DOCUMENT RESUME

ED 293 876 TM 011 466

AUTHOR Cloud-Silva, Connie; Denton, Jon J.
TITLE The Development of a Low-Inference Observation Instrument To Assess Instructional Performance of Teaching Candidates.
PUB DATE Apr 88
NOTE 42p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; Elementary Secondary Education; Evaluation Methods; Interrater Reliability; *Minimum Competencies; Reliability; *Teacher Evaluation; Teacher Supervision; Teaching Skills; Validity
IDENTIFIERS *Teacher Candidates

ABSTRACT

A prototype low inference observation instrument to measure minimal teaching competencies of teaching candidates was deductively developed. Focus is on determining if observers could be trained to use the observation instrument with a high degree of reliability and validity. The instrument, entitled Classroom Observation and Assessment Scale for Teaching Candidates (COAST), was developed from appraisal systems of the states of Florida, Georgia, and Texas. The initial draft was field-tested, and a panel of 11 experts in teacher evaluation reviewed the indicators. The final form of the instrument covered: (1) instructional planning; (2) classroom management and organization; (3) instructional techniques and materials; (4) communication and presentation techniques; (5) motivation of students; and (6) assessment of students. The instrument was tested on 27 teacher candidates (categorized in three grade level subdivisions of grades 1-12) with seven observers (six graduate students and one university faculty member). Inter-rater agreement was acceptable. Results suggest that COAST is an instrument that can be used across subject areas and grade levels with a high degree of content validity. Seven tables are presented, and an outline of the COAST Instrument Indicators is appended. (SLD)

The Development of a Low-Inference Observation Instrument to Assess Instructional Performance of Teaching Candidates

Connie Cloud-Silva and Jon J. Denton

Texas A&M University

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, April 6, 1988

The Development of a Low - Inference Observation Instrument to Assess Instructional Performance of Teaching Candidates

Intense efforts have been made at the national and state levels to develop comprehensive observation instruments based on empirical research findings. The state of Texas, for example, has had to respond to legislation that requires the State Board of Education to adopt an appraisal process and to establish criteria to be used in the appraisal of classroom teachers. This system consists of a generic observation instrument used with all subject area and grade level teachers. Because the appraisal system is based on the premise that teachers consider evaluation systems that are clearly linked to teacher effectiveness literature as more fair and credible (Knapp, 1982), the instrument is comprised of teaching behaviors that have been found to be valid and reliable predictors of student achievement (Texas Education Code, Section 13.302(b).

Purpose of the Study

The purpose of the present study was to develop a prototype, low-inference observation instrument deductively derived from teacher effectiveness research that would measure minimal teaching competencies of teaching candidates. The study was designed to determine if observers could be trained to use the observation instrument with a high degree of reliability and validity. It was thought that the findings of the study might also provide an indication of how well candidates exhibit instructional behaviors required of inservice teachers.

The research questions that guided the development of the observation instrument included: (1) Will the instrument be characterized by a high

degree of content and face validity? (2) Can different observers use the instrument with a high degree of criterion-related agreement? (3) Can an observer use the observation instrument on different occasions with a high-degree of intraobserver agreement? (4) Will applications of the instrument reveal a high degree of interobserver agreement? and (5) Will different contextual conditions, such as subject area and grade level, affect the generalizability of the instrument?

METHOD OF THE STUDY

Initial Development of the Instrument

The development of the observation instrument, entited Classroom Observation and Assessment Scale for Teaching Candidates (COAST), was comprised of several steps. The first step was to examine existing observation systems and teacher effectiveness literature to identify low-inference teaching behaviors associated with gains in student achievement. The appraisal system for the states of Florida (Florida Performance Measurement System, 1983) and Georgia (Teacher Performance Assessment Instrument, 1984), and preliminary drafts of the Texas appraisal system (Texas Teacher Appraisal System, 1986) gave direction to the format and structure of the instrument. To simplify the task of examining and structuring the content of the literature, research findings were categorized into six domains of teaching behaviors. Teaching behaviors representative of each domain were identified and served as indicators of teaching performance on the observation instrument. Behavioral descriptions and specific teaching examples for each of the indicators were developed. This comprised the intial draft of the instrument.

The intial draft was then field-tested at both the elementary and secondary levels with a variety of different subject areas. Field testing

served a dual purpose: first, to investigate the occurrence of the behaviors in classrooms; and second, to assist in the completion of a list of indicators and indicator descriptions that could be used in determining content validity.

Judges' Determination of Content Validity

A panel of 15 judges were asked to evaluate the indicators and the indicator descriptions. Eleven of the 15 agreed to participate. Of these eleven, five were experts in teacher effectiveness literature and six were Texas school administrators active in evaluating inservice teachers. Judges were asked to indicate if they agreed or disagreed that each specific indicator was important to the teaching act, and also to rate the wording of each indicator as clear or unclear. A criterion level of .75 was established for placement of indicators on the observation instrument prior to receiving feedback from the judges. The indicators received high ratings in terms of importance and clarity. In response to written suggestions made by the judges, redundant indicators were removed, the length of the instrument was reduced, and some of the behavioral descriptions of the indicators were rewritten in an attempt to remove ambiguous terminology and to more clearly define the behaviors.

Final Form of the Instrument

The observation instrument in its final form was comprised of six domains of teaching behaviors: (a) Instructional Planning; (b) Classroom Management and Organization; (c) Instructional Techniques and Materials; (d) Communication and Presentation Techniques; (e) Motivation of Students; and (f) Assessment of Students. The domains of Classroom Management and Organization and Instructional Techniques and Materials were subdivided into three teaching behavior subdomains. Each of the domains and the subdomains included five teaching behaviors, or indicators, except for the Lesson Cycle

- Instructional Behaviors subdomain, which had ten, and the Assessment of Students domain which had two behaviors.

COAST is a structured sign system designed to determine the presence or absence of 62 low-inference teaching indicators in a teaching episode. Directions for use of the instrument indicate that observers are to check instrument indicators as they occur during a lesson. A tally is made at the conclusion of a lesson of the indicators that did occur. The stipulation was made in the coding procedures that if a behavior did not occur during the lesson because of the inappropriateness of the behavior to the lesson, this was to be indicated by the observer.

Procedures of the Study

Training Observers

Videotapes wre made of the actual instructional sequences of three elementary teachers. All three tapes were coded by the instrument developer (the criterion person) prior to the training session. A training manual, consisting of an explanation of the purpose of the study, behavioral descriptions of the indicators, specific teaching examples of the indicators, and the procedures to follow in using the instrument to code instructional classroom processes, was developed and discussed with the seven observers who participated in the study.

One practice tape was used to demonstrate specific examples of the instrument indicators and to aid in the explanation of the indicators. Observers then watched the second practice videotape and practiced coding without direct assistance from the instrument developer. Coding tabulations were compared to those of the instrument developer (who also served as the criterion person in the study) and clarification was given as necessary. The observers then viewed the third and last training tape, and their coding

was again compared to that of the criterion person. Because criterion-level agreement on videotapes representative of actual classrooms should be .75 or higher (Frick & Semmel 1978; Medley & Norton, 1971), this level was established prior to training and training was to continue until this level was reached by each of the observers.

The seven observers who participated in the study included six graduate students and one faculty member from a major Southwestern university. All seven observers had had prior experience in evaluating preservice teachers. Four of the observers participated in the study for two semesters and three participated in the study for one semester.

Recruitment of Teaching Candidates

A total of 27 teaching candidates volunteered to be observed four times during the course of their student teaching semester. Two of the observations were to be simultaneously coded by the criterion person and an observer. For the purpose of analyzing the data, and to test the generalizability of the instrument, teaching candidates were categorized into three grade-level subdivisions and represented different subject areas. The grade level subdivisions included grades 1-5, 6-8, and 9-12. Fourteen teaching candidates represented the basic skill areas of mathematics and language arts, and the other 13 teaching candidates represented other subject areas.

Statistical Design of the Study

Measures were taken to answer the research questions posed in the study. These questions dealt with content validity, interobserver agreement, intraobserver agreement, criterion-related agreement, and generalizability of COAST. Content validity was established by having a panel of eleven judges evaluate the instrument indicators with respect to

their importance and clarity.

Criterion-related agreement, defined as the percentage of agreement between the observers and the criterion-person, was calculated using videotaped and actual classroom data. Four of the observers collected data across two semesters and coded the training tape before, during, and after the observation cycle. These three codings were compared to the codings of the criterion person. Three of the observers participated in the study for one semester and coded the training tape before and after data collection. These two codings were also compared to codings of the criterion person.

Intraobserver agreement, defined as the consistency between observer coding on separate occasions of the same videotape of actual classroom sequences, was also calculated. Intraobserver agreement of the four observers who participated across two semesters was determined by comparing their codings before, during, and after the data collection. The intraobserver agreement obtained by the three observers who participated for one semester was determined by comparing the codings made before and after data collection.

Interobserver agreement, also known as interrater agreement, was also determined in this study in an effort to examine the extent to which the disagreement between the observers and the criterion person limited the reliability of the instrument. Interobserver agreement was determined by calculating the agreement between each observer's coding and the criterion person's coding on two simultaneously coded classroom observations for each of the 27 teaching candidates.

The generalizability of the observation instrument was determined by comparing the observer agreement coefficients obtained in each of the three grade level subdivisions and the two subject area categories. The frequency of instrument indicator occurrence for each grade level subdivision and

subject are category was also calculated to investigate if some behaviors occurred more often at some grade levels and/or with certain subject areas.

All observer agreement measures were calculated using Scott's (1955) coefficient. Scott's coefficient is one of the percent agreement indices that has been used to evaluate the interrater consistency of many teacher and student variables at many grade levels (Anderson, Evertson, & Brophy, 1978; Good & Grouws, 1977). Scott's coefficient is appropriate for calculating reliability estimates because the method is suitable for calculating observer agreement on the total categories of an observation instrument, comparing two observers' codings at a time. The method is also unaffected by low frequencies across categories and this was particularly suitable for the data analyses in this study.

Table 1 provides an outlay of the study indicating the data collection procedures and data treatment procedures that were utilized to answer the research questions posed in the study.

RESULTS OF THE STUDY

A criterion level of .75 was established prior to receiving responses from the judges. All behaviors reached this level regarding their importance to the teaching act. Agreement levels ranging between 91% and 100% were found for 51 of the 60 indicators. The level of agreement was 82% on the other nine indicators. Pertaining to clarity of the indicators, 49 of the 60 indicators had an agreement level of 91% to 100%. Nine of the indicators had 82% agreement, and two of the indicators had 73%.

Changes were made in the instrument based on the judges' scoring of the indicators and their comments regarding the behavioral descriptions of the indicators. Efforts were made to reduce the length of the instrument, to eliminate redundant indicators, and to eliminate indicators that required

Table 1

Outlay of Data collection Procedures and Data Treatment Procedures

Research Questions	Data Collection	Treatment of the Data
#1 - Will the instrument be characterized by a high degree of content validity?	Before the Data Collection. A panel of 15 judges.	Decisions rules made prior to receiving feedback from the judges. Criterion-level set at .75.
#2 - Can different observers use the instrument with a high degree of criterion-related agreement?	Before, during, and after data collection by using videotaped examples of actual classroom processes. Before training, the criterion-level of intraobserver agreement was set at .75.	Used Scott's coefficient to determine percentage of agreement. Coders must reach a minimal level agreement of .75.
#3 - Can an observer use the observation instrument on different occasions with a high degree of intraobserver agreement?	Videotapes of actual classroom processes used during training were coded on three separate occasions.	Used Scott's coefficient to determine percentage of agreement.
#4 - Will a test of the instrument reveal a high degree of intraobserver agreement?	Observer and criterion person ratings were gathered after data collection.	Used Scott's coefficient to determine the level of agreement between the ratings of the observers and the criterion person.
#5 - Will different contextual conditions, such as subject-matter area and grade level, affect the generalizability of the instrument?	Observer and criterion person ratings were gathered after data collection for each of the three grade level divisions and the subject-matter areas.	(1) Comparisons of inter-observer agreement coefficients across grade levels and subject areas was made. (2) Percentages of occurrence of each behavior across grade level divisions and subject-matter areas was calculated.

observers to make high-level inferences. Although changes were made in the isntrument indicators, most of the changes involved clarification of indicators and behavioral descriptions of the indicators. The substance of the indicators remained basically unchanged, as the judges' ratings clearly reflected a high level of content validity. Table 2 presents the percentage of agreement of each of the instrument indicators.

Criterion-related agreement was calculated on all seven observers before data collection began (following training) and after the data collection. Criterion-related agreement was also calculated between the two semesters for the four observers who collected data both semesters. The criterion-related agreement on the videotapes exceeded .80 for each of the seven observers, which has been considered an acceptable level of accuracy for sign systems (Frick & Semmel, 1978). Table 3 presents the criterion-related agreement.

The results shown in Table 4 indicate that the intraobserver agreement measures on the videotaped tests demonstrated acceptable observer consistency and compared favorably with the consistency of observers reported in other experimental studies (Frick & Semmel, 1978; Medley & Norton, 1971). The observers were also consistent among themselves in their coding, and their levels of agreement indicated that their coding did not deteriorate during the data collection process.

Interobserver agreement was determined by having the seven observers code simultaneously with the criterion-person on two separate occasions in each of the 27 teaching candidates' classrooms. There was a different number of simultaneously coded classroom observations for each of the seven observers. This occurred because the observers, who were also serving as student teacher supervisors, had different numbers of teaching candidates

Table 2

Content Validity of Instrument Indicators

Importance			Indicators	Clarity	
Agree	Disagree			Clear	Unclear
		2.0	Classroom Management and Organization		
		2.1	Management of Instructional Time		
91%	9%	2.1.1	begins class promptly	100%	0%
91%	9%	2.1.2	engages students in learning activities for a majority of the available class time	73%	27%
100%	0%	2.1.3	routines are established for recurring activities	91%	9%
100%	0%	2.1.4	provides sufficient time for students to complete instructional activities	100%	0%
100%	0%	2.1.5	makes smooth transitions between activities	91%	9%
		2.2	Instructional Organization of the Classroom		
100%	0%	2.2.1	materials and facilities are ready for at the start of the lesson	100%	0%
82%	18%	2.2.2	arrangements for seating/grouping are appropriate for instructional activities	82%	18%
82%	18%	2.2.3	schedules for assignments are made available to students	82%	18%
91%	9%	2.2.4	revises schedules as necessary	82%	18%
91%	9%	2.2.5	arranges classroom for a variety of instructional purposes	73%	27%

Table 2 (Continued)

		2.3	Management of Seatwork/Homework		
100%	0%	2.3.1	gives clear directions for seatwork/ homework assignments	100%	0%
100%	0%	2.3.2	indicates expected completion time for assignments	100%	0%
100%	0%	2.3.3	circulates and provides assistance to students	91%	9%
91%	9%	2.3.4	provides a variety of meaningful seatwork/homework assignments	91%	9%
82%	18%	2.3.5	provides opportunities for students to interact while completing assignments	82%	18%
		2.4	Management of Student Conduct		
91%	9%	2.4.1	establishes classroom management procedures appropriate for the students in the classroom	91%	9%
91%	9%	2.4.2	administers established rules and procedures for classroom behavior	91%	9%
100%	0%	2.4.3	monitors classroom activities	100%	0%
82%	18%	2.4.4	reinforces appropriate classroom behavior	91%	9%
82%	18%	2.4.5	establishes physical arrangement of the classroom to promote good behavior	82%	18%

Table 2 (Continued)

		3.0	Instructional Techniques and Materials		
100%	0%	3.1.1	gains attention of students to begin instruction	91%	9%
100%	0%	3.1.2	provides anticipatory set/gives focus of lesson	91%	9%
100%	0%	3.1.3	links lesson to previous experiences	100%	0%
91%	9%	3.1.4	determines starting points of students	82%	18%
100%	0%	3.1.5	provides lesson objectives to students	100%	0%
		3.2	Lesson Cycle - Instructional Behaviors		
100%	0%	3.2.1	presents accurate stimulus information and materials	91%	9%
100%	0%	3.2.2	provides learner guidance of new information	91%	9%
100%	0%	3.2.3	checks student understanding of new information	91%	9%
82%	18%	3.2.4	addresses specific needs of all students	91%	9%
100%	0%	3.2.5	checks the work and/or responses of all students	100%	0%
100%	0%	3.2.6	provides guided practice of new learnings by students	100%	0%
100%	0%	3.2.7	provides independent practice of new learnings by student	100%	0%
100%	0%	3.2.8	provides feedback concerning student responses	100%	0%
100%	0%	3.2.9	reteaches/remediates during lesson if necessary	100%	0%
100%	0%	3.2.10	reviews and summarizes information in closure of lesson	100%	0%

Table 2 (Continued)

		3.3	Use of Instructional Material		
100%	0%	3.3.1	uses materials appropriate for the attainment of lesson objectives	92%	8%
100%	0%	3.3.2	provides directions for use of the instructional materials	100%	0%
91%	9%	3.3.3	uses audio-visual resources effectively	91%	9%
100%	0%	3.3.4	uses supplementary material/ goes beyond text when necessary	91%	9%
91%	9%	3.3.5	provides sufficient quantities of materials	100%	0%
		4.0	Communication and Presentation Techniques		
100%	0%	4.1.1	uses correct syntax in oral discourse	100%	0%
100%	0%	4.1.2	uses correct syntax and spelling in written communication	100%	0%
91%	9%	4.1.3	varies voice characteristics during delivery of lesson	91%	9%
82%	18%	4.1.4	communicates effectively with special learners in the classroom	82%	18%
100%	0%	4.1.5	emphasizes important points in lesson	91%	9%
100%	0%	4.1.6	encourages students to express ideas and interests	100%	0%
100%	0%	4.1.7	communicates enthusiasm for subject matter	91%	9%
100%	0%	4.1.8	accepts and incorporates ideas of students into the presentation	100%	0%
100%	0%	4.1.9	varies mode of presentation	91%	9%
100%	0%	4.1.10	demonstrates appropriate nonverbal communication	91%	9%

Table 2 (Continued)

		5.0	Motivation of Students		
100%	0%	5.1.1	provides options for students and/or allows students to choose and make options	82%	18%
100%	0%	5.1.2	provides opportunities for students to initiate ideas, discussion, and activities	100%	0%
100%	0%	5.1.3	creates a relaxed, comfortable atmosphere for learning	91%	9%
91%	9%	5.1.4	shows appreciation for individual differences	91%	9%
100%	0%	5.1.5	provides opportunities for students to share and display work efforts	100%	0%
		6.0	Academic Assessment of Students		
100%	0%	6.1.1	provides formative feedback to students on oral and written tests, assignments, homework and classwork	100%	0%
82%	18%	6.1.2	uses student performance on tests and assignments	91%	9%
82%	18%	6.1.3	uses a variety of assessment techniques to determine the learning patterns of students	91%	9%
100%	0%	6.1.4	uses evaluation techniques appropriate for stated objectives	91%	9%
91%	9%	6.1.5	provides reinforcement and feedback for learning efforts of students	100%	0%

Table 3

Criterion - Related Agreement of the Observation Instrument

Observers	Before Data Collection	During Data Collection	After Data Collection
1	.85	.98	.86
2	.88	.91	.87
3	.85	.93	.91
4	.88	.88	.94
5	.91		.93
6	.88		.85
7	.86		.92

Table 4

Intraobserver Agreement

Observers	During Data Collection	After Data Collection
1	.90	.90
2	.85	.85
3	.87	.85
4	.91	.90
5		.92
6		.95
7		.93

who volunteered to participate in the study. Table 5 provides the interobserver coefficients for each of the seven observers. The interobserver agreement coefficients for each of the seven observers were as follows: Observer #1 = .91; Observer #2 = .89; Observer #3 = .85; Observer #4 = .92; Observer #5 = .94; Observer #6 = .90, and Observer #7 = .82. The total average of interobserver agreement was .89, indicating that different observers can use the instrument with a high degree of consistency and skill.

Two steps were taken to determine the generalizability of the instrument. First, the levels of interobserver agreement were compared across grade levels and subject matter areas. Second, frequency of indicator occurrence across grade levels and subject area categories was calculated. As previously mentioned, there were three grade level subdivisions, grades 1-5, 6-8, and 9-12, and also two subject area categories, the basic skill areas of mathematics and language arts and "other" subject areas.

At the elementary level, grades 1-5, five teaching candidates were observed teaching mathematics or language arts; two were observed teaching social studies; and two were observed teaching science. At the junior high level, grades 6-8, four teaching candidates were observed teaching mathematics or language arts; two were observed teching social studies; and one was observed teaching physical science; and one other was observed teaching history. At the high school level, grades 9-12, five teaching candidates were observed teaching mathematics or language arts; two were observed teaching government; two were observed teaching biology; and one was observed teaching chemistry.

Table 6 presents the interobserver agreement coefficients for each of

Table 5

Interobserver Agreement of Simultaneously Coded Observations

Observers													
1		2		3		4		5		6		7	
.98		.74		.92		.89		.92		.86		.85	
.97		.88		.91		.98		.91		.84		.84	
.91		.96		.88		.87		.93		.94		.77	
.86		1.00		.81		.94		.99		.94		.82	
.89		.88		.87									
.91		.84		.35	Avg.=	.92	Avg.=	.94	Avg.=	.90	Avg.=	.82	
.89		.91		.78									
.88		.94		.80									
.87				.78									
.81	Avg.=	.89		.80									
.88				.79									
.92				.94									
.87				.94									
.94				.89									
.95													
1.00			Avg.=	.85									
Avg.= .91													

Total Average = .89

Table 6

Interobserver Agreement and Averages Across Grade Levels and Subject Areas

	Mathematics and Language Arts						Other Subject Areas					
*	1	2	3	4	5	Avg.	6	7	8	9	Avg.	
Grades 1-5	.85	.92	.80	.92	.98	= .89	.85	.85	.87	.98	= .88	
	10	11	12	13	Avg.		14	15	16	17	Avg.	
Grades 6-8	.90	.89	.85	.88	= .88		.86	.85	.79	.79	= .82	
	18	19	20	21	22	Avg.	23	24	25	26	27	Avg.
Grades 9-12	.90	.87	.89	.94	.93	= .91	98	.96	.94	.91	.92	= .94

* Note. Numbers 1 to 27 represent the 27 teaching candidates being observed during the study.

the three grade level subdivisions and also for the two subject categories. The coefficients were relatively high and consistent across the grade level subdivisions and subject areas. At the elementary subdivision, the average interobserver agreement for areas of mathematics and language arts was .89, and .88 for the subject areas of health and social studies. At the junior high level, the average interobserver agreement coefficient was .88 for mathematics and language arts, and .82 for the subject areas of social studies, physical science, and history. Interobserver agreement coefficients were higher at the high school level as .90 was obtained for mathematics and language arts and .94 was obtained for the subject areas of government, chemistry, and biology.

The second measure taken to determine the generalizability of COAST dealt with determining the frequency with which the instrument indicators occurred across subject areas and grade levels. Frequency percentages indicate that, overall, the consistency of indicator occurrence was stable across the contextual settings respresented in the study. These frequency percentages are shown in Table 7.

Frequency results did suggest, however, that a few of the indicators appeared more at some grade levels than at others. Secondary teaching candidates, for example, did not reinforce students for good classroom behavior as frequently as did teaching candidates at either the elementary or junior high level. Secondary teaching candidates provided students with more options to initiate and discuss ideas and activities than did elementary or junior high teaching candidates. Teaching candidates at the elementary and junior high levels used audio-visual resources less effectively and used supplementary materials (materials outside of the textbook) less frequently than did teaching candidates at the secondary level.

Table 7

Frequency of Instrument Indicator Occurrence

Across Grade Levels and Subject Areas

1.0 INSTRUCTIONAL PLANNING

1.1 Planning-Developed in Written Lesson Plans

	Instrument Indicators				
Grades 1-5	1.1.1	1.1.2	1.1.3	1.1.4	1.1.5
Math and Lang. Arts	69%	94%	94%	94%	56%
Other Subject Areas	81%	100%	94%	100%	100%
Avg.	75%	97%	94%	97%	78%
Grades 6-8					
Math and Lang. Arts	88%	94%	94%	94%	88%
Other Subject Areas	58%	83%	58%	75%	67%
Avg.	73%	89%	76%	85%	78%
Grades 9-12					
Math and Lang. Arts	92%	92%	83%	92%	83%
Other Subject Areas	88%	88%	100%	69%	100%
Avg.	90%	90%	92%	81%	92%

Table 7 (Continued)

1.2 Planning Implemented

	Instrument Indicators				
Grades 1-5	1.2.1	1.2.2	1.2.3	1.2.4	1.2.5
Math and Lang. Arts	56%	88%	81%	69%	38%
Other Subject Areas	81%	100%	81%	94%	88%
Avg.	69%	94%	81%	82%	63%
Grades 6-8					
Math and Lang. Arts	81%	94%	81%	94%	81%
Other Subject Areas	50%	75%	58%	67%	50%
Avg.	66%	85%	70%	81%	66%
Grades 9-12					
Math and Lang. Arts	83%	75%	75%	67%	75%
Other Subject Areas	69%	81%	94%	63%	100%
Avg.	76%	78%	85%	65%	88%

Table 7 (Continued)

2.0 CLASSROOM MANAGEMENT AND ORGANIZATION

2.1 Management nstructional Time

	Instrument Indicators				
Grades 1-5	2.1.1	2.1.2	2.1.3	2.1.4	2.1.5
Math and Lang. Arts	100%	95%	95%	95%	80%
Other Subject Areas	94%	100%	81%	100%	88%
Avg.	97%	95%	88%	98%	84%
Grades 6-8					
Math and Lang. Arts	88%	94%	81%	75%	75%
Other Subject Areas	81%	81%	75%	75%	75%
Avg.	85%	88%	78%	75%	75%
Grades 9-12					
Math and Lang. Ar	100%	95%	90%	80%	85%
Other Subject Areas	100%	100%	95%	85%	100%
Avg.	100%	98%	93%	83%	93%

Table 7 (Continued)

2.2 Instructional Organization of the Classroom

	Instrument Indicators				
Grades 1-5	2.2.1	2.2.2	2.2.3	2.2.4	2.2.5
Math and Lang. Arts	95%	95%	100%	55%	70%
Other Subject Areas	100%	88%	94%	38%	13%
Avg.	98%	92%	97%	47%	42%
Grades 6-8					
Math and Lang. Arts	100%	94%	81%	69%	19%
Other Subject Areas	88%	81%	88%	44%	13%
Avg.	94%	88%	85%	57%	16%
Grades 9-12					
Math and Lang. Arts	95%	90%	95%	65%	5%
Other Subject Areas	100%	100%	95%	80%	50%
Avg.	98%	95%	95%	73%	28%

21

Table 7 (Continued)

2.3 Management of Student Conduct

	Instrument Indicators				
Grades 1-5	2.3.1	2.3.2	2.3.3	2.3.4	2.3.5
Math and Lang. Arts	100%	95%	90%	70%	90%
Other Subject Areas	100%	94%	100%	75%	63%
Avg.	100%	95%	95%	73%	77%
Grades 6-8					
Math and Lang. Arts	94%	81%	75%	56%	75%
Other Subject Areas	81%	81%	75%	50%	63%
Avg.	88%	81%	75%	53%	69%
Grades 9-12					
Math and Lang. Arts	80%	90%	80%	50%	80%
Other Subject Areas	90%	100%	80%	25%	85%
Avg.	85%	95%	80%	38%	83%

Table 7 (Continued)

3.0 INSTRUCTIONAL TECHNIQUES AND MATERIALS

3.1 Establishes Learning Set

	Instrument Indicators				
Grades 1-5	3.1.1	3.1.2	3.1.3	3.1.4	3.1.5
Math and Lang. Arts	80%	70%	90%	90%	40%
Other Subject Areas	75%	81%	100%	81%	50%
Avg.	78%	75%	95%	86%	45%
Grades 6-8					
Math and Lang. Arts	75%	75%	100%	63%	44%
Other Subject Areas	69%	50%	81%	63%	25%
Avg.	72%	63%	91%	63%	35%
Grades 9-12					
Math and Lang. Arts	95%	95%	100%	85%	60%
Other Subject Areas	85%	95%	95%	75%	65%
Avg.	90%	95%	98%	80%	63%

22

Table 7 (Continued)

3.2 Lesson Cycle - Instructional Behaviors

	Instrument Indicators				
Grades 1-5	3.2.1	3.2.2	3.2.3	3.2.4	3.2.5
Math and Lang. Arts	95%	95%	95%	90%	90%
Other Subject Areas	81%	94%	88%	100%	88%
Avg.	88%	95%	92%	95%	89%
Grades 6-8					
Math and Lang. Arts	100%	94%	94%	94%	94%
Other Subject Areas	63%	75%	69%	63%	75%
Avg.	82%	85%	82%	79%	85%
Grades 9-12					
Math and Lang. Arts	100%	100%	90%	90%	80%
Other Subject Areas	90%	100%	100%	100%	90%
Avg.	95%	100%	95%	95%	85%

Table 7 (Continued)

3.3 Use of Instructional Material

	Instrument Indicators				
Grades 1-5	3.3.1	3.3.2	3.3.3	3.3.4	3.3.5
Math and Lang. Arts	90%	80%	45%	65%	95%
Other Subject Areas	94%	81%	56%	69%	81%
Avg.	92%	81%	51%	67%	88%
Grades 6-8					
Math and Lang. Arts	100%	81%	43%	56%	56%
Other Subject Areas	81%	56%	63%	69%	75%
Avg.	91%	69%	53%	63%	66%
Grades 9-12					
Math and Lang. Arts	90%	75%	75%	70%	60%
Other Subject Areas	100%	75%	75%	85%	100%
Avg.	95%	75%	75%	78%	80%

23

Table 7 (Continued)

4.0 COMMUNICATION AND PRESENTATION TECHNIQUES

	Instrument	Indicators			
Grades 1-5	4.1	4.2	4.3	4.4	4.5
Math and Lang. Arts	90%	95%	60%	75%	70%
Other Subject Areas	94%	94%	88%	81%	81%
Avg.	92%	95%	74%	78%	76%
Grades 6-8					
Math and Lang. Arts	100%	100%	81%	100%	69%
Other Subject Areas	75%	75%	63%	50%	63%
Avg.	88%	88%	72%	75%	66%
Grades 9-12					
Math and Lang. Arts	100%	100%	85%	65%	80%
Other Subject Areas	100%	100%	80%	80%	85%
Avg.	100%	100%	83%	73%	83%

Table 7 (Continued)

5.0 MOTIVATION FOR LEARNING

	Instrument	Indicators			
Grades 1-5	5.1	5.2	5.3	5.4	5.5
Math and Lang. Arts	25%	60%	95%	85%	80%
Other Subject Areas	31%	75%	94%	94%	81%
Avg.	28%	68%	95%	90%	81%
Grades 6-8					
Math and Lang. Arts	38%	63%	94%	94%	75%
Other Subject Areas	19%	50%	56%	56%	31%
Avg.	29%	57%	75%	75%	53%
Grades 9-12					
Math and Lang. Arts	5%	95%	90%	80%	70%
Other Subject Areas	40%	80%	80%	70%	50%
Avg.	23%	88%	85%	75%	60%

24

Table 7 (Continued)

6.0 ACADEMIC ASSESSMENT OF STUDENTS

	Instrument Indicators	
	6.1	6.2
Grades 1-5		
Math and Lang. Arts	94%	81%
Other Subject Areas	100%	94%
Avg.	97%	88%
Grades 6-8		
Math and Lang. Arts	100%	100%
Other Subject Areas	83%	75%
Avg.	92%	88%
Grades 9-12		
Math and Lang. Arts	100%	83%
Other Subject Areas	100%	94%
Avg.	100%	89%

Frequency of occurrence of a few of the instrument indicators was consistently low across subject area categories and grade levels. The teaching candidates did not arrange their classrooms to accommodate a variety of instructional purposes. Their classroom arrangements were inflexible and static throughout the semester. Teaching candidates also did not provide learners with instructional objectives in the set inductions of their lessons. Another indicator that was conistently low in frequency was the provision of a review or summary of main points in the conclusion of the lesson. These teaching candidates also infrequently provided students with options during their lessons.

Results did indicate, however, that these 27 teaching candidates were proficient in exhibiting several competency skills. The teaching candidates appeared to have been good instructional planners; good managers of the utilization of instructional time; effective organizers of instruction in that materials and equipment were ready at the start of the lesson and the seating/grouping arrangements were appropriate for instructional activities; and good managers of student conduct in that classroom management procedures were established and administered and student conduct was monitored. These teaching candidates used correct syntax and spelling in oral and written discourse, communicated enthusiasm, and demonstrated appropriate nonverbal behaviors.

Overall, teaching candidates were also effective in utilizing the instructional techniques in the Lesson Cycle - Instructional Behaviors subdomain. Accurate stimulus information was presented, key features of the lessons were pointed out, guided practice and independent practice opportunities were provided, clear directions for assignments were given, and feedback concerning student performance was provided. The teaching candidates' classrooms were comfortable atmospheres for learning and

teaching candidates personalized instruction by recognizing individual differences of students. Frequency percentages also indicated that these teaching candidates were proficient in providing assessment of student learning, either formal or informal, that was appropriate for the lesson content.

CONCLUSIONS AND RECOMMENDATIONS

Collectively, the results to the research questions suggest that COAST is a generic observation instrument that can be used across subject areas and grade levels. The instrument has a high degree of content validity, and it can be used by observers with a high degree of accuracy and consistency. Based on the frequency of indicator occurrence across the small number of subject areas and grade levels represented in the study, the instrument appears to be generalizable across different contextual conditions. This conclusion requires further testing as some of the indicators were rather low in frequency. Further testing might determine if this low frequency was due to the proficiency level of the teaching candidates participating in the study, or the inappropriateness of the indicators on a generic observation instrument.

The findings of the study also indicate that the instrument can be used to evaluate lessons that fit the direct-instruction model of teaching. Forms completed by the observers concerning the grouping structures and teaching methods utilized in the observed lessons indicated that out of the total 108 lessons observed, 107 lessons utilized a whole-class grouping structure for at least part of the lesson. Eight lessons utilized a small-group structure that was teacher-led, and seven of these eight lessons were in elementary reading group situations. There were no lessons observed that incorporated student-to-student interaction or allowed students to work

together to complete instructional tasks. Most of these lessons appeared to have incorporated the direct instruction model of teaching which limits the types of instructional strategies observed during this study.

The findings of the study do indicate that COAST may be suitable for inclusion in teacher preparation programs. Because COAST is comprised of behaviors identified by research as effective and because these behaviors are often included on observation instruments that are used to evaluate inservice teachers, an instrument such as COAST might be valuable in the assessment of teaching candidates.

COAST could be used in a teacher preparation program to improve teaching candidate performance. Teaching candidates could first be pre-tested over the behaviors incorporated in COAST, then provided in-depth instruction on each of the instrument indicators, then assessed in actual classrooms with COAST to determine if teaching performance improved as a result of the training. Results obtained by using COAST could be used to diagnose specific areas of teaching that need to be remediated in the teaching candidate's preparation.

Several recommendations for further study are made. First, this study should be replicated in other situations to increase the total number of teaching candidates observed with COAST. By having larger numbers of teaching candidates observed, the generalizability of the instrument could be further verified. Second, replication studies should focus on specific content areas to determine how well teaching candidates from just one subject area compare in their evaluations. This would provide more information concerning the appropriateness of instrument indicators for evaluating teacher performance in specific subject areas. Third, a replication study should be conducted that extends the number of subject

areas and grade levels represented. The number of teaching candidates at different grade levels and subject areas, as well as specific grade levels and subject areas, were limited in this study, and by extending these more information should be obtained concerning the external validity and generalizability of the instrument. And fourth, a replication study should be conducted that addresses a common criticism raised about the ability of observation instruments to adequately and fairly assess teacher performance when different instructional strategies are utilized. A deliberate attempt should be made in a replication study to assess a wide range of instructional strategies to determine if significantly different ratings will be obtained from those obtained in direct instruction lessons.

References

Anderson, L., Evertson, C., & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. Elementary School Journal, 79, 193-223.

Florida Performance Measurement System. (1983). Coalition for the Development of Performance Evaluation System, Office of Teacher Education: Certification & Staff Development: Tallahassee, FA.

Frick, T., & Semmel, M.I. (1978). Observer agreement and reliabilities of classroom observational measures. Review of Educational Research, 48(1), 157-184.

Georgia Department of Education. (1984). Teacher Performance Assessment Instrument. Georgia Department of Education: Division of Staff Development.

Good, T.L., & Grouws, D. (1977). Teaching effects: A process-product study in fourth grade classrooms. Journal of Teacher Education, 28, 49-54.

Knapp, M.S. (1982). Toward the study of teacher evaluation as an organizational process: A review of current research and practice. Educational and Human Services, Research Center, SRI International, Menlo Park, CA.

Medley, D.M., & Norton, D.P. (1971). The concept of reliability as it applies to behavior records. Paper presented at the meeting of the American Psychological Association, Washington, D.C.

Scott, W.A. (1955). Reliability of content analysis: The case of nominal case coding. Public Opinion Quarterly, 19, 321-325.

Texas Teacher Appraisal Instrument. (198[illegible]. Texas Education Agency: Austin, TX.

INSTRUMENT INDICATORS OF COAST

1.0 INSTRUCTIONAL PLANNING

1.1 Developed in Lesson Plans

1.1.1 lesson plans specify the set induction of the lesson
1.1.2 lesson plans specify the instructional materials to be used
1.1.3 lesson plans specify the guided practice techniques to be used
1.1.4 lesson plans specify the independent practice techniques
1.1.5 lesson plans specify the assessment techniques to be used

1.2 Implemented as Planned

1.2.1 the set induction of the lesson is implemented as planned
1.2.2 specified instructional materials are used during lesson
1.2.3 specified guided practice techniques are used during lesson
1.2.4 specified independent practice techniques are used
1.2.5 specified assessment techniques arae implemented during lesson

2.0 CLASSROOM MANAGEMENT AND ORGANIZATION

2.1 Management of Instrucitonal Time

2.1.1 begins class promptly
2.1.2 engages students in learning activities for a majority of the available classtime
2.1.3 routines are established for recurring activities
2.1.4 provides sufficient time for students to complete instructional activities
2.1.5 makes smooth transitions between activities

2.2 Instructional Organization of the Classroom

2.2.1 instructional materials are ready for use at the start of the lesson
2.2.2 instructional materials and equipment are ready for use at the start of the lesson
2.2.3 arrangements for seating/grouping are appropriate for instructional activities
2.2.4 schedules for assignments are made available to students
2.2.5 arranges room for a variety of instructional purposes

2.3 Management of Student Conduct

2.3.1 establishes classroom management procedures appropriate for the students in the classroom
2.3.2 administers established rules and procedures for classroom behavior
2.3.3 monitors student conduct
2.3.4 reinforces appropriate classroom behavior
2.3.5 establishes physical arrangement of classroom to promote good classroom behavior

3.0 INSTRUCTIONAL TECHNIQUES AND MATERIALS

3.1 Establishes Learning Set

3.1.1 gains attention of students to begin instruction
3.1.2 provides anticipatory set/gives focus of lesson
3.1.3 links lesson to previous experiences
3.1.4 determines starting points of students
3.1.5 provides lesson objectives to students

3.2 Lesson Cycle - Instructional Behaviors

3.2.1 presents stimulus information and materials
3.2.2 points out key features of new information
3.2.3 provides for guided practice of new information by students
3.2.4 provides for independent practice of new information by students
3.2.5 gives clear directions for seatwork/homework assignments
3.2.6 indicates expected completion time for assignments
3.2.7 circulates and provides assistance to students
3.2.8 provides feedback concerning student performance
3.2.9 reteaches/remediates during lesson if necessary
3.2.10 reviews/summarizes information in closure of lesson

3.3 Use of instructional Material

3.3.1 uses materials appropriate for the attainment of lesson objectives
3.3.2 provides explanations and directions to students for use of the material
3.3.3 uses audio-visual resources effectively
3.3.4 uses supplementary material/goes beyond text
3.3.5 provides sufficient quantities of materials for the class

4.0 COMMUNICATION AND PRESENTATION TECHNIQUES

4.1 uses correct syntax in oral discourse
4.2 uses correct syntax and spelling in written materials
4.3 communicates enthusiasm for the subject matter
4.4 varies mode of presentation
4.5 demonstrates appropriate nonverbal behaviors

5.0 MOTIVATION FOR LEARNING

5.1 provides options to students during lesson/allows students to take responsibility for their lesson
5.2 provides opportunities for students to initiate ideas, discussion, and/or activities
5.3 creates a relaxed, comfortable atmosphere for learning
5.4 shows appreciation for individual differences
5.5 provides opportunities for students to share ideas and work efforts

6.0 ACADEMIC ASSESSMENT AND EVALUATION TECHNIQUES

6.1 evaluates student understanding either formally or informally during the lesson
6.2 evaluation technqiues used are appropriate for the content of the lesson