks to fully and frankly disclose findings, while remaining particularly sensitive to the need for confidentiality. No names or other identifying information related to teachers or students who are the subjects of study are revealed in reporting findings. The only exception to the responsibility to fully and frankly disclose findings is where such disclosure is likely to be detrimental to teachers of handicapped students and/or to handicapped students.

Basic moral obligations which many accept as prima facie duties are obligations to be honest and truthful. In accepting the responsibility for validating competencies which contribute to effective teaching in special education, this researcher in effect has agreed to try to discover

what is fact and what is myth regarding effective teaching. She has agreed to look for the truth in so far as it is possible to determine truth, and to base her conclusions on that truth. She further has accepted the obligation to be truthful when she seeks to ensure that her study is valid. These commitments are, in effect, commitments to be honest and truthful.

Chapter II

Review of Research

Can teacher educators, school administrators, and others responsible for formal evaluation of teachers validly assess teaching? Do teacher evaluators know what characteristics, knowledge, behaviors, and attitudes are indicative of "good" teaching; and, can they measure these and make valid judgments about who is an effective teacher and who is not? Are tests which measure a teacher's knowledge of subject matter or methods of teaching valid? What can be said of the validity of classroom observations or measures of pupil achievement for assessing teacher effectiveness? These are not questions of mere academic interest. Evaluation by definition involves assessment of the worth of something. At root, teacher evaluation requires that the evaluator assess the worth of individuals as teaching professionals, although such decisions may have a variety of purposes, many of which are not related to individual accountability. For example, substantive policy decisions related to pre-service and in-service training, accountability to the public, instructional improvement, and school status (vis-a-vis

certification or accreditation) are made on the basis of information gathered through the evaluation of teachers (Darling-Hammond et al., 1983). Whether the information gathered through teacher evaluation is used directly to judge an individual's worth as a teacher or to make administrative decisions, administrators and teachers have both a right and a responsibility to demand that the process yield valid information. Indeed, in the introduction to their recent book outlining a systematic, measurement-based approach to teacher evaluation, researchers Donald Medley, Homer Coker, and Robert Soar express the view that teacher evaluation today is in a "chaotic state" and consists of little more than "obtaining someone's subjective judgment of how 'good' a teacher is" (1984, p. 4). If there is an accurate assessment of the state of the art of teacher evaluation, questions about the validity of current approaches are well founded, and systematic review can only be viewed as helpful and healthy. Questions about the validity of the teacher evaluation process should be raised and concerns addressed if necessary. It is, therefore, the purpose of this chapter to review the research related to validity of current

approaches to teacher evaluation and to ask under what conditions and in what contexts they may be valid.

Judgments about the validity of any approach to teacher evaluation cannot be made without first understanding what it is that teaching is supposed to encompass. Robert Travers has aptly pointed out that "as tasks prescribed for the teacher vary, so too do the criteria for evaluating effectiveness" (1981, p. 14). In other words, one's definition of teaching determines to some extent the dimensions which are the basis for judging effectiveness. Dimensions are then operationally defined in terms of specific criteria which in turn dictate what data are gathered and how. Validity may be compromised if the basic construct, teaching, is ill defined, if the criteria for evaluation are not accurate indicators of the dimensions of teaching as it is defined, or if procedures or instruments for gathering data are themselves unreliable and consequently invalid. Thus it seems reasonable to:

1. Examine definitions of teaching which underlie approaches to teacher evaluation;

2. Examine the criteria which are used to judge teacher effectiveness;

3. Examine general approaches to gathering data related to teacher effectiveness criteria. Such an approach will provide some answers to basic questions which have been posed concerning the validity of teacher evaluation processes. More to the point, however, specific problems may be more readily identified and discussed, and perhaps some implications for research or current practice may emerge.

Definitions of Teaching

Over the course of history definitions of teaching have varied. In Ancient Greece, the teacher was one who gathered about him any who wished to engage in discussion or disputation. The method was the dialectic, and the pupils were generally adults. The effective teacher was quite simply the one who could attract students (Travers, 1981). With the emergence of the grammar school in the late Middle Ages, the role ascribed to the teacher was that of manager. Teaching effectiveness was judged on the basis of ability to carry out such managerial functions as keeping order, hearing pupils recite, organizing work of pupils, and giving new assignments. The

teacher provided opportunities for pupils to learn, but all responsibility for learning rested with pupils. This definition of teaching as essentially managerial, with responsibility for learning the pupil's own, persisted until the present century when some fundamental assumptions about teaching changed dramatically.

In England in the late Victorian era, the British government introduced "payment-by-results," a system in which teachers were paid according to how high their students scored on end-of-the-term examinations. Implicit in the system was the assumption that teachers could influence and were responsible for student learning (Travers, 1981). In spite of a body of research which suggests that pupil learning is a function of many factors, of which competent teaching is only one, current conceptions of teaching work hold to the belief that the teacher is able to influence learning and is somehow at fault if students fail to learn.

In the past teachers were looked upon primarily as managers. More recently the work of teachers has been compared to that of craftsmen, laborers, bureaucrats, professionals, and artists. Darling-Hammond, Wise, and Pease suggest, in

general, four conceptions of teaching work. One of these four they see as underlying every teacher evaluation system, since every system must "embody a definition of the teaching task" (1983, p. 291). One conception of teaching which may underlie a system of evaluation is that of teaching as labor. When teaching is thought of as labor, teaching activities are in the form of a standard set of operating procedures defined by administrators. The job of the teacher is to adhere to the prescribed routines and procedures.

A perhaps more common conception of teaching underlying evaluation systems is that of teaching as craft. This definition of teaching assumes that teaching requires a repertoire of specialized techniques and knowledge of specialized rules for applying techniques. The teacher who properly uses the prescribed set of rules for applying specific techniques will perform satisfactorily. The conception of teaching as a profession requires that the teacher master a body of theoretical knowledge and a range of techniques and, further, that the teacher exercise judgment about when to apply specialized techniques. The teacher as a professional is expected to "follow what research findings or professional judgment

suggest to be the best practice in a given circumstance" (Soar et al., 1983, p. 240). Finally, teaching as art implies that teaching techniques and their application "may be novel, unconventional, or unpredictable" (Darling-Hammond et al. 1983, p. 292). Techniques and best practice are not ignored, but are thought to be uniquely influenced by the personality of the teacher as well as by the interactions of teacher with students (Darling-Hammond et al., 1983).

Influence of Definitions on Evaluation

There is no general consensus among either researchers or practitioners as to how best to conceptualize the work of teaching for purposes of evaluation. There can be no doubt, however, that definitions of teaching influence teacher evaluation systems. For example, when teaching is viewed as labor, evaluation typically involves inspection of lesson plans, monitoring of classroom performance, and evaluation of pupil outcomes. On the other hand, when teaching is thought of as a profession, evaluation is based on the degree to which the teacher can professionally solve problems given certain standards of professional knowledge and practice (Darling-Hammond et al., 1983). Such definitions

of teaching and criteria for evaluation may or may not be consistent with what research suggests is effective teaching. Changes in assumptions about what is teaching without changes in what is known about teaching hardly increase the validity of teacher evaluation (Travers, 1981). A good definition helps determine the dimensions upon which evaluation will focus, but it is research which helps to operationally define criteria within certain dimensions.

Criteria of Teaching Effectiveness

There are three sets of variables which researchers have examined in their efforts to determine what constitutes teacher effectiveness (Medley et al., 1984). These variables can be arranged along a continuum from direct to indirect:

Product Process Presage

For nearly half a century, researchers sought to link teacher characteristics, such as age, intelligence, experience, and scores on personality tests, with principals' ratings of performance (Gage, 1971). In an elaborate study, the "Teacher Characteristics Study," conducted over 6 years with 6000 teachers in 1,700 schools and 450 school systems, teachers' classroom

performance ratings were correlated with such variables as age, sex, extent of teaching experience, and marital status. Most correlations were nonsignificant. Further, the relation between these variables and student growth was a presumed one (King, 1981). In a classic review of the research published in 1960, Mitzel concluded that this study and hundreds like it on a smaller scale had failed to establish significant correlations between what he termed presage criteria and measures of teaching effectiveness. He noted, "Presage criteria, so called here because of their origin in guessed predictions are from a logical standpoint completely removed from the goals of education....In a sense they are pseudocriteria, for their relevance depends upon an assumed or conjectured relationship to other criteria, either process or product" (p. 1484). Other reviewers (Gage, 1971) have agreed with Mitzel's assessment of the research related to presage variables.

Process and Product Variables

More recently, researchers have sought to link process and product variables. They have attempted to determine specific relationships between what a teacher does in the classroom and

what pupils learn (Rosenshine, 1971). While the results of process-product research are by no means definitive, knowledge about effective teaching is substantially greater today than it was 20 years ago (Medley et al., 1984).

Researchers investigating correlations between teaching behavior and student achievement typically use a structured observation system to tally frequencies of certain teacher behaviors and then attempt to determine relationships between frequencies of teacher behavior and measures of adjusted pupil achievement (Rosenshine, 1971). Examination of some of the major studies (Soar & Soar, 1972; Brophy & Evertson, 1974; McDonald & Elias, 1976) reveals that hundreds of variables in classroom instruction have been investigated. Though many gaps in the teacher effectiveness research remain (Medley et al., 1984), certain relevant, important, reliable, and generalizable findings have been clearly documented.

In a classic monograph published in 1977, Donald Medley culled the most significant findings from 289 empirical studies of process-product relationships. He reported 613 significant correlations. In deciding whether a relationship should be reported or not, he used four criteria:

(a) the results had to be legitimately generalizable to teachers other than those in the sample studied, (b) a strong and reliable relationship equivalent to a linear correlation of .39, significant at the 5% level had to have been obtained between process and product measures, (c) product measures had to relate to the kinds of outcomes, generally gains in reading or arithmetic or changes in attitude toward school or in pupils' perceptions of the self, that teachers are hired to accomplish, and (d) measures of teacher behavior had to be specified clearly enough to be reproducible. Findings which met these criteria were reported in 40 tables designed to provide educators direct access to the results of process-product research. It is not within the scope of this paper to reiterate the research findings reported by Medley. The interested reader is referred to the original monograph. It is important to note, however, that Medley's review confirms the assumption that what the teacher does in the classroom does influence student learning. The question then becomes one of how to measure what it is teachers do in their classrooms. Scholars and researchers generally agree that current procedures have serious flaws,

though there is major controversy over how, indeed, evaluators are to measure teaching effectiveness (Coker et al., 1980; Haefele, 1980; Harris, 1981; Darling-Hammond et al., 1983; Soar et al., 1983).

Measuring Teaching Effectiveness

Current approaches to measuring teaching effectiveness can be categorized under three broad headings: (a) paper-pencil tests which measure abilities, knowledge, skills, and values, (b) supervisors' or administrators' ratings of performance based on classroom observations, and (c) standardized test scores of students in a teacher's class which measure mean gains in learning (Haefele, 1980; Soar, et al., 1983).

Paper-Pencil Tests

There is a growing trend towards the use of paper and pencil tests for teacher evaluation, at least as part of the initial certification process. This is due primarily to the public demand for accountability (Harris, 1981). Some states have developed their own teacher competency tests, among them Georgia, Florida, and South Carolina (Darling-Hammond et al., 1983). The National Teacher Examinations, however, which include the Common Examinations and the Teaching

Area Examinations are by far the most widely used teacher competency tests.

Validity of the NTE.

Quirk, Witten, and Weinberg (1973) have done an extensive review of studies of the concurrent and predictive validity of the NTE. Concurrent validity has been studied in relation to scores on the Graduate Record Examinations, grade point average of undergraduate and graduate students, and personal characteristics of candidates. Although Quirk et al. seriously question the statistical procedures of most of the studies of concurrent validity of the NTE, there is some evidence of positive correlations between scores on certain subtests of the NTE and GPA and GRE scores.

Predictive validity of the NTE has been studied most frequently in relation to administrators' ratings. Two studies were found which correlated NTE scores with pupil ratings, two with pupil residual gain scores, and one which correlated NTE scores with measures of teacher behavior gathered through structured classroom observation. Based on their review, Quirk, et al. conclude that NTE scores are poor predictors of a teacher's on-the-job ratings. They note, however,

that "on-the-job ratings are notoriously unreliable, and their reputation is well-deserved" (1973, p. 108). Because so few studies have been done which correlate NTE scores with measures other than supervisors' ratings, these reviewers conclude that more studies need to be done before conclusions about the predictive validity of the NTE can be drawn. They do caution against the use of fixed cutoff NTE scores as a criterion for certification, for considering raises in salary, for contract assignment, and the like, while noting that the validity of the NTE should be judged by the accuracy with which it measures what it is designed to measure and not "the total complex of teaching ability" (1973, p. 109).

