DOCUMENT RESUME

ED 247 266 TM 840 437

AUTHOR Rose, Janet S.; Huynh, Huynh
TITLE Technical Issues in Adopting the APT for Districtwide Teacher Evaluation.
PUB DATE Apr 84
NOTE 34p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 24-26, 1984).
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; School Districts; *Teacher Evaluation; Test Bias; *Test Reliability
IDENTIFIERS *Assessments of Performance in Teaching; *Charleston County School District SC

ABSTRACT

As part of a new teacher evaluation program initiated by the local school board, the Charleston County School District (South Carolina) adopted the Assessments of Performance in Teaching (APT) as a major evaluation tool to assess the teaching performance of annual contract teachers. Since evaluation procedures can ultimately lead to teacher dismissal, it was incumbent upon the district staff to ensure the appropriateness of the APT and its technical quality for a population of teachers wider than those for whom the instrument was designed. A study was conducted on approximately 250 teachers to examine the inter-observation and inter-rater reliability of the APT for various groups of teachers: special education teachers, Chapter 1 teachers, elementary, middle and high school teachers, black teachers and white teachers. Agreement indices were calculated for individual items to identify teacher behaviors which reduced reliability and for which observers need additional training and practice. Other local concerns addressed by the study focused on differences in the ratings of principals versus district staff and ratings of observers evaluating teachers within their own field of certification versus observers evaluating teachers in fields outside their own. (Author)

TECHNICAL ISSUES IN ADOPTING THE APT FOR DISTRICTWIDE TEACHER EVALUATION

Janet S. Rose
Charleston County (S.C.) School District

Huynh Huynh
University of South Carolina

NCME Symposium: Teacher Performance Assessment: An Examination of Technical Issues from an Employment Decision Context

New Orleans
April, 1984

Symposium: Teacher Performance Assessment: An Examination of Technical Issues from an Employment Decision Context

Presentation: Technical Issues in Adopting the APT for Districtwide Teacher Evaluation

Authors: Janet S. Rose, Charleston County (S.C.) School District
Huynh Huynh, University of South Carolina

As part of a new teacher evaluation program initiated by the local school board, the Charleston County School District adopted the APT as a major evaluation tool to assess the teaching performance of annual contract teachers. Since evaluation procedures can ultimately lead to teacher dismissal, it was incumbent upon the district staff to ensure the appropriateness of the APT and its technical quality for a population of teachers wider than those for whom the instrument was designed.

A study was conducted on approximately 250 teachers to examine the inter-observation and inter-rater reliability of the APT for various groups of teachers: special education teachers, Chapter I teachers, elementary, middle and high school teachers, black teachers and white teachers. Agreement indices were calculated for individual items to identify teacher behaviors which reduced reliability and for which observers need additional training and practice. Other local concerns addressed by the study focused on differences in the ratings of principals versus district staff and ratings of observers evaluating teachers within their own field of certification versus observers evaluating teachers in fields outside their own.

Responding to the public's outcry for educational accountability, the Board of Trustees of the Charleston County School District (CCSD) enacted a new policy on teacher evaluation in June of 1982. The intent of this policy was to "strengthen" evaluation practices in such a way that incompetent teachers would be identified and either remediated, or if remediation failed, dismissed from the school system. Board members became attuned to the need for a change in current teacher evaluation practices upon hearing of incidents which caused them and the commmunity to question the quality of instruction students were receiving in the classroom. To assure the community that poor teachers would no longer be permitted to continue being employed in the schools, they decided to replace the current program with one that could be used to remove teachers who lacked basic teaching competencies. This action altered the focus of teacher evaluation in Charleston County. Whereas the older program was based upon a model of clinical supervision and assumed not only competence but also the need for all teachers, regardless of their level of competence, to prepare and implement improvement plans, the new program, upon request from the School Board, was designed to determine whether teachers *were* competent, i.e., whether they possessed skills important for successful performance as a teacher.