Supervisors' and Administrators' Ratings

Quirk and his colleagues refer to the failure of supervisors' ratings to reliably measure teaching effectiveness. The limitations of rating systems, e.g. observer bias, poor instrumentation, susceptibility to halo effect, lack of interrater reliability, have been well documented in the literature (Haefele, 1980; Ellet, Capie & Johnson, 1980; Evertson & Holley 1981; Glass, 1981; Soar et al., 1983; Medley et al., 1984). Typically, ratings of a teacher, sometimes called high

inference measures, are done by the principal or a supervisor whose task is to infer a rating from whatever is observed in the classroom. The teacher behaviors which are rated are usually poorly defined, and often evidence of any relationship between the behaviors that are rated and student outcomes is lacking. Holley, in a review of formally documented teacher evaluation systems, found only 19% which used any kind of competency based ratings or any type of expanded behavior descriptors (1981). "This approach is shot through with validity and reliability problems" (Haefele, 1980, p. 350). Some researchers, however, have demonstrated that the reliability and validity of rating scales can be improved by carefully training observers, by incorporating into the scales characteristics of teaching which research has established can be reliably observed and which bear some significant relationship to desired pupil outcomes, and by including specific performance indicators for each characteristic included in the scale (Evertson & Brophy, 1974; Manatt, Palmer, & Hildebaugh, 1976; Ellett et al., 1980).

Improving the validity of rating scales.

In a report of the findings of the first year of a two-year study attempting to isolate correlates of effective teaching, Evertson and Brophy (1974) provide evidence for the reliability and validity of high-inference behavioral correlates of teaching effectiveness. In their study, high-inference measures of teacher process variables were taken on a sample of 31 teachers selected for their demonstrated consistency in producing gains in student learning. Two types of high-inference measures were used. One was a 5-point rating scale which included ratings of student attention, teacher enthusiasm, clarity, positive and negative affect, task orientation, cognitive level of questions, student passivity, pupil-pupil interaction, and percentage of time spent in lectures and demonstration. Teachers in the study were observed 4 times by trained coders, and scales were marked several times during each observation. The second measure included 41 high-inference ratings and 15 high-inference checklists and percentage estimates which coders filled out following their last two visits to each teacher's class. These high-inference measures were correlated with residual pupil gain scores

which had been correlated with low-inference behavioral codings of such variables as teacher vs. student initiation of contacts, types of interactions (academic, procedural, or behavioral-disciplinary), difficulty level of teacher questions, quality of student responses, quantity and quality of teacher feedback and evaluative reactions to student response and student work, and the teacher's method and general effectiveness in handling classroom management and disciplinary problems (Evertson & Brophy, 1974). Although the high-inference ratings showed evidence of halo effect and logical error, in general they supported the findings from the correlation of low-inference measures and residual gain scores. Although this study offers some support for the use of rating scales for teacher evaluation, it should be noted that the high-inference measures in the study were performance-based, observations were done by trained coders, and there were measures of interrater reliability.

The successful use of classroom observation in this and other teacher effectiveness studies leaves no doubt that observation systems can measure effective teaching. Two conditions, however, are essential for classroom observations

to be valid: (a) observers must be trained to use a systematic approach and (b) a reliable and valid instrument must be used to gather data during an observation. Of utmost importance is the use of an instrument that has an acceptable degree of reliability and validity (Evertson & Holley, 1981, p. 101). These conditions are not met in observation systems currently used by most school systems. In most instances, school systems have one instrument which is used systemwide. The observer watching the performance of a kindergarten teacher is watching for the same behaviors as the observer watching the performance of a high school physics teacher (Evertson & Holley, 1981, p. 97). Research suggests that teaching behavior positively correlated with effective teaching in one setting may be negatively correlated with effective teaching in another (Medley, 1977). There simply is not one set of effective teaching behaviors. A rating system which presumes that all effective teachers engage in the same behaviors regardless of the context is not going to provide valid measures of effective teaching. A further threat to validity is the fact that a rating is usually assigned at the end of the observation. This requires a high

degree of interpretation on the part of the observer with the result that reliability is often sacrificed.

A number of researchers and experts are advocating the use of classroom observation systems for teacher evaluation similar to those used successfully by the teacher effectiveness researchers. They point to the fact that category, sign, and multiple coding systems, which allow for recording of particular teaching behaviors, have been validated through research (Rosenshine, 1970; Flanders, 1970; Medley et al., 1984). The instruments, they suggest, can be used by teacher evaluators to gather reliable and valid information about the performance of the teacher in the classroom. The assumption is that inferences about teacher effectiveness can be made from valid information about a teacher's classroom performance. This assumption, however, is rejected by those who argue that information about the performance of the teacher in the classroom, no matter how reliably and validly measured, cannot be a measure of teacher effectiveness. The only real measure of a teacher's effectiveness, they suggest, is student achievement data.

Student Achievement Data

Since student learning is the true test of teacher effectiveness, a number of scholars and researchers suggest that direct measures of student achievement are apt to be more valid than proxies (Millman, 1981). Theoretically, standardized test scores of students should be the most valid measure of teacher effectiveness. Arguments against the use of student achievement data for evaluating teachers, however, are based on the fact that factors over which the teacher has no control affect student learning. "The best teacher in the world would not fare very well if faced with slow learners, unmotivated students, a poor learning environment, and an achievement measure out of harmony with the teacher's goals," argue the critics (Millman, 1981, p. 157). On the other hand, it would appear that differences among classrooms and differences in student characteristics can be controlled for statistically (Veldman & Brophy, 1974). In a study in a large Southwestern school system, Veldman and Brophy were concerned with "methodological considerations involved in obtaining unbiased estimates of teacher influence on pupil achievement..." (1974, p. 320). Two

hundred seventy-five second and third grade teachers were considered for inclusion in the study. Selected were those who (a) had at least five years of experience at their grade level, (b) had taught the same grade level during the three years for which data were gathered, and (c) had at least 14 children with available data for each of the three years. Pupil scores on the Metropolitan Achievement Test were obtained for each of four successive years from pupil records. A series of regression models were then compared using (a) pre-test, (b) squared pre-test, (c) pupil sex, (d) year of testing, and (e) teacher as predictor of post-test performance. The strongest predictors of post-test scores, by a considerable margin, were usually pupil pre-scores. Inclusion of the teacher variable also yielded a significant and often substantial increase in predictive efficiency. Veldman and Brophy concluded that their data showed that reasonably stable estimates of teacher influence could be obtained from standardized achievement measures of pupil performance. They noted, however, that it was necessary to eliminate new teachers and teachers who had recently switched grades from their sample. They also cautioned that the stability

obtained were not high enough to justify the use of residual gain scores on standardized tests for teacher accountability purposes (Veldman & Brophy, 1974, p. 323). Other studies by Bennett, Harris, and Brophy (Glass, 1974) which have examined the reliability of using standardized test scores to measure teachers' effectiveness have obtained few stability-reliability coefficients significant at the .05 level. Though reliability does not ensure validity, no measure can be considered valid if it is not highly reliable.

Conclusions and Implications

From the foregoing review, it seems safe to say that there is no approach to teacher evaluation which can be said to be truly valid. A systematic approach to classroom observation using a validated instrument to record specific behavior of teachers seems the most valid approach, relatively speaking. Such an approach does have serious limitations of which the evaluator should be cognizant, however. Low-inference measures are unlikely to measure the complex which is teaching and likely will have to be supplemented with other information. Further, it appears that low-inference measures suggest a definition of teaching as labor or craft. This seems a fair

teaching as labor or craft. This seems a fair statement in spite of the argument of researchers Medley, Coker, and Soar (1984) for a low-inference system of evaluation and a definition of teaching as a profession. Finally, it must be noted that such an approach does not measure effectiveness. Donald Medley (1984) has offered definitions of four terms which are often used synonymously. Distinguishing among these terms, teacher competency, teacher competence, teacher performance, and teacher effectiveness, is critical to an understanding of teacher evaluation and to an appreciation of the limitations of even the best classroom observation. Teacher competency refers to any single knowledge, skill, or professional value which is believed to be relevant to the successful practice of teaching; competencies refer to things teachers know, do, or believe but not to the effects of these things. Teacher competence refers to the repertoire of competencies a teacher has; overall competence is a measure of the degree to which a teacher has mastered a set of competencies, some of which are more important in teaching than others. Teacher performance refers to what the teacher does on the job rather than to how competent he or she is;

teacher performance depends on the competence of the teacher, the context in which he or she works, and ability to apply competencies in specific situations. Teacher effectiveness refers to the effect that the teacher's performance has on pupils; effectiveness depends not only on competence and performance but also on the responses pupils make. From a valid assessment of a teacher's performance in the classroom, the evaluator can make some judgment about the teacher's ability to apply certain knowledge and skills and perhaps to use his professional judgment. Inferences about competence can be made. Inferences about effectiveness cannot.

More research needs to be done to establish the correlates of effective teaching. Most research thus far has been done on low SES children at the elementary level, primarily because this has been where the funding has been. In spite of studies on effective teaching which number in the hundreds, little is known about effective teaching at the secondary level. Almost nothing is known about the correlates of effective teaching and the handicapped. What little is known about effective teaching and the handicapped has come from the applied behavior analysis

Until a great deal more research is done, those responsible for teacher evaluation should proceed with considerable caution in designing and implementing teacher evaluation systems.

Chapter III

Methodology

Population and Sample

The accessible population for this study included elementary special education teachers of mildly handicapped students from Petersburg City and Chesterfield County Schools in Virginia. All teachers in the population held the Collegiate Professional Certificate. All were certified to teach mildly handicapped students.

The sample from the accessible population was a volunteer one. All teachers of learning disabled, emotionally disturbed, and educable mentally retarded students in southern Chesterfield and Petersburg City were apprised of the study and asked to participate. Consistent with the assumptions of noted researchers in special education, it was assumed that there were no significant statistical differences in the characteristics of these children typically identified as mildly handicapped (MacMillan et al. 1986). Algozzine, Morsink, and Algozzine, referring to the work of Hallahan and Kauffman, Edgar and Hayden, and others, provide logical evidence for considering that children who fall into the "three categories of mildly handicapped--

learning disabled, emotionally disturbed, and educable mentally retarded..." represent "essentially the same population..." (1987, p. 4). It should be noted, however, that while there appear to be no statistically significant differences among the various categories of mildly handicapped learners, categories themselves are significantly heterogeneous. Perhaps the only identifying characteristic of children labeled mildly handicapped is that they are "inefficient school learners whose deviations in school achievement, and possibly social adjustment, are so marked as to necessitate specialized intervention" (MacMillan et al., 1986, p. 686).

The requirement that volunteers attend three half-day training sessions to learn curriculum based measurement (CBM) techniques, and the further requirement that volunteers take curriculum based measures of reading of students in their classes was carefully explained. The benefits to volunteers of learning and using CBM, particularly as a tool for monitoring educational progress of handicapped students, was explained as an incentive to potential volunteers. Additionally, the practical importance of the research for special education teachers was

explained in an effort to increase the rate of volunteering and reduce volunteer bias (Borg & Gall, 1983). From the total population from the two districts, 18 teachers volunteered. Though this was a small sample, statistical significance at the .05 and .01 levels was obtained.

Procedures

Two types of data were gathered for this study, classroom observations of teacher behavior and curriculum based measures of student achievement in reading.

Classroom observation data were collected using a standard observational schedule developed by the Program Development Team for the Beginning Teacher Assistance Program in Virginia. For BTAP, observations of each beginning teacher are completed by independent observers recruited by a Regional Support Unit. Observers receive a total of four days of training spaced over three weeks to allow time between training sessions for practice. Training sessions for observers include: (a) an orientation to the Beginning Teacher Assistance Program; (b) an orientation to the procedures of the observation system including scheduling, reporting, and school building protocol; (c) an orientation to the

instrumentation; and (d) extensive practice and feedback using the observation schedule. All BTAP observers are experienced educators who observe in an area of their expertise, all have participated in observer training and successfully passed a final, practical examination, and all are objective observers who do not know the teachers they observe nor do they observe in a division in which they work (Beginning Teacher Assistance Program, Phase II Final Report, 1984). All observations of subjects in this study were completed by the researcher, who is a trained BTAP observer.

A total of three observations of each teacher were completed between May and June of 1988. The duration of each observation was approximately 35 to 40 minutes, consistent with BTAP procedures. Each observation provided a clear, low-inference record of the teacher's classroom behavior. Specific teaching behaviors were operationally defined, and this observer recorded occurrences of behaviors during seven, three-minute periods scattered throughout an observation. For purposes of analysis, the number of periods in which a teaching behavior was observed was used to estimate the frequency of

that behavior. Subsequently, the frequency of these teaching behaviors was compared with mean student achievement in reading.

Data on student achievement in reading were gathered by subjects using curriculum based measurement. Curriculum based measurement is a type of direct measurement in which brief samples, from one to three minutes, of student performance are obtained using the curriculum as the source of items for testing. Typically, tests are developed from the regular curriculum used in a district and are grade appropriate. For example, passages for reading are obtained by sampling the regular curriculum at the appropriate grade level. All subjects were trained to use the same curriculum based procedures and reading samples developed and validated by the Minneapolis Public Schools for measuring student achievement.

Intervention

All subjects were trained to administer curriculum based measures of reading. Consistent with similar training of teachers of mildly handicapped students conducted by researchers at the University of Minnesota, subjects received one half-day of training, with follow-up to provide them feedback on the accuracy of their

implementation (Skiba et al., 1983). Training was based on the monograph, Considerations for Designing a Continuous Evaluation System (Mirkin, Fuchs & Deno, 1982), and on the Curriculum Based Reading Measures Manual for the Elementary Special Education Program (Minneapolis Public Schools, Special Education Department, 1986). A copy of the manual is included in Appendix A.

Specifically, subjects were trained to (a) select reading passages randomly from appropriate Ginn reading materials; (b) measure using a difficulty level approximating the student's age-grade appropriate level or a level between the age-grade appropriate and instructional levels; (c) administer measures under standardized conditions; and (d) score and record number words correct. Particular attention was given to efficiency of measurement, given constraints on teacher time.

To ensure that subjects accurately administered and scored measures, a random sample of four subjects was observed implementing procedures. Teachers were observed to ensure that passages were (a) randomly selected from the validated reading passages provided by the researcher, (b) at a difficulty level approximating the age-grade appropriate level,

(c) administered for 1 minute with the teacher supplying a word only after the student waited 3 seconds without responding and not saying the correct word after the student said an incorrect word. It was noted that teachers followed procedures as they were trained to do, except that they were reluctant to administer measures approximating the age-grade appropriate level. In the follow-up training session, the importance of measuring at the age-grade appropriate level was stressed.

Ethical Considerations

Informed consent and confidentiality were the principle ethical considerations in this study. This investigator recognized that subjects must be fully informed of all responsibilities that were to be placed on volunteers. The investigator accepted the obligation to inform all potential volunteers that they must attend an inservice training session with follow-up and, further, that they must take curriculum based measures of reading achievement of all students in their classes. Reasonable estimates of the time and energy involved in carrying out these responsibilities were provided to all potential volunteers. Freedom to decline to participate or

to withdraw from the study at any time was ensured. See Appendix B for a copy of the letter soliciting volunteers.

In addition to informed consent of subjects, informed consent of school divisions employing subjects was obtained. The necessity for each subject to take curriculum based measures of achievement of students and for the researcher to observe each subject was explained. The right of any division to decline to have an employee participate was assured. See Appendix C for a copy of the letter to school division principals.

Confidentiality of research data was maintained by randomly assigning an identification number to each subject. All observation data and student achievement data were identified by these numbers. Absolutely no names were used except in securing informed consent of subjects.

Instrumentation

A standard observational schedule and curriculum based measures of student achievement in reading were used to collect data for study. A description of the standard observational schedule and curriculum based measurement of reading achievement, including reliability and validity

evidence follows. Instrumentation is described separately for each measure.

Classroom Process Record

The standard observational schedule, or Classroom Process Record, used in this study was developed by the BTAP Program Development Team and validated during the 1984-1985 school year (Beginning Teacher Assistance Program, Phase III Products and Activities, 1985). There were three validation activities: (a) validation of the indicators of competence, (b) instrument pilot test and revision, and (c) an instrument norming study.

Professional judgment of Virginia educators was used to validate the BTAP indicators of competence. The process for validating competencies included review and revision of twenty-six competencies identified in a "meta" review of reviews on teaching by the BTAP Program Development Team, followed by review and revision by constituent groups, including the Virginia Department of Education and the BTAP Program Advisory Committee, with final review and revision by the BTAP Program Development Team. Instrument specifications including specific instrument items

were validated through essentially the same process.

The Classroom Observation Record was piloted in the Fall of 1984. A group of 90 beginning teachers was observed three times by trained observers. Each teacher was observed once while presenting new material, once while leading a discussion, and once while supervising independent classwork. From the pilot testing, it was determined that three observations were sufficient to produce satisfactory reliability measures. Information gathered during training of observers for pilot testing also indicated satisfactory interrater agreement.

The BTAP instrument norming study included two phases. In the first phase, a stratified random sample of 1500 teachers in Virginia was selected according to geographical region, school district size, and teaching level, including elementary, late elementary, middle school, and high school. Each teacher in the sample was observed once. Information was gathered on settings and on how teachers performed on the competency indicators in given settings. In the second phase, 300 beginning teachers were randomly selected and observed once. Again, information on settings and

how teachers performed on competency indicators in given settings was gathered. Cut scores or minimum levels of performance were determined based on information about the population of teachers and the population of beginning teachers in Virginia. Special education teachers were included in the random samples for both phases of the norming study.

Curriculum Based Measures

Useful systems for curriculum based measurement of student achievement in reading have been developed, and studies suggest the systems are both valid and reliable. One such system has been developed and validated by professionals in the Minneapolis Public Schools in cooperation with researchers at the University of Minnesota. This system of curriculum based measurement is based on the Ginn Reading Series which is the adopted reading curriculum of the two school divisions participating in this study. The system, which includes twenty reading passages at each grade level 1-6, has been shared with this researcher and was used to assess reading achievement of handicapped students in this study.

Early studies at the University of Minnesota of the curriculum based measurement system used in

this study found a significant, positive correlation between composite scores of reading rate and composite scores of reading comprehension (Mirkin et al., 1982). More recent studies investigated the relationship between simple, direct measures of reading achievement, including reading in context, reading words in isolation, and cloze procedures, and scores on standardized reading tests. Deno, Mirkin, Chiang, and Lowry (1980) found reading aloud from text, reading words in isolation, and a cloze test all positively correlated with the Reading Comprehension subtest of the Stanford Diagnostic Reading Test and the Word Identification and Word Comprehension Tests of the Woodcock Reading Mastery Tests. Correlations ranged from +.73 to +.91, with reading aloud from text showing the highest correlation with standardized measures. Similar correlations between direct measures and standardized measures have been obtained by other researchers (Mirkin et al., 1982), with again the best predictor of achievement test scores being reading aloud from text. Test-retest reliability has also been established for reading aloud from text using the curriculum based reading system which was shared with this researcher and used in

this study. Fuchs, Deno, and Marston (1982) tested 30 students in grades 1-6 across four occasions on reading in context measures. Coefficients ranged from +.92 to +.96. Similar coefficients for reading in context measures across occasions have been found by other researchers (Mirkin et al., 1982).

For curriculum based measures of academic performance to be sensitive, valid, and reliable, they must meet certain criteria. These include: absolute unit measurement on a functionally important task (Mirkin et al., 1982). In the case of the measures used in this study, the functionally important task is reading aloud from passages selected from the school division's adopted reading curriculum, the Ginn Reading Series. The absolute unit of measurement is the number of words read correctly. The second criterion is number correct in fixed time (Mirkin et al., 1982). All curriculum based measures in this study were administered for one minute, and scores reported as number of words correct per minute. A third criterion is a difficulty level which remains the same across tests (Mirkin et al., 1982). Twenty reading passages were selected at each grade level 1-6, and screened using the

Fry Readability Index to ensure equivalence within identified levels. To ensure a difficulty level which remains the same across tests, any passage rated one year above or below the grade level of the Ginn basal text from which it was drawn was discarded. The final criterion is a measurement domain which is limited to the age-grade appropriate level, or if the age-grade appropriate level is too difficult, at a level between the instructional and age-grade appropriate level (Mirkin et al., 1982). During training in the use of curriculum based measurement, teachers were instructed to select passages for administration to individual students that were age-grade appropriate, except in cases where students are reading three grade levels down. In these cases, teachers were instructed to select passages at a level between the age-grade appropriate and instructional levels. Teachers were further instructed to select all passages for administration to individual students from the same level across all tests.

The curriculum based measures of reading achievement used in this study meet all of the above criteria. Using the Ginn 720 Reading Series, Marston and Magnusson (1985) randomly

selected 30 passages from each basal reader for grades 1-6. Passages were then screened using the Fry Readability Index to ensure equivalence within identified levels. Passages rated 1 year above or below the grade level of the basal from which they were drawn were discarded. Appropriate grade-level and/or instructional level passages were administered individually each week for 4 months to 309 students in one elementary school in Minneapolis. Mean and slope of words read correctly were calculated for each student, and mean scores of 26 third-graders were correlated with reading and vocabulary subtests of the Stanford Achievement Test, SRA Achievement Series, and Ginn 720 Reading Series. Coefficients ranged from $+.80$ to $+.90$.

The Ginn Reading Series is the adopted reading curriculum in Petersburg City and Chesterfield County Schools. The measures constructed by Magnusson and Marston for the Minneapolis Public Schools have been shared with this researcher and were used to assess reading achievement of students in this study. All subjects in this study were trained to administer the measures according to the procedures developed in the Minneapolis Public Schools. Pre-test

measures were administered individually on three consecutive days to each student by subjects. Subjects waited three weeks, then administered post-test measures individually on three consecutive days to each student.

Sample and Data Gathering Procedures

The purpose of this study is to validate the BTAP indicators of competence for teachers of mildly handicapped students at the elementary level. With the permission of Petersburg City and Chesterfield County Schools in Virginia, the population from which the sample for this study was drawn included all special education teachers of mildly handicapped students in grades 1-5 in Petersburg and the southern portion of Chesterfield. Teachers in the accessible population were apprised of the study and invited to participate. It was explained that subjects would be observed three times between May and June, and would be trained to administer and would administer curriculum based measures of reading achievement to mildly handicapped students in their classes. Training in the use of curriculum based measures of student achievement was scheduled for a half-day session and a one hour follow-up session at sites in Petersburg and

Chesterfield convenient for subjects. The right of subjects to decline to participate or withdraw from the study at any time was guaranteed.

Following training in the use of curriculum based measurement, subjects administered measures of reading achievement to students in their reading classes. The membership in a class averaged 8 students. For purposes of this study, subjects administered measures for three consecutive days, waited three weeks, then administered measures again for three consecutive days. This repeated measurement, pre and post, was consistent with procedures used in the Minneapolis Public Schools for administering these measures to assess reading achievement gains of students. Each measure took approximately 2 minutes to give and 2 minutes to score.

During the three weeks between measures, each subject was observed by the researcher three times. Duration of observations was approximately 35 minutes each consistent with the procedures established by BTAP for Classroom Process Observations (Beginning Teacher Assistance Program, Phase II Final Report, 1984). Observation data were collected using the Classroom Process Record.

Research Design

The research design for this study was causal-comparative. Subjects' scores on the Classroom Process Record were computed to determine frequency of use of effective teaching behaviors as defined by BTAP. A pre and post, mean reading achievement score was then computed for students in each subject's class. With these data, the magnitude of the relationship between teaching behaviors of subjects and reading achievement of subjects' students was computed using the product-moment correlation. Finally, correlation statistics and subjects' scores on the Classroom Process Record were used to form two contrasting groups, a high frequency of effective teaching behaviors group and a low frequency of effective teaching behaviors group. Observed differences between reading achievement scores of the two groups were analyzed for statistical significance, using the t test.

Specific Directional Hypotheses

The purpose of this study was to determine whether there is a significant relationship between the BTAP indicators of competence and reading achievement of mildly handicapped

students. To achieve this purpose, the following directional, null hypotheses were tested:

1. There is no significant difference between pre and post, curriculum based, reading achievement scores of mildly handicapped students.
2. There is no relationship between the teaching behavior of special education teachers, as measured by the Beginning Teacher Assistance Program, and curriculum based, reading achievement scores of mildly handicapped students.
3. There is no significant difference between mean, curriculum based, reading achievement scores of mildly handicapped students in classes of special education teachers who exhibit effective teaching behaviors, as measured by the Beginning Teacher Assistance Program, more and less frequently.

Statistical Analysis Techniques

All data were analyzed using SPSS-X (SPSS Inc., 1986). Hypothesis #1 was tested using the t test. Mean reading achievement scores were computed, pre and post. Observed differences between pre and post, mean scores were then

analyzed for statistical significance. Hypothesis #2 was tested using Pearson's product-moment correlation. Mean frequency of subjects' teaching behavior in 17 categories measured by the Classroom Process Record were computed. These data were analyzed with mean reading achievement scores to determine the magnitude of relationships. Finally, Hypothesis #3 was tested using the t test. Two contrasting groups were formed using correlation statistics and subjects' scores on the Classroom Process Record. Reading achievement scores of the contrasting groups were then analyzed for statistical significance of observed differences.

Summary

The purpose of this study is to validate the indicators being used in Virginia to assess the competence of beginning special education teachers. For purposes of study, two types of data were collected, classroom observation data on teaching behavior of subjects and reading achievement data on mildly handicapped students in subjects' classes. Data were analyzed for relationships between teaching behaviors of subjects in the competency areas specified by BTAP

and reading achievement of students, using a causal-comparative research design.

CHAPTER IV

Overview of Methodology

One of the requirements of the Beginning Teacher Assistance Program in Virginia is that every beginning teacher demonstrate functional knowledge of fourteen generic indicators of competence in actual performance in the classroom. The underlying assumption is that effective teachers of all students share certain teaching behaviors which are indicators of professional competence. The appropriateness of BTAP for evaluating the competence of special education teachers has not been empirically validated, however. It was the purpose of this study to examine the validity of the BTAP indicators of competence for special education teachers of mildly handicapped students in Virginia.

The accessible population for this study included a total of 32 elementary special education teachers of mildly handicapped students in Petersburg City and in the following schools in Chesterfield County: Matoaca Elementary, Ettrick Elementary, Enon Elementary, Wells Elementary, Harrowgate Elementary and Curtis Elementary. This population was made up of both beginning and experienced teachers of special education, all

certified to teach their respective, mildly mentally retarded, emotionally disturbed, and/or learning disabled students. Some were resource and some self-contained teachers. All were sent letters apprising them of this study and asking them to participate (see Appendix for a copy of a letter to building principals explaining the study and asking for their cooperation, and a copy of a letter to teachers requesting their voluntary participation). A total of 18 teachers from the accessible population of 32 volunteered.

The researcher met with volunteers in after school sessions and explained the purposes of the study and the procedures for collecting data on teacher behavior and student outcomes. The requirement that each volunteer attend a half-day inservice training session plus follow-up to learn curriculum based measurement of reading achievement was carefully explained. A total of 4 sessions were conducted, two in Petersburg and two in Chesterfield. Training was continued until all teachers were able to give and score measures with 100% accuracy. All volunteers attended the required training.

The week following training, all volunteers in the sample administered curriculum based

The week following training, all volunteers in the sample administered curriculum based measures of reading achievement to handicapped students in their classes. Each student was administered measures individually on three consecutive days. Self-contained ED and LD teachers administered measures to all students. Self-contained EMR teachers administered measures to all students who were able to read on at least a first grade level. Resource teachers administered measures to all students in at least one reading class. These were considered pre-test scores. Teachers then waited three weeks and administered measures again, individually, on three consecutive days. These were considered post-test scores. Teachers recorded all scores for each student on the Reading Progress Record form, see Figure 1. Forms were then mailed to the researcher.

To obtain mean pre and post-test scores for each teacher's class, pre-test scores of all students in a class were summed and divided by the total number of pre-measures administered. Mean post-test scores were similarly derived.

Observational data on each teacher in the sample was gathered, using the Classroom Process

Record. The Classroom Process Record is a standard observational schedule used in the Beginning Teacher Assistance Program to gather data on 67 specific teaching behaviors in 18 categories. All observations were done by the researcher, who is a trained BTAP observer. Observations were completed between the middle of May and the middle of June, 1988. Consistent with the requirements of BTAP for independent observation to reduce observer bias, the researcher did not know or work previously with any volunteer in the sample.

Each teacher in the sample was observed during 21 different intervals over three observations of 35 to 40 minutes duration. All observations were done during the three weeks between pre and post-testing. For purposes of analysis, total frequency of teaching behaviors in 18 categories, see Figure 2, were computed for each teacher. Mean frequency of behavior in each category for each teacher was then computed by dividing total frequency by total number of observations.

Findings

The purpose of this study was to examine the validity of the BTAP indicators of competence for

Figure 2

Classroom Process Record: Categories of Teacher Behavior

1. Teacher Response
2. Teacher Response: Learner Does Not Answer Question
3. Teacher Follow-up Response
4. Teacher Responses: Learners Who Are:
Inattentive/Disruptive
5. Groups or Individuals Without Teachers
6. Groups With Teacher
7. Began Instructional Activity
8. Stated Expectations
9. Teacher Behaviors During Period
10. Rebuked On-task Learner
11. Interruptions
12. Relationships
13. Changed Activity
14. Made Assignment
15. End of Instructional Activity
16. Environment
17. Participation
18. Off-task

evaluating special education teachers in Virginia. To achieve this purpose, the following specific, directional null hypotheses were tested:

1. There is no significant difference between pre and post, curriculum based, reading achievement scores of mildly handicapped students.

2. There is no relationship between the teaching behavior of special education teachers, as measured by the Beginning Teacher Assistance Program, and curriculum based, reading achievement scores of mildly handicapped students.

3. There is no significant difference between mean, curriculum based, reading achievement scores of mildly handicapped students in classes of special education teachers who exhibit effective teaching behaviors, as measured by the Beginning Teacher Assistance Program, more and less frequently.

Hypothesis 1

There is no significant difference between pre and post, curriculum based, reading achievement scores of mildly handicapped students.

In order to determine if there was significant difference between pre and post, curriculum based, reading achievement scores of mildly handicapped students in this study, the

appropriate analysis was a t-test. To obtain an average or sample mean pre-test score, scores of students in classrooms of participating teachers were summed and divided by the number of cases or 18. The sample mean post-test score was similarly derived. The mean pre-test score was 54.6383. The mean post-test score was 58.9239, for a difference of 4.2856. The t value was 2.90. With 17 degrees of freedom, this difference was significant at the 1% level of probability. (See Table 1.) This finding suggests that curriculum based measures of reading achievement of mildly handicapped students were sensitive to gain in this study.

Hypothesis 2

There is no relationship between the teaching behavior of special education teachers, as measured by the Beginning Teacher Assistance Program, and curriculum based, reading achievement scores of mildly handicapped students.

In order to determine if there was a relationship between the teaching behavior of subjects and curriculum based, reading achievement scores of handicapped students, several analyses were conducted. First, a range and frequency

Table 1

Differences Between Pre and Post-tests of Reading
Achievement

	Mean	Standard Deviation	Standard Error		
Pre-test	54.6383	21.750	5.127		
Post-test	58.9129	24.991	5.890		
Difference Mean	4.2856	Standard Deviation 6.271	Standard Error 1.478	T-Value 2.90	Degrees of Freedom 17
					2-Tail Prob. .010

distribution for each teaching variable, or category of teaching behaviors, measured on the Classroom Process Record was computed. Table 2 summarizes the data on range and frequency distribution of teaching behaviors observed. Ranges for the variables, Teacher Response, Teacher Response: Unanswered Question, Teacher Follow-up Response, Responses to Inattention/Disruption, Groups /Individuals Without Teachers, Groups With Teacher, Teacher Behaviors During Period, Changed Activity, and Made Assignment, were adequate to meet the assumptions necessary for statistical analysis using Pearson's product moment correlation and the t-test. Mean frequencies for the variables, Teacher Response, Teacher Follow-up Response, Responses to Inattention/Disruption, and Teacher Behaviors During Period were also adequate to meet the assumptions necessary for statistical analysis using Pearson's product moment correlation and the t-test. Frequency distributions for other variables were skewed.

Following computation and analysis of frequency distributions, each teaching variable on the Classroom Process Record was analyzed for its relationship to reading achievement of handicapped

Table 2

Ranges and Means of Frequency of Teaching
Behaviors of Subjects

Teaching Behavior	Range	Mean
Teacher Response	6-17	11.222
Teacher Response:		
Unanswered Question	9-14	10.944
Teacher Follow-up Response	0-7	2.722
Responses to Inattention/ Disruption	0-17	6.222
Groups/Individuals Without Teachers	0-7	4.167
Groups With Teachers	0-7	3.278
Began Instructional Activity	0-2	.889
Stated Expectations	0-5	1.833
Teacher Behaviors During Period	7-39	24.556
Rebuked On-task Behavior	1-10	9.111
Interruptions	8-11	9.611
Relationships	0-2	.722
Changed Activity	0-5	1.833
Made Assignment	0-5	1.833
End of Instructional Activity	0-1	.111
Environment	0-1	.167

students. The method of analysis used to test for relatedness was Pearson's product-moment correlation. The results of the Pearson's test are summarized in Table 3. As can be seen, the following variables are positively related to reading achievement of handicapped students: Teacher Response, Participation, Teacher Follow-up Response, Groups/Individuals Without Teachers, Began Instructional Activity, Stated Expectations, Teacher Behaviors During Period, Rebuked On-task Learner, Interruptions, Relationships, Changed Activity, Made Assignment, and End of Instructional Activity. All correlations are in the expected direction except those for Rebuked On-task Learner and Interruptions.

According to the Beginning Teacher Assistance Program, Phase II Final Report, "the beginning teacher should know that students tend to learn best in an affectively neutral and non-punitive environment. The teacher who knows this avoids threats, does not show anger (1984, Appendix C, p. 17). Further, the Phase II Final Report states that, "the beginning teacher should know that students learn best when activities flow smoothly and continuously. The teacher who knows this

Table 3
Correlations Between Teacher Behaviors and Reading
Achievement of Handicapped Students

Teacher Behavior	Coefficient	Probability
1. Teacher Response	.3990	.050
2. Teacher Response: Unanswered Question	-.1201	.317
3. Participation	.1778	.240
4. Off-task	-.3004	.113
5. Teacher Follow-up Response	.3378	.085
6. Teacher Response to Inattention/Disruption	-.3144	.102
7. Groups/Individuals Without Teachers	.2098	.202
8. Groups With Teacher	-.1203	.317
9. Began Instructional Activity	.1797	.238
10. Stated Expectations	.2548	.154
11. Teacher Behaviors During Period	.5648	.007
12. Rebuked On-task Learner	.1198	.319
13. Interruptions	.4737	.024
14. Relationships	.2505	.158
15. Changed Activity	.3106	.105

16. Made Assignment	.0981	.349
17. End	.1067	.337
Instructional Activity		
18. Environment	-.2431	.167

seldom interrupts a presentation or an engaged student or small group" (Beginning Teacher Assistance Program, Phase II Final Report, 1984, Appendix C, p. 3). By definition, the two categories of teacher behavior, Rebuked On-task Learner and Interruptions, should be negatively correlated to reading achievement of students. Table 3 shows negative correlations for the following categories of teacher behavior: Teacher Response Learner Does Not Answer Question, Off Task, Teacher Responses Learners Who Are Inattentive/Disruptive, Groups With Teachers, and Environment. Off Task is the only category which should be negatively correlated with reading achievement. Other categories should show positive relationships.

According to the Beginning Teacher Assistance Program, Phase II Final Report, "(t)he beginning teacher should know that high teacher expectations can increase student learning. The teacher who knows this prompts and encourages a student who falls to answer a question..." (1984, Appendix C, p. 9). Also, "(t)he beginning teacher should know that different children within the same cultural group may learn at different rates and in different ways and respond to different kinds of

motivation. The teacher who knows this organizes the class in different groups for different purposes..." (1984, Appendix C, p. 13). And, "(t)he beginning teacher should know that it is important to be aware of everything that is going on in various parts of the room. The teacher who knows this frequently asks for status, makes contact when a student not in contact with the teacher is off task" (1984, Appendix, p. 15) Further, "(t)he beginning teacher should know that disruption is minimal when rules of behavior are clear and consistently enforced. The teacher who knows this reminds a deviant learner of rules instead of stating them, sometimes by quietly calling the student by name" (1984, Appendix C, p. 16). Finally, (t)he beginning teacher should know that a classroom functions best if the physical environment is adapted to learning activities. The teacher who knows this rearranges furniture for different activities..." (1984, Appendix C, p. 14).

By BTAP definitions, the categories, Teacher Response Learner Does Not Answer Question, Teacher Responses Learners Who Are Inattentive/Disruptive, Groups With Teachers, and Environment, describe effective teacher behaviors. Also by definition,

effective teacher behaviors are those that correlate positively with desirable student outcomes (Medley, 1977). Therefore, reading achievement of handicapped students and the teaching behaviors described by the above categories should be positively related.

Of those categories which showed correlations in the expected direction, two, Teacher Response and Teacher Behaviors During The Period, reached statistical significance, $p=.050$ and $p=.007$, respectively.

Some mention should be made of the statistically significant negative relationship between the category of teaching behaviors described by Interruptions and curriculum based reading achievement scores of handicapped students in this study, $p=.024$. This correlation is highly suspect. Data (see Table 3) suggest that the range and mean frequency for the variable, Interruptions, may not meet the assumptions necessary for analysis using Pearson's product-moment correlation. This correlation, therefore, was discarded in further analysis.

Hypothesis #3

There is no significant difference between mean, curriculum based, reading achievement scores

of mildly handicapped students in classes of special education teachers who exhibit effective teaching behaviors, as measured by the Beginning Teacher Assistance Program, more and less frequently.

In order to determine if there was significant difference between the reading achievement scores of students in classes of special education teachers who exhibited effective teaching behaviors more and less frequently, the results of the Pearson's test were subjected to further analysis. Special education teachers who exhibited effective teaching behaviors more frequently were defined as those who scored above the mean (Rating > 35.78) in the BTAP categories of Teacher Response and Teacher Behaviors During The Period (see Table 4). Ten of 18 subjects fell into this more effective group. Those who scored below the mean on the same categories (Rating < 35.77) were defined as the less effective group. Eight of 18 subjects fell into this group. A t-test was then used to analyze the difference between mean reading achievement scores of the two groups. Results are reported in Table 4. With 15.34 degrees of freedom, a t value of 3.09 was

Table 4

T-test for Difference in Reading Achievement With
Independent Variables of Teacher Behavior During
Period and Teacher Response

	# of Cases	Mean	Standard Deviation	Standard Error	T Value	Degrees of Freedom	2-Tail Prob.
Group 1	10	7.6210	5.190	1.641	3.09	15.34	.007
Group 2	8	.1275	5.030	1.778			

obtained. The probability that this difference is due to random error is .007.

Analysis was also done to determine if special education teachers who used all of the BTAP categories of teaching behavior more frequently had significantly greater reading achievement scores than teachers who used all of the categories less frequently. Table 5 summarizes the results of a t-test. With 9 cases in each group and 15.75 degrees of freedom, a t value of .80 was obtained. There was no statistically significant difference in mean reading achievement scores of the more effective and less effective groups ($p=.437$). Since only two of the teaching behavior categories on the Classroom Process Record showed a positive correlation with reading achievement of handicapped students, it not unexpected that there was no significant difference in reading achievement of handicapped students when all of the BTAP categories of teaching behavior were used in analysis.

Table 5

T-test for Difference in Reading Achievement for All Independent Variables

	# of Cases	Mean	Standard Deviation	Standard Error	T Value	Degrees of Freedom	2-Tail Prob.
Group 1	9	5.4778	5.923	1.974			
					.80	15.75	.437
Group 2	9	3.0933	6.729	2.243			

Chapter V

Summary

In the early 1970's, researchers began to examine what it is that effective teachers do. Through extensive studies correlating teacher behavior with measures of student achievement, specific competencies of effective teachers have been identified and validated (Veldman & Brophy, 1974; Medley, 1977). The state of Virginia has mandated that fourteen such competencies be used to evaluate the professional competence of its beginning teachers through the Beginning Teacher Assistance Program.

Those experts who advocate using specific competencies for evaluating teachers insist that the teaching behaviors which are thought to be indicators of competence must be empirically tested to verify relationships to student outcomes, if, indeed, evaluations are to be valid (Medley et al., 1984; Soar, Medley, & Coker, 1983). Much of the research on effective teaching has been done at the elementary level. Very little research has examined effective teaching in special education. The fourteen competencies used by the state of Virginia to evaluate beginning teachers were taken primarily from research done

in classrooms of regular teachers at the elementary level (Beginning Teacher Assistance Program, Phase II Final Report, 1984). Therefore, certain questions can be raised regarding the validity of these indicators of competence for evaluating special education teachers. It was the purpose of this study to verify relationships between the BTAP indicators of competence and achievement of mildly handicapped students in classes of special education teachers in Virginia. The following null hypotheses were tested to verify relationships:

1. There is no significant difference between pre and post, curriculum based, reading achievement scores of mildly handicapped students.

2. There is no relationship between the teaching behavior of special education teachers, as measured by the Beginning Teacher Assistance Program, and curriculum based, reading achievement scores of mildly handicapped students.

3. There is no significant difference between mean, curriculum based, reading achievement scores of mildly handicapped students in classes of special education teachers who exhibit effective teaching behaviors, as measured

by the Beginning Teacher Assistance Program, more and less frequently.

Observational data on 18 special education teachers of learning disabled, emotionally disturbed, and educable mentally retarded students were gathered, using the Classroom Process Record. The Classroom Process Record is a standard observational schedule used in the Beginning Teacher Assistance Program to gather data on 67 specific teaching behaviors in 18 categories. All observations were done by the researcher, who is a trained BTAP observer. This has to be considered a limitation of the study; though, it should be noted that, consistent with the requirements of BTAP for independent observation to reduce observer bias, the researcher did not know or work previously with any volunteer in the sample.

Each teacher in the sample was observed during 21 different intervals over three observations of 35 to 40 minutes duration. For purposes of analysis, total frequency of teaching behaviors in 18 categories was computed for each teacher. Mean frequency of behavior in each category for each teacher was then computed by dividing total frequency by total number of observations. These data were analyzed for

observations. These data were analyzed for relationships to curriculum based measures of reading achievement of handicapped students in subjects' classes. The three hypotheses under study were tested using t-tests and Pearson's product-moment correlation.

Conclusions

Hypothesis 1

The limitations of this study, which include a small sample size and achievement data on students collected solely by subjects, make all conclusions about the difference between pre and post-test scores subject to further validation. Recognizing these limits, however, the results do provide some support for the conclusion that curriculum based measures of reading achievement of handicapped students are sensitive to gain. This suggests that curriculum based measurement can be used to measure academic gains of mildly handicapped students in a process-product study of teacher effectiveness. This finding is consistent with previous evidence that curriculum based measurement is highly sensitive to short-term gains in academic achievement of handicapped students (Deno et al., 1982; Marston et al., 1983). It is feasible that the use of

curriculum based measures in process-product studies may overcome one of the documented obstacles to validating competencies of special education teachers.

It has been noted previously in this study that "questions about the appropriateness of extant instruments have been raised in regard to diagnosis and assessment for clinical purposes, as well as about research and evaluation efforts where achievement, self-concept, sociometric methods, and even IQ serve as dependent variables or program outcomes to be tapped" (MacMillan, 1986, p. 693). Research efforts to validate competencies of special education teachers using standardized tests of pupil achievement as the outcome measure have been fraught with difficulty. Standardized tests are designed to provide information about the achievement of children in the middle range. The validity of such tests for children outside this range is questionable, "since test scores become unreliable at the extreme ends of the score distribution" (MacMillan et al., 1986, p. 694). Curriculum based measurement techniques correlate highly with standardized achievements tests, yet are sensitive to short-term growth in students with learning

1983). The measures are closely linked to the student's curriculum and are administered repeatedly, thus increasing their reliability. It has been argued that curriculum based measures are preferable to standardized tests for measuring academic progress of handicapped students (Marston et al., 1983). This study presents evidence that these measures may be a viable alternative to standardized tests in research on special education teacher effectiveness.

Hypothesis 2

Within the limitations of this study, which include observation data on teaching behaviors collected solely by the researcher, results of a Pearson's product-moment test suggest that there may be a relationship between the reading achievement of handicapped students and the teaching behavior of special education teachers. The Classroom Process Record of the Beginning Teacher Assistance Program yields data on 67 teaching behaviors in 18 categories. All 67 behaviors are low-inference measures of 14 indicators of competence used to evaluate beginning teachers in Virginia. Teaching behaviors in two of the 18 categories measured by

the Classroom Process Record were significantly related to reading achievement of mildly handicapped students in this study. Several conclusions can be drawn from these results, considering, of course, the limitations of the study.

First, there is evidence that BTAP measures at least two categories of teaching behaviors which are indicators of competence of special education teachers. Among these are behaviors in the categories of Teacher Response and Teacher Behaviors During the Period. The thirty-two teaching behaviors in these categories are low-inference measures of the following BTAP indicators of competence: Academic Learning Time, Accountability, Evaluation, Consistent Rules, Affective Climate, Learner Self-Concept, Meaningfulness, Questioning Skills, Reinforcement, Close Supervision, and Awareness.

Second, there is evidence that there may be indicators of competence which are critical for all teachers. An underlying assumption of BTAP is that "mastery of this set of indicators of competence should...equip a teacher of any subject or grade to perform the basic functions or tasks

describable as 'safe practice'" (Beginning Teacher Assistance Program, Phase II Final Report, 1984, p. 49). While this study certainly does not validate this assumption, it offers some support for the validity of at least two indicators of competence for evaluating special education teachers, similar to those used in regular education. It should be noted, however, that these indicators of competence do not appear to occur in the same context as they do in regular education classes. For example, none of the categories on the Classroom Process Record which measure teacher behavior during group instruction were related to the reading achievement of handicapped students. These categories include Groups with Teachers, Began Instructional Activity, Stated Expectations, Relationships, Changed Activity, Made Assignment, and End of Instructional Activity. In the classrooms observed in this study, these teaching behaviors simply did not occur frequently enough to yield sufficiently valid values and ranges for reliable statistical analysis. This finding is consistent with similar findings in studies of teacher behavior in special education classrooms.

In a review of the research on special education classrooms, Morsink, Soar, Soar, and Thomas (1986), reported a study conducted in 44 special classrooms, including 38 educable mentally handicapped, learning disabled, and emotionally disturbed classes. A low-inference description of teacher behavior in these classrooms suggested that teachers "were engaged in supervising/directing learning experiences, rather than in giving direct instruction..." (Morsink et al., 1986, p. 37). In the current study, teachers in similarly labeled categorical programs for educable mentally retarded, learning disabled, and emotionally disturbed students engaged in teaching behaviors which occurred while they supervised learning rather than provided direct instruction. If effective teaching takes place in groups and is teacher directed as the researchers on effective teaching in regular classrooms suggest (Rosenshine, 1986), this study and others indicate that special education teachers may not demonstrate competencies in this context. On the other hand, this study suggests that in the context of supervising learning activities, special education teachers may demonstrate many of the competencies of teachers in regular

classrooms. Further, within the limitations which have already been described, this study provides some evidence that these competencies, whether demonstrated in the context of group instruction or supervising learning activities, are related to reading achievement of handicapped students.

Hypothesis 3

A t-test of the reading achievement scores in classes of teachers who demonstrated teaching behaviors in the categories of Teacher Response and Teacher Behaviors During the Period more and less frequently revealed a statistically significant difference. A similar t-test of the reading achievement scores of teachers who demonstrated teaching behaviors in all categories of the Classroom Process Record more and less frequently yielded no statistically significant difference. As before, these findings must be interpreted given the limitations of the study. Several conclusions, however, are suggested.

Again, only those teaching behaviors in the categories of Teacher Response and Teacher Behaviors During the Period are related to reading achievement of mildly handicapped students. Indeed, frequency data related to teaching behaviors in other categories appears to mask the

behaviors in other categories appears to mask the effects of teaching behaviors in these two categories. This may suggest a problem with instrumentation. That is, the Classroom Process Record of the Beginning Teacher Assistance Program may measure teaching behaviors that are not related to competencies of special education teachers. It must be emphasized, however, that the BTAP formulas for evaluating the competence of teachers in varying contexts differ by context. Certain teaching behaviors are more heavily weighted in one context than in another. While the formulas are not available, it is reasonable to assume that teaching behaviors in the categories of Teacher Response and Teacher Behaviors During the Period may be more heavily weighted by BTAP for teachers in special education.

Implications

This study has implications for researchers in special education, for those who construct and use systems like BTAP for evaluating special education teachers, and for special education teachers themselves. For researchers, the evidence that there is a relationship between the teaching behavior of special education teachers

and reading achievement of handicapped students suggests that this is a viable area for further research. Indeed, the study raises two immediate questions for future research. One, does increasing the frequency of BTAP or similar indicators of competence among a group of teachers who do not use them as frequently lead to increases in student achievement? That is, is there a causal relationship between the BTAP indicators of competence and the achievement of handicapped students? And two, what are the indicators of competence for special education teachers that are different from those of regular education teachers? This study has offered very little in the way of answer to this second question, except to suggest that the indicators of competence for special education teachers may occur in a context which is different from that of regular education. These two questions are critical. Finally, the study has offered some evidence for the feasibility of using curriculum-based measurement of reading achievement of handicapped students in research on teacher effectiveness in special education. Of course, it must be recognized that constructing

such measures, where they are not already available, is a difficult and lengthy process.

For evaluators, some evidence has been offered to suggest, within the limitations of the study of course, that indicators of competence for regular and special education teachers are similar. An underlying assumption of BTAP, that "mastery of this set of indicators of competence should...equip a teacher of any subject or grade to perform the basic functions or tasks that any teacher is expected to perform at a level describable as 'safe practice'" (Beginning Teacher Assistance Program, Phase II Final Report, 1984, p. 49), may be valid. The questions must still be asked, however, "What is this set of indicators of competence?" and "In what context must they be demonstrated?" The two subjects in this study who had been evaluated by BTAP explained to this researcher that they felt they had to alter the context in which they typically taught, in order to pass BTAP. Specifically, both of these teachers structured their classes, for the BTAP observers, so that they provided direct instruction to a group rather than supervised learning activities of students as was their typical mode of instruction. When this researcher

asked one of these teachers if she ever used this mode of instruction other than for BTAP, she said, "Absolutely not, it wouldn't work with all the students at different reading levels in the class." The results of this study and this vignette may suggest to evaluators that the context in which special education teachers demonstrate competencies may be a critical variable. Evaluations which ignore context or which force special education teachers to alter the context in which they typically teach in order to demonstrate competency may not be valid. This certainly has implications for instrumentation, as well as for definitions of indicators of competence. Finally, for evaluators, the question still remains, "What are the competencies of special education teachers that are different from those of regular education?" Until this question is answered, the validity of evaluations of competence of special education teachers will be questionable.

For special education teachers, many more questions remain than there are answers. There may be a link between the BTAP indicators of competence and reading achievement of mildly handicapped students. No evidence, however, has

been offered to suggest that this is a causal link. Special education teachers who demonstrate these competencies may or may not have students who achieve more in reading. Also, for teachers, many questions remain about the validity of evaluations of their competence. In what context should they demonstrate their competence? What are the indicators of competence for special educators that are different from those of regular educators? Until these questions are answered, it would appear that special educators and their students are best served when the teachers examine their own teaching behavior, seek to increase the frequency of apparently effective teaching behaviors, and carefully and systematically evaluate the learning outcomes of students, using a system like curriculum based measurement. Of course, many variables remain uncontrolled, and conclusions about effective teaching may be highly subjective. With repeated and direct measurement of student learning using curriculum based measurement, conclusions about learning are less subjective. This may be the most practical implication of this research for teachers concerned about their competence and effectiveness.

Limitations of the Study

In a causal-comparative study undertaken in a naturalistic setting, there can be threats to internal and external validity. Among the threats to validity in this study are instrumentation, differential selection, the researcher as sole observer, and subjects as evaluators of reading achievement of students. The wide range of instructional objectives and curricular materials in special education and the failure of standardized tests to be sensitive to changes in achievement of handicapped students preclude the use of standardized tests of achievement for measuring the dependent variable. The use of curriculum based measures of reading achievement is relatively new in process-product studies. Caution, therefore, must be used in interpreting these measures of achievement. The fact that teachers were invited to participate in the study and all were volunteers is a further threat to validity. This threat can only be acknowledged and accepted since it simply was not possible to require teachers to participate. Other threats to validity are the researcher as observer and subjects as evaluators of reading achievement in

their classes. These are threats to validity which must be recognized in interpreting all findings. Observer and evaluator bias cannot be ruled out, though it should be noted that both researcher and subjects were independently trained to administer the various measures used in this study in valid and reliable ways. These threats to validity, however, make findings tentative and subject to further validation.

Recommendations for Future Research

The following recommendations are suggested for future research. First, the causal relationship between BTAP or similar indicators of competence and learning outcomes of mildly handicapped students needs to be examined. Second, the relationship between context and indicators of competence for special education teachers needs to be explored. Along the same line of inquiry, the relationship between the context of teaching in special education and achievement of students warrants study. Finally, much more research needs to be done to determine particular competencies of special education teachers that are different from those of regular education teachers. Until this research is done,

systems designed to evaluate the competence of special education teachers may not be valid.

References

- Adams, K. A. (1983). When to "hold em" and when to "fold em": Ethical problems of internal evaluators. Ohio State University: The National Center for Research in Vocational Educational. (ERIC Document Reproduction Service No. ED 231 858).
- Algozzine, B., Morsink, K. V., & Algozzine, K. M. (1986). Analysis of special education. Gainesville: University of Florida, Department of Special Education.
- Algozzine, K. M., Morsink, C. V., & Algozzine, B. (1986). Classroom ecology in categorical special education classrooms: and so, they counted the teeth in the horse! The Journal of Special Education, 20(2), 209-216.
- Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first grade reading groups. The Elementary School Journal, 79, 193-222.
- Beach, C. M., & Reinhartz, J. (1984). Using criteria of effective teaching to Judge teacher performance. NASSP Bulletin, 68, 31-37.

- Beginning Teacher Assistance Program. (1984).
Phase II Final Report. Virginia Department of
Education: Author.
- Beginning Teacher Assistance Program. (1985).
Phase III Products and Activities. Virginia
Department of Education: Author.
- Borg, W. R., & Gall, M. C. (1983). Educational
Research: An Introduction. New York:
Longman Inc.
- Brophy, J. E. & Good, T. (1986) Teacher behavior
and student achievement. In M. C. Wittrock
(Ed.) Handbook of research on teaching, (pp.
328-375). New York: Macmillan Publishing
Co.
- Brophy, J. E. (1973). Stability of teacher
effectiveness. American Educational Research
Journal, 10(3), 245-252.
- Coker, H., Medley, D. M., & Soar, R. S. (1980).
How valid are expert opinions about effective
teaching? Phi Delta Kappan, 62, 131-149.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R.
(1983). Teacher evaluation in the
organizational context: A review of the
literature. Review of Educational Research,
53(3), 285-328.

- Deno, S. L., Mirkin, P. K. , Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.
- Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehne, K. (1980). Relationships among simple measures of spelling and performance on standardized achievement tests. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.
- Donald, T. M. (1982). The politics of education. Salt Lake City, Utah: International Conference for Parent/Citizen Involvement in Schools. (ERIC Document Reproduction Service No. ED 222 458).
- Ellet, C. D., Capie, W., & Johnson, C. E. (1980). Assessing teacher performance. Educational Leadership, 38(3), 219-220.
- Englert, C. S. (1984). Effective direct instruction practices in special education settings. Remedial and Special Education, 5(2), 38-47.
- Evertson C.M., Anderson C. W., Anderson L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in

Junior high mathematics and English classes. American Educational Research Journal, 17(1), 43-60.

Evertson, C., & Brophy, J. (1973). High-inference behavioral ratings as correlates of teaching effectiveness. Austin: Research and Development Center for Teacher Education, University of Texas. (ERIC Document Reproduction Service No. ED 095 174).

Evertson, C., & Holley, F. (1981). Classroom observation. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 90-109). Beverly Hills: Sage Publications.

Flanders, N. (1970). Analyzing teacher behavior. Reading, Mass.: Addison-Wesley.

Frick, T. W., Polsgrove, L. J., & Reith, H. J. (1986). Executive summary: A special project to develop and implement a computer-based special teacher education and evaluation laboratory. Bloomington: Center for Innovation in Teaching the Handicapped, School of Education, Indiana University.

Fuchs, L. S., Deno, S. L., & Marston, D. (1982). Use of aggregation to improve the

reliability of simple direct measures of academic performance. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.

- Gage, N. L. (1971). The microcriterion. In I. O. Westbury & H. A. Bellack (Eds.), Research in classroom processes (pp. 175-181). New York: Teachers College Press.
- Glass, G. V. (1974). Teacher effectiveness. In H. J. Walberg (Ed.), Evaluating educational performance (pp. 11-32). Berkeley, California: McCutchan Publishing Corporation.
- Haefele, D. L. (1980). How to evaluate thee, teacher--Let me count the ways. Phi Delta Kappan, 61(5), 349-352.
- Hallahan, D. F. & Kauffman, J. M. (1977). Labels, categories, behaviors: ED, LD, & EMR reconsidered. Journal of Special Education, 11, 139-149.
- Harris, W. U. (1981). Teacher command of subject matter. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 58-72). Beverly Hills: Sage Publications.

- Hunter, M. (1984). Increasing your teaching effectiveness. Belmont, CA: Pitman Learning Incorporated.
- King, J. A. (1981). Beyond classroom walls: Indirect measures of teacher competence. In J. Millman (Ed.), Handbook of teacher evaluation (pp. 167-179). Beverly Hills: Sage Publications.
- Larrivee, B. & Vacca, J. M. (1982). Training teachers to apply teaching behaviors which provide for the successful integration of the mildly handicapped. Identifying effective teaching behaviors for mainstreaming. Providence, RI: Rhode Island College, Department of Special Education. (ERIC Document Reproduction Service No. ED 222 036).
- MacMillan, D. L., Keogh, B. K., & Jones, R. L. (1986). Special education research on mildly handicapped learners. In M. C. Wittrock (Ed.) Handbook of research on teaching, (pp. 686-724). New York: Macmillan Publishing Co.
- Manatt, R. P., Palmer, K. L., & Hildebaugh, E. (1976). Evaluating teacher performance with improved rating scales. NASSP Bulletin, 60(401), 21-24.

- Manatt, R. P., & Stow, S. B. (1984). Clinical manual for teacher performance evaluation. Ames, Iowa: Iowa State University Research Foundation.
- Marston, D. & Deno, S. L. (1982). Implementation of direct and repeated measurement in the school setting. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.
- Marston, D., Deno, S. L., & Tindal, G. (1983). A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.
- Medley, D. M. (1977). Teacher competence and teacher effectiveness: A review of process-product research. Washington, D. C.: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED 143 629).
- Medley, D. M., Coker, H., & Soar, R. S. (1984). Measurement-based evaluation of teacher performance. New York: Longman, Inc.
- Millman, J. (1981). Student achievement as a measure of teacher competence. In J. Millman

- (Ed.), Handbook of teacher evaluation (pp. 146-166). Beverly Hills: Sage Publications.
- Mirkin, P. K., Deno, S., Fuchs, L., Wesson, C., Tindal, G., Marston, D., & Kuehle, K. (1981). Procedures to develop and monitor progress on IEP goals. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.
- Mirkin, P. K., Fuchs, L. S., & Deno, S. L. (1982). Considerations for designing a continuous evaluation system: An integrative review. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.
- Morsink, C. V., Dykes, M. K., Algozzine, K. & Fardig, D. B. (1985). Personnel competencies research: Educable mentally handicapped, specific learning disabilities, emotionally handicapped, trainable mentally handicapped, visual disabilities, and motor disabilities. Gainesville: University of Florida, Department of Special Education.
- Morsink, C. V., Soar, R. S., Soar, R. M., & Thomas, R. (1986). Research on teaching: Opening the door to special education classrooms. Exceptional Education, 53(1), 32-40.

- Nagel, S. B. (1982). Ethical dilemmas in policy/program evaluation. Baltimore, Md: Annual Meeting of the Evaluation Research Society. (ERIC Document Reproduction Service No. ED 231 848).
- Pietrowicz, J. (1982). Ethics and educational concerns. In L. H. Orzack & A. L. Simcoe (Eds.), Professions Forum Proceedings. (ERIC Document Reproduction Service No. ED 221 087).
- Quirk, T. J., Witten B. J., & Weinberg, S. F. (1973). Review of studies of the concurrent and predictive validity of the National Teacher Examinations. Review of Educational Research, 43(1), 89-113.
- Rosenshine, B. (1970). Evaluation of classroom instruction. Review of Educational Research, 40(2), 279-297.
- Rosenshine, B. (1971). Teaching behavior related to pupil achievement: A review of research. In I. O. Westbury & H. A. Bellack (Eds.), Research in classroom processes, (pp. 51-98). New York: Teachers College Press.
- Rosenshine, B. (1976). Classroom instruction. In N. L. Gage (Ed.), The Psychology of Teaching

- Methods (pp. 335-371). Chicago, ILL: The University of Chicago Press.
- Rosenshine, B. (1986). Synthesis of research on explicit teaching. Educational Leadership, 43(7), 61-69.
- SPSS Incorporated. (1986). User's Guide to SPSSX (2nd ed.). Chicago: Author.
- Shinn, M., & Marston, D. (1985). Differentiating mildly handicapped, low achieving, and regular education students: A curriculum-based approach. Remedial and Special Education, 6(2), 31-38.
- Skiba, R., Marston, D., Wesson, C., & Sevcik, B., & Deno, S. L. (1983). Characteristics of the time-series data collected through curriculum-based reading measurement. Minneapolis: The Minnesota Institute for Research on Learning Disabilities.
- Skiba, R., Sevcik, B., Wesson, C., King, R., & Deno, S. L. (1983). The non-effect of process-product variables in resource classrooms. (1983). Minneapolis: The Minnesota Institute for Research on Learning Disabilities.

- Soar, R. S., Medley, D. M., & Coker, H. (1983).
Teacher evaluation: A critique of currently
used methods. Phi Delta Kappan, 65(4),
239-246.
- Soar, R. S., & Soar, R. M. (1972). An empirical
analysis of selected Follow Through programs:
An example of a process approach to
evaluation. In I. Gordon (Ed.), Early
childhood education (pp. 229-259). Chicago:
National Society for the Study of Education.
- Stallings, J. A. (1974). Follow Through classroom
observation 1972-1973--Executive summary.
Menlo Park, CA: Stanford Research Institute
(ERIC Document Reproduction Service No. ED
104 970).
- Stallings, J., Needels, M. & Stayrook, N.
(1979). The teaching of basic reading skills
in secondary schools, phase II and phase III.
Menlo Park CA: SRI International.
- Tindal, G., Germann, G., Marston, D., & Deno, S.
(1983a). The effectiveness of special
education: A direct measurement approach.
Minneapolis: The Minnesota Institute for
Research on Learning Disabilities.
- Tindal, G., Marston, D., & Deno, S. L. (1983b).
The reliability of direct and repeated

measurement. Minneapolis: The Minnesota
Institute for Research on Learning
Disabilities.

Travers, R. M. W. (1981). Criteria of good
teaching. In J. Millman (Ed.), Handbook of
teacher evaluation (pp. 14-22). Beverly
Hills: Sage Publications.

Wolking, W. (1985). Pilot probe project: Direct
and continuous learning outcomes.
Gainesville: University of Florida,
Department of Special Education.

Veldman, D. J., & Brophy, J. E. (1974). Measuring
teacher effects on pupil achievement. Journal
of Educational Psychology, 66(3), 319-324.

Ysseldyke, J. E., Thurlow, M. L., Mecklenburg, C.
& Graden, J. (1984). Opportunity to learn for
regular and special education students during
reading instruction. Remedial and Special
Education, 5(1), 29-37.

Appendices

Appendix A

CURRICULUM BASED
READING MEASURES MANUAL

FOR THE ELEMENTARY
SPECIAL EDUCATION PROGRAM

This manual includes all the materials and procedures necessary for monitoring the progress of students in Reading using Curriculum Based Measurement Procedures.

The materials in this manual have been purchased from the Minneapolis Public Schools, Special Education Department, c1986. Permission is granted to the purchaser for reproduction of the student answer sheets for the purpose of noncommercial and individual use, only. Further duplication is prohibited. All Rights Reserved.

SECTION ONE

PROCEDURES FOR MONITORING
READING PROGRESS USING
CURRICULUM BASED MEASUREMENT

Curriculum Based Measurement of Reading Progress

Purpose

The purpose of curriculum based measurement is to monitor student progress in reading toward Individualized Education Program (IEP) goals.

Overview

A major premise underlying curriculum based measurement is that traditional standardized achievement tests do not meet the requirements of a satisfactory progress measuring system. Standardized tests do not adequately measure learning. They may lack content validity, that is they often do not measure what is taught in a typical reading curriculum. Curriculum based measurement provides for repeated measurement over time using measures selected from the adopted curriculum. The measures in this manual were selected from the Ginn Reading Series. All passages were screened using the Fry Readability Formula. Validity, reliability, and stability of the measures have been extensively investigated with coefficients ranging from .80 to .96.

Materials

The following materials are included in Section Two of this manual:

GINN 5	GRADE 1	1-20
GINN 7	GRADE 2	1-20
GINN 9	GRADE 3	1-20
GINN 10	GRADE 4	1-18
GINN 11	GRADE 5	1-18
GINN 12	GRADE 6	1-18

Procedures for Administering Curriculum Based Measures

Progress monitoring will be done weekly. Measures should be collected approximately 6 to 7 days apart.

To begin measuring, choose a numbered passage (see Section Two) from the student's AGE appropriate grade level UNLESS the student is reading more than three years below his age appropriate grade level. In this case, choose a numbered passage BETWEEN the student's instructional level and his/her age appropriate grade level.

Put the unnumbered copy in front of the student. Use the numbered copy to monitor performance and mark errors while the student reads aloud to you.

1. Supply a word ONLY after the student has waited 3 seconds without responding.
2. DO NOT say the correct word after the student has said an incorrect word.
3. Record errors as follows:
 - (a) Put a slash (/) through words read incorrectly (errors). These are recorded as errors: Teacher supplied words, mispronunciations, omissions, words read out of sequence (transpositions) and substitutions.

Repetitions are NOT errors.
Self-corrections are NOT errors.
Dialect characteristics are NOT errors.
 - (b) Circle unusual proper nouns (names such as Séigo) which are incorrectly read.
 - (c) Mark insertions with a caret (^).

Directions

Tell the student: "When I say start, you may begin reading aloud at the top of this page."

"Read across the page." (Demonstrate by pointing.)

"Try every word. If you wait too long on a word, I will tell you the word. Be sure to try every word. If you finish the page, start reading again at the top. Keep on reading until I tell you to stop. DO YOUR BEST READING."

"Do you have any questions?" Ready? Begin."

Start the stopwatch simultaneously as you say "Begin."

After 1 minute, say, "Thank you" and remove the reading passage.

Put a bracket (f) after the last word read.

Scoring

1. Write the number of words read within the minute interval. EXAMPLE: Billy
150 words read
2. Subtract the circled words (incorrectly read unusual nouns) from the words read. These are not counted as errors.
150 words read
- 2 circled words
148 total words read
3. Find the Total Words Correct (TWC) by subtracting errors from the words read.
148 words read
- 10 errors
138 Total Words Correct
4. Mark any insertions with (caret marks). Add insertions to the errors to find the total number of errors.
10 errors
+ 2 insertions
12 total errors

Billy read 138 words correct with 12 errors. This should be written as:
138 TWC

Recording Measures

Record the student's TOTAL WORDS CORRECT (TWC) on the READING PROGRESS RECORD (see Section Three of this manual).

DO NOT record the student's name or TWC in the blanks marked NAME and CORRECT on the reading passages.

SECTION TWO

READING PASSAGES

Note: Included in this section are samples of the measures used at each grade level, 1-6. There are 20 such measures available at each level.

" Eat away! " said City Mouse.	5	1 - 1
" You will like this food. "	10	
Country Mouse said, " I do like it. I may not go back to the country. "	17 25	<hr/> Name
City Mouse said, " Don't go back! You can live here with me. "	31 37	<hr/> Correct
When they were eating,	41	
City Mouse saw something big.	46	
He said, " Run! Run, Country Mouse. And don't stop. "	52 55	
Away went City Mouse.	59	
And away went Country Mouse.	64	
They ran out of the house.	70	
City Mouse called, " Come back, Country Mouse!	75 77	
There is no danger now.	82	
The cat went back into the house. "	89	
But Country Mouse did not stop.	95	
He called, " No, I don't like to live where there is danger.	103 107	
I'm going home. "	110	
Country Mouse ran up a hill and into the country.	116 120	
When he got home, he said,	126	
" At last I can stop!	131	

One night his mother came in when he had just gone to bed and was ready to read.

"Scuffie," she said. "I don't think it is good for you to read so much in bed. And look at your candle. It has dripped down on your bed."

"Please let me read for a little while. I have all these books to read."

"All right," said his mother, "but don't read a long book. And don't forget to shut your outside door. You know why, don't you?"

"Yes, Mother. I know why," Scuffie said.

"You really should shut it right now," his mother said. And she gave him a small good-night mouse kiss.

When she was gone. Scuffie thought, "I really should get up and shut the door right now. I will do it in a little while."

He started to read. Then he heard something outside.

"What is that?" he thought.

It was his neighbors. Big Beetle and Little Beetle, who lived nearby.

"Hello," said Big Beetle. "I see that you are reading in bed."

"And with your door wide open," said Little Beetle. "You really should NOT do that, you know."

The beetles crawled away.

2-1

9
18

27

38

47

57

62

70

79

86

93

101

110

113

121

131

138

146

147

152

160

164

173

176

184

192

193

197

Name

Correct

"This is your home now, my children. We
 welcome you, I and the stars, for we enjoy
 watching you dance." And the boys went on
 dancing, and strangely enough, they found that
 they did not grow tired at all. The bear's song
 grew louder and sweeter. Behind each boy a bright
 star grew, and the moon smiled at their dance.

Then the smallest star boy heard a tiny voice
 from far away. Someone was crying and calling
 his name. Over the sound of the bear's song and
 of his brothers' dancing feet he listened, and he
 heard the distant voice again. It was his mother's
 voice. The smallest boy began to run as fast as he
 could go, with the bright star he was wearing
 making a shining trail behind him.

"Come back, come back," cried his brothers and
 the moon, but the little boy raced away from them.
 Down he flew, past the eagle's nest, past the
 clouds, and closer and closer to the earth, as the
 sound of his mother calling him grew louder and
 louder.

Soon he could see her. She could almost touch
 his hand. Then he landed on the earth. But where
 he landed there was no boy. There was only a
 hole, the kind a star makes when it falls. His
 mother cried still harder when she saw the fallen
 star. Then she looked up and saw her other boys
 dancing in the sky.

8
 17
 25
 32
 42
 51
 60
 69
 77
 87
 96
 105
 116
 125
 131
 139
 149
 158
 168
 177
 178
 187
 197
 207
 217
 226
 236
 240

3-2

 Name

 Correct

In the end they decided on "Teddy" because the puppy looked so much like a small teddy bear, and he even squeaked.

He squeaked and cried—especially at night. No matter how cozy Mary Jo made his bed in the kitchen or how many times Teddy yawned at bedtime, he always woke as soon as everyone was in bed and the house was still. He woke and cried as if his heart would break. Mary Jo put a night-light in the kitchen, in case he was afraid of the dark. She gave him a little snack at bedtime, in case he was hungry. She put an old toy dog in bed with him, hoping he would think it was another puppy. But he didn't.

Mary Jo walked sleepily from her warm bed out to the kitchen a dozen times a night to see Teddy. As long as she was there, he was happy. He tried to get her to play as if it were the middle of the day instead of the middle of the night, and he licked her with his loving puppy tongue. As tired as she was, Mary Jo could never feel angry with him because he was so happy each time she appeared at the kitchen door.

But by the end of the first week she could hardly get up in the mornings. She was almost late for school. Everyone looked tired because although Mary Jo was the one who got up to soothe him, Teddy woke the others with his loud, sad little cries.

8 3-20

17

22

30

Name

40

48

Correct

56

67

78

88

99

111

121

125

133

144

154

167

178

187

197

207

212

222

231

238

248

256

260

 Name

 Correct

The policemen peered after Petros as he wandered along the quay, stopping every few feet to gobble fish. Then they got bored and turned around and dozed off again.	8 17 26 29
Fish by silver fish, and of his own free will, Petros wandered toward the little beach. There Vassili waited, silent, in the little boat.	40 48 53
"Petros?" Vassili whispered.	56
He held up the last of the fish.	64
Petros hopped aboard. The fish curved through the air. Petros' bill clattered. The fish disappeared.	72 79
And then once more Vassili began to row.	87
The next morning was Sunday. Everyone on the island was getting ready to go to church when suddenly Fofa appeared on the waterfront, barking wildly.	95 104 112
"What is Fofa barking about? Doesn't she know it's Sunday?" the Chief Official of the island said as he sat at breakfast.	121 131 134
But Fofa went right on barking. And then, when everyone came out of doors to see what was the matter, she began dashing toward the Cats' Church.	143 154 161
"There must be a cat inside that's just had kittens," the Chief of Police announced. He buttoned up his uniform. "I suppose I'll just have to go along to make sure."	171 180 191 192
A crowd had already gathered outside the Cats' Church when the Chief of Police arrived.	200 207
He blinked.	209
There, standing by the open door of the Cats' Church, was Petros.	218 221
"A miracle!" everyone cried. "Petros is back!" And they ran to tell Thodori.	229 234
Thodori was still asleep.	238
Vassili, from his cot in the corner of the room, saw the Chief of Police and the priest and the Chief Official of the island standing over Thodori's bed.	249 260 267

 Name

 Correct

As the boat disappeared toward the shore James tried 9
 to shake off a feeling of gloom. Would he ever get off 21
 the *Jersey*? How much longer could he endure the 30
 horrors of this ship? 34

James Forten remained a prisoner of the *Jersey* for 43
 three more months, seven months in all. Then, with 52
 the war almost over, he was released in a general 62
 exchange of prisoners. Without shoes and clothed in 70
 rags, he walked the whole distance to Philadelphia. 7-
 After he had rested a few days, he looked up his friend 90
 Daniel Brewton and was overjoyed to find him safe 99
 and restored to health. 103

Daniel Brewton never forgot what James Forten had 111
 done for him. Fifty-six years later Brewton told his 121
 story to William Nell, an early black historian. "With 130
 tears raining down his face," Nell wrote, "the old man 140
 (Brewton) told how James Forten had saved his life 149
 when they were both captives on the prison ship 158
Jersey." 159

As for James Forten, his courage and compassion 167
 were splendidly rewarded. After making a trip to 175
 England as a seaman on an American vessel, James 184
 returned to his own country where he became a worker 194
 in the shop of a famous Philadelphia sailmaker. He 203
 continued to read everything he could get his hands 212
 on, and he tried to be the best worker in the shop. 224
 After a while, he became head man in the sailmaking 234
 shop. After a few more years, he was able to buy the 246
 business for himself. The money for this purchase 254
 was loaned to him by a rich merchant who had known 265
 James all his life and had faith in him. 274

James Forten lived to an old age, a rich and highly 285
 respected citizen. 287

SECTION THREE

READING PROGRESS RECORD

Appendix B



PETERSBURG PUBLIC SCHOOLS

SCHOOL ADMINISTRATION OFFICES
141 EAST WYTHE STREET
PETERSBURG, VIRGINIA 23803-4594
(804) 732-0510

SUPERINTENDENT
SHIRL E. GILBERT II, Ph.D.

EXECUTIVE DIRECTORS
DR. JAMES S. HEYWOOD, *Instruction*
MRS. JOANNE W. NORMAN, *Business & Finance*
ROBERT O. WALLS, *Administrative Operations*

SCHOOL BOARD
WILLIAM A. ARCHIE, *Chairman*
Ms. HEDY ROSENSTOCK, *Vice Chairman*

DR. READING W. BLACK
OLLIE BROWN
MRS. JOAN T. CURLING
DR. PERRY A. MASSEY
REV. ARLESTER NEWSOME, JR.
JOHN W. SCHOEB, SR.
CLYDE T. YANDLE

April 29, 1988

Dear Special Education Teacher:

The purpose of this letter is to request your valued assistance in collecting data for a study designed to investigate the validity of the BTAP competencies for special education teachers. As you know, the State of Virginia requires that all beginning teachers demonstrate their competence in the Beginning Teacher Assistance Program. The BTAP competencies have been validated for regular education teachers. They have not been validated for teachers of special education. As a former special education teacher and program specialist in Chesterfield, I know, like you, that there are differences between handicapped and nonhandicapped students. It is only logical to assume that competencies of regular education and special education teachers may differ. Yet, the State of Virginia evaluates all beginning teachers on the same competencies. As a volunteer in this study, you can make a valuable contribution to our knowledge of special education teacher competencies. You can help to ensure that beginning special education teachers are evaluated on competencies important for success in special education, not regular education.

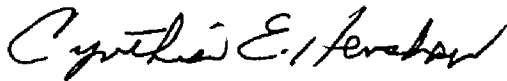
As with all research, there are certain things that I need to ask of volunteers. First, I must ask that volunteers attend a two hour, after school workshop to learn curriculum based measurement for monitoring the progress of handicapped students. Second, I am asking that volunteers administer curriculum based measures to students in one of their reading classes on three consecutive days in May and again on three consecutive days in June. Each measure takes approximately two minutes to administer and two minutes to score. Finally, I am asking volunteers to allow me to do three, 35-minute observations of their teaching over the course of four weeks in May and June.

I realize that this research will have an impact on the time and energy of volunteers. In return for your participation, you will receive a \$25.00 gift certificate for dinner at Steak 'n Ale and a copy of a Curriculum Based Reading Measures Manual. Curriculum based measurement is the best researched approach to monitoring students' progress on IEP goals and objectives. Measures takes only two minutes to give and score, they can be administered as often as necessary, and they are highly valid indicators of progress. Upon request, volunteers may also receive a copy of the "Personnel Competencies Research Project" report, which identifies and documents competencies of special education teachers by category. Finally, all volunteers will have the reward of knowing they have contributed to current knowledge in the field of special education.

If you are willing to be a volunteer in this study, please sign the attached consent form. Return the form to your school secretary by Wednesday, May 11, 1988. I will contact the secretary to get the names of volunteers. I will then contact all volunteers by Friday, May 13th.

Thank you for your attention and consideration.

Sincerely,

A handwritten signature in cursive script that reads "Cynthia E. Henshaw".

Cynthia E. Henshaw
Staff Development Specialist

CEH:ask

Attachment

BTAP VALIDITY STUDY

Assurance of Voluntary Participation

Participation in this study is strictly voluntary. The right of an individual to decline to participate or to withdraw in part or whole at any time is guaranteed. Refusal to participate will not result in any penalty, bias, or loss of benefits.

Availability of Results

Results of this study will be available from:

Cynthia E. Henshaw, Researcher
3208 Osborne Road
Chester, VA 23831
804-796-5212

or

Dr. Douglas Prillaman, Sponsor
Department of Special Education
College of William and Mary
Williamsburg, VA 23185
804-253-4607

Informed and Voluntary Consent to Participate

I have been fully informed and agree to participate in the study. My right to decline to participate or to withdraw in whole or part at any time has been guaranteed.

Volunteer _____

Date _____

CONSENT FORM

The purpose of this form is to request your voluntary participation in a study which will be conducted from September to November, 1988. Federal and state guidelines, as well as moral and ethical obligations, demand that all subjects of research be informed participants. Please read carefully the following information. Then sign in the section marked "Informed and Voluntary Consent to Participate" if you are willing to participate in this study.

Purpose of the Research

The purpose of this study is to validate certain indicators of competence currently used to assess beginning teachers in Florida and Virginia. The indicators of competence which will be investigated have been validated for teachers of regular education. They have not been validated for teachers of special education.

Amount of Time Involved for Subjects

The length of the study is six weeks, from September to November of 1988. All volunteers will be asked to attend a 2 and 1/2 hour inservice training session and a 1 hour follow-up session to learn Curriculum-Based Measurement procedures for monitoring the progress of students in reading. Training sessions will be scheduled in September. Following training, volunteers will be asked to take individual reading achievement measures of 5 to 8 students in their classes for three consecutive days, wait four weeks, and take measures again for three consecutive days. Each reading measure takes approximately 4 minutes to administer and score. For 5 to 8 students, the impact on subjects' time is estimated to be 20-32 minutes per day for a total of 6 days. All volunteers will also be asked to allow a trained observer to do a weekly classroom observation of approximately 35 minutes duration during three of the four weeks between reading achievement measures.

Assurance of Confidentiality

All data collected for purposes of study will be kept strictly confidential. Classroom observation data on volunteers and reading achievement data of students in volunteers' classes will be identified by Social Security number only. The researcher, and only the researcher, will have access to data collected on individual volunteers. For purposes of analysis, only group data will be used. No data will be

used for any purpose except that expressly specified in this study.

Assurance of Voluntary Participation

Participation in this study is strictly voluntary. The right of an individual to decline to participate or to withdraw in part or whole at any time is guaranteed. Refusal to participate will not result in any penalty, bias, or loss of benefits.

Availability of Results

Results of this study will be available from:
Cynthia E. Henshaw, Researcher
3208 Osborne Road
Chester, Virginia 23831
804-796-5212

or

Dr. Douglas Prillaman, Sponsor
Department of Special Education
College of William and Mary
Williamsburg, Virginia
804-253-4607

Informed and Voluntary Consent to Participate

I have been fully informed and agree to participate in the study outlined above. My right to decline to participate or to withdraw in whole or part at any time has been guaranteed.

Volunteer _____ Date _____

Appendix C



PETERSBURG PUBLIC SCHOOLS

SCHOOL ADMINISTRATION OFFICES
141 EAST WYTHE STREET
PETERSBURG, VIRGINIA 23803-4594
(804) 732-0510

SUPERINTENDENT
SHIRL E. GILBERT II, Ph.D.

EXECUTIVE DIRECTORS
DR. JAMES S. HEYWOOD, *Instruction*
MRS. JOANNE W. NORMAN, *Business & Finance*
ROBERT O. WALLS, *Administrative Operations*

April 29, 1988

SCHOOL BOARD
WILLIAM A. ARCHIE, *Chairman*
Ms. HEDY ROSENSTOCK, *Vice Chairman*

DR. READING W. BLACK
OLLIE BROWN
MRS. JOAN T. CURLING
DR. PERRY A. MASSEY
REV. ARLESTER NEWSOME, JR.
JOHN W. SCHOEB, SR.
CLYDE T. YANDLE

Dear Principals:

The purpose of this letter is to ask your permission to contact the LD, ED, and EMR teachers in your school to solicit their participation in a study designed to validate BTAP competencies for special education teachers. I have shared the intent and procedures for this study with Dr. Finkler and have his approval to approach you.

The study will not impact on your time, at all, other than the decision to agree that your teachers may participate. The impact on students will be minimal. Teachers in the study will administer a total of six one-minute measures of reading progress to a group of six (6) to eight (8) of their students.

As a former special education teacher and program specialist, I feel it is important to demonstrate that the competencies used to certify special education teachers are related to achievement of handicapped students. I know that you want the new special education teachers you hire to be certified on the basis of competencies that are correlated to a teacher's success with special education students.

I will call your office on May 8th to answer any questions you may have and seek your approval to ask your teachers for their voluntary participation. If you do approve, would you have your secretary place a copy of the attached consent letter in the box of each of your LD, ED, and EMR teachers. I am asking volunteers to return consent forms to your secretary. I will then obtain their names from your secretary and contact those who indicate a willingness to participate.

Thank you for your consideration and valuable contribution.

Sincerely,

Cynthia E. Henshaw
Staff Development Specialist

CEH:ask

Attachment

Abstract

In Virginia, the State Board of Education has specified fourteen areas in which every teacher who is granted a Collegiate Professional Certificate must be competent and has established the Virginia Beginning Teacher Assistance Program (BTAP) for the purpose of assessing the competence of beginning teachers. The basis of BTAP is a "set of measurable or observable indicators through which beginning teachers can demonstrate their competence in each competency area specified by the Board of Education." In studies of teacher effectiveness in regular education, these competency areas have been consistently related to increased student achievement. Few studies in special education have sought to demonstrate that these competency areas relate to increased achievement for handicapped students. The purpose of this study was to validate the BTAP indicators competence of for special education teachers. Classroom observation data on teaching behavior of subjects and reading achievement data on mildly handicapped students in subjects' classes were collected. Data were analyzed for relationships between teaching behaviors of subjects and reading achievement of students.