The School Board approached this task logically and recommended to the Superintendent that experienced teachers (i.e., those with continuing contracts) meet, at minimum, South Carolina's new requirements for beginning teachers. They proposed that the instrument developed under Act 187, the "Assessments of Performance in Teaching" or APT, be administered to all teachers in Charleston County. Administration of the APT became one component of the new teacher evaluation program. The APT is described in its manual as follows:

> The APT instrument is divided into five Performance Dimensions. Each Performance Dimension is measured through eight to eleven statements. The observation statements are dichotomous (yes/no) decisions that indicate whether or not a teaching skill or behavior was demonstrated during the observation. Specific evidence in the form of a statement citing one or more incidents describing the

> demonstration is required. Each Performance Dimension is an essential area of teaching competence and each must be satisfactorily demonstrated to successfully complete the APT. The assessment is the composite of three observers' ratings (APT Manual, page 1).

Recent litigation in the area of teacher evaluation forced district staff to consider several issues regarding use of the APT, or any evaluation procedure, on experienced teachers, under conditions where the results of that procedure could be used to make critical employment decisions. First, the courts have acknowledged that teachers given tenure or continuing contract status are presumed competent and have legitimate expectations of future employment. Consequently, they have protected property interests and can insist upon due process procedures. Evidence for removing experienced teachers from the system must be rather persuasive, and the burden of proof is upon school authorities. Second, if more minority teachers are terminated as a result of evaluation practices, the school district must demonstrate that evaluation criteria are non-discriminatory and related to job performance. And, finally, it is essential to the validity of the instrument that users demonstrate that it yields objective data and can be applied to all teachers in a consistent or standardized manner. Failure to use an evaluation instrument for which there is sufficient evidence of reliability, as Allen and Jarvis (1983) warn, can have significant legal ramifications.

Since the S.C. State Department of Education developed the APT for beginning teachers (i.e., those with provisional contracts), it was necessary to investigate more thoroughly the extent to which the APT could be applied accurately and fairly to experienced teachers by the 130 observers trained to assist in this evaluation effort. Educators from various segments of the district expressed their concerns regarding the reliability of the APT. Trained observers expressed concern over training and practice in using the APT instrument. District staff, principals and assistant principals were trained

via videotapes in the administration of the APT and had met the criteria to become certified as "endorsed observers," but they did not have sufficient practice, if any, observing teachers in the field. Would the reliability demonstrated under controlled "videotaped" conditions be maintained in real-life teaching situations? The School Board shared this concern and had questions about the reliability of ratings gathered by principals. Board members wanted to know whether principals would tend to be more lenient than district staff in their observations of teachers employed in their own school. From a similar perspective, district staff were concerned about differences between observers rating teachers in their own field of expertise and those rating teachers outside their field. Teachers voiced their opinions regarding the observation of teachers in unique types of teaching arrangements, such as special education and Chapter I teachers. Could these teachers be evaluated objectively and fairly using an instrument which was applied universally to all teachers in the district? Finally, there were accusations made regarding the discriminatory nature of the instrument. Some teachers alleged that evaluation would result in the dismissal of proportionally more black teachers and erroneously concluded that, if this indeed occurred, the instrument would be biased against minority teachers.

To protect the school district and the rights of teachers, the Evaluation and Research (E&R) office recommended that a study be conducted to address some of the concerns raised by teachers, administrators and School Board members. The School Board agreed to the study and to delay until the 1983-84 school year use of APT data in making decisions regarding the employment status of experienced teachers. (This decision coincided with the State Department's recommendation to delay the application of the APT to decisions regarding the employment of beginning teachers.) Due to the limited availability of funds and human resources, the study was designed to

respond only to reliability issues, though at the same time E&R would be able to produce for the Office of Teacher Evaluation and Staff Development estimates of the percentage of teachers who would fall below the state minimum standard and descriptions of teachers' performance on individual competencies and the five Performance Dimensions. Although local educators questioned the extent to which some of the 51 competencies were necessary for successful performance as a teacher, validity issues were excluded from the study and left to the State Department to tackle.

Questions addressed by CCSD's Reliability Study were as follows:

1. What are the score distributions for the APT total score and the five Performance Dimensions or PDs for teacher subpopulations and racial groups? What percentage of teachers fall below the state's minimum standard on the total score?
2. What percentage of teachers demonstrate each of the 51 teaching competencies?
3. Is the APT instrument reliable for concerned sub-populations of teachers? (I.e., do the ratings of two observers evaluating the behavior of a teacher at the same time agree? Is the intra-observer variability sufficiently small to yield reliable ratings?) Which competencies appear to contribute to measurement error?
4. Are there differences between the ratings of principals and district staff; that is, does one group tend to score teachers lower than the other?
5. Do district staff observing teachers within their own field of expertise rate teachers differently than staff observing teachers outside their field? This question was asked only for middle and high school teachers.

Evertson and Holley (1981) remind us of the three causes of unreliability: (a) unstable phenomena being observed; (b) disagreement between observers about what they see occur; and (c) inconsistency in the way the instrument measures

teacher behavior. Though we would not be able to distinguish among the causes of measurement error, if found, we would at least be able to identify where problems exist.

Overview of Study

A study of the APT was conducted during the 1982-83 school year on samples of teachers from various teacher populations - special education, Chapter I and "regular" teachers (elementary, middle and high). A group of beginning teachers was included for comparative purposes only. Analyses were performed for these groups as well as for racial groups. Different sets of observation patterns or "schemes" were used for teacher samples. These were counterbalanced and used with equal frequency as much as possible. Observers were classified as Principals (principals and assistant principals) or District Staff (central staff and area superintendents). A third category, "Other," consisting of Principals, District Staff and peer teachers who had been endorsed as observers, was necessary for some parts of the study. Observers were then assigned to three APT observations which occurred during a two-week interval.

Scheduling and implementation of the study were executed by the Teacher Evaluation staff under the guidance of the Evaluation and Research office. Observation designs, data entry and data analysis were contracted to Dr. Huynh Huynh of the University of South Carolina.

Study Designs

Regular teachers. Ninety teachers (54 elementary, 18 middle and 18 high school teachers) were selected to participate in the study. A stratified random sampling procedure was used to ensure that teachers participating in the study were representative of all teachers in Charleston County, with the exception of special education and Chapter I teachers. The three types of observers (Principals, District Staff and Others) were paired according to the three possible combinations, and pairs were assigned to either the first, second or third

observation, according to the same six observation schemes used in the State Department's study. (See Table 1.) Each observer was assigned to two teachers. Fifteen teachers were assigned to each scheme. At the middle and high school levels District Staff observers were assigned to one teacher in the same field and one teacher in a different field than themselves.

Special education teachers. Eighteen pairs of Principals and District Staff (special education consultants) were assigned to three teachers, yielding a total of 54 special education teachers participating in the study. A stratified random sampling procedure was used to select teachers representative of resource and self-contained placements and the various handicapping conditions which reflected the composition of the teaching population. The first two observations were conducted individually, the first by either the Principal (scheme 1) or District Staff (scheme 2), while for third observation both observers were present. The first two observations were conducted on Tuesday and Thursday of the first week, and the third observation was conducted on Thursday of the following week. Half the observer-pairs observed two teachers according to scheme 1 and one teacher according to scheme 2, while the other half observed one teacher according to scheme 1 and two teachers according to scheme 2. (See Table 2.)

Chapter I teachers. Twelve language arts and 12 math teachers were selected for the study. Pairs of Principal/District Staff observers were assigned to two teachers each. Teachers were observed according to the two observation schemes used for special education teachers. Each observer-pair observed one teacher according to each scheme. (See Table 3.)

Beginning teachers. Forty eight beginning teachers were selected for the study and were observed according to the six observations schemes used for regular teachers. Eight teachers were observed according to each scheme.

Observers were assigned to one teacher only.

Description of Participating Teachers

Although we anticipated that a total of 216 teachers (118 experienced and 48 new) would be participating in all studies, the actual number of teachers observed was 214. Two teachers (one regular and one beginning) had incomplete observation data. The number of teachers from each group and the number of observations are given in Table 4. Table 5 gives a breakdown of the sample by race, sex, age group and education.

Presentation of Results

Since primary interest focuses on experienced teachers as a group and on differences among regular, special education and Chapter I teachers and between black and white teachers, the results of the reliability study are presented for groups selected from the list below according to the questions asked in the study:

a. Beginning teachers (47 teachers, 282 observations)

b. All regular teachers (elementary, middle, high) (89 teachers, 534 observations)

c. Regular black teachers (36 teachers, 216 observations)

d. Regular white teachers (51 teachers, 307 observations)

e. Special education teachers (54 teachers, 216 observations)

f. Chapter I teachers (24 teachers, 96 observations)

The unit of analysis was observations, rather than teachers.

Results

Question 1: Score Distributions and Percentages of Teachers Below Standard

APT scores below standard. Table 6 presents the percentage of total APT scores below the state standard of 44 (out of 51 competencies). These figures were used to estimate the percentage of teachers who were expected to score less

than 44 on the APT administered in 1983-84. Beginning teachers had the most scores below standard (26.6%). A higher percentage of black teachers scored below standard, compared with white teachers. Special education and Chapter I teachers obtained the highest APT scores.

APT score distributions. Table 7 contains the frequency distributions for total APT scores. Cumulative percentages are provided for each teacher group. Very few teachers obtained scores of 40 or below.

PD score distributions. The number of competencies (and score range) for the five Performance Dimensions are: 8 for Planning; 11 for Instruction, Management and Communication; and 10 for Attitude. Cumulative percentages of scores on the PDs are located in Table 8. The most noticeable characteristic of the frequency distributions is the lack of variability of the scores. For example, only 16.5% of the scores for Management obtained by all regular teachers were 10 or less of a possible 11 points, while 83.5% were 11. For the other PDs, about half the teachers demonstrated all competencies, and the other half demonstrated all but one. Very few teachers failed to demonstrate two or more competencies within a particular PD. This lack of variability was not a surprise, considering that the APT assesses basic teaching competencies.

Question 2: Percentage of Teachers Who Demonstrated Each of the 51 Competencies

The percentage of teachers who demonstrated each of the 51 competencies is the percentage of observation sheets on which the item was coded as demonstrated. Table 9 presents this information. Below is a summary of the competencies demonstrated by fewer than 75% of the teachers (noted by "X"):

Competency	Beg.	Reg. Tot.	Reg. Blk.	Reg. Wht.	Spec. Ed.	Chapter I
PD 1: PLANNING						
f. differences planned	X	X	X			
g. objectives assessed	X		X		X	X
h. progress provided	X					

Competency	Beg.	Reg. Tot.	Reg. Blk.	Reg. Wht.	Spec. Ed.	Chapter I
PD 2: INSTRUCTION						
c. needs accomodated	X	X	X			
PD 3: MANAGEMENT						
(none)						
PD 4: COMMUNICATION						
i. written communication	X	X	X	X	X	X
PD 5: ATTITUDE						
d. learning personalized	X	X	X			X
e. value communicated	X	X	X	X	X	X
f. humor acknowledged	X	X	X	X	X	X

Question 3: Reliability of the APT

Indented sections have been extracted verbatim from a memorandum from Huynh (1983) summarizing the results of the study.

> Table 10 reports the inter-observation and inter-rater reliabilities for the various teacher and race groups. For this table, the index of inter-observation reliability was taken as the correlation between the two total APT scores assigned on two occasions by the same observer. For beginning and regular teachers, there were three observers *(for a total of 267 score pairs and 141 score pairs, respectively)*. As for special education and Chapter I teachers, there were only two observers *(for a total of 108 score pairs and 48 score pairs, respectively)*.
>
> For the two groups of beginning and regular teachers, there were 12 indices of inter-rater reliability. Each index was represented by the correlations between the two total APT scores assigned by two different observers. The term average inter-rater reliability of Table 10 denotes the average of these correlations (Huynh, pages 4-5).

Inter-observation reliability. Inter-observation reliability is an index of the degree to which an observer rates a teacher consistently from one observation session to the next. These indices are expected to be high, though not perfect, due to minor variations and true inconsistences in a teacher's behaviors from one day to the next. The indices listed in Table 10 are moderately high (greater than .60), with a few exceptions. Highest reliability was found for special education teachers and lowest for Chapter I teachers. Principals were most consistent in their ratings of regular teachers and least consistent observing beginning teachers. District staff were most consistent in their

ratings of special education teachers and least consistent with regard to Chapter I teachers.

Inter-rater reliability. Inter-rater reliability is an index of agreement between the total scores assigned by two observers simultaneously rating behaviors of the same teachers. Indices for beginning and regular teachers were .57, while other indices were .69 for special education teachers and .30 for Chapter I teachers. A difference of .05 was found between indices for black and white regular and beginning teachers. Though the difference is minor, the lower index for black teachers may be attributed to the significantly fewer black teachers observed.

The Reliability Training Program developed to train and certify APT observers sets a minimally acceptable reliability standard of .80. None of the reliability estimates obtained in this study reached that figure. The estimates for Chapter I teachers are extremely low compared to those found for other teacher groups and suggests problems with using the APT for this group of teachers without some further investigation.

The overall reliability of the total APT scores and the associated standard error of measurement (SEM) are documented in Table 11.

> In this table, the standard deviation (SD) was obtained by combining all total APT scores for each teacher group in one sample. For each teacher group, reliability was taken as the average of all the inter-observation and inter-rater correlations. (For beginning and regular teachers, there are 15 such correlations. As for special education and Chapter I teachers, these correlations number at 6.) The standard error of measurement was computed via the formula
>
> $$SEM = SD\ (1 - \text{reliability})^{\frac{1}{2}}.$$
>
> Table 11 also reports the reliability and standard error of measurement for all teachers. The overall reliability (.589) was derived by taking the weighted average of the reliabilities of the four teacher groups with each reliability weighted by the number of teachers in the group. The overall standard error of measurement (2.19) was computed via the formula listed in the last paragraph (Huynh, page 5.)

Based upon the results, the standard error of measurement can be estimated at two for the APT. The overall reliabilities again identify

potential problems with the use of the APT for Chapter I teachers.

Item reliabilities. The extent to which observers agreed on their ratings of individuals items is documented in Table 12.

> To combine all the data for the purpose of examining the reliability of each item (skill), the observations made by the category "Other" ware deleted from the two groups of beginning and regular teachers. Thus, in the combined data, there were four observations made on each item for each teacher. Each observation was coded as 0 (no evidence of the skill) or 1 (evidence of the skill).
>
> For each item, the reliability was taken as the percentage of times in which two separate observations made by the category "Other" were both zero or one. Thus the item reliability was taken as the raw agreement index taken over the observers and for the group of teachers under consideration.
>
> ...In the interpretation of the item reliability, please note that its chance level is .50. This level will occur if all observers randomly assigned their scores to the items (Huynh, page 5).

In general, raw agreement was lowest for competencies demonstrated by fewer than 75% of the teachers. No doubt the greater variability in the degree to which teachers demonstrated these behaviors contributed to the lower agreement indices.

Question 4: Differences Between Principals and District Observers

Table 13 lists the mean and standard deviation and the percentage of cases below 44 for the total APT scores assigned by the principals and district observers for each teacher group.

> Overall, the mean difference between the total APT scores assigned by the principals and by the district observers was .35 on the 51-score APT scale. In terms of the percentage of observations below the state passing score of 44, the difference between the two groups of observers was one percent. Judging from both the mean and the percent of cases below 44, the data indicated that district observers tended to score lower (be "harder") than the principals when all teacher groups were combined. This trend, however, was not consistent across the four individual teacher groups (Huynh, page 8).

Question 5: Differences Between In-Field and Out-of-Field Observers

Below are listed the mean, standard deviation and percentage of cases below 44 for the total APT scores assigned by the in-field and out-of-field district observers. The data were compiled from the group of middle and high school teachers.

Observer	Number of Teachers	Number of Observations	Mean	S.D.	Percent below 44
In-field	18	36	47.06	3.22	11
Out-of-Field	18	36	46.36	4.04	22

The data indicate that district staff observing teachers within their own field assigned higher scores (thus failing less teachers) than when observing teachers outside their own field. Although the mean scores are similar, out-of-field observers failed an additional four of the 36 observations.

Conclusions

The reliability indices found in this study were much lower than the index of .80 used to endorse APT observers. Data comparing raw agreement indices with percentages of teachers demonstrating each competency suggest that the inter- and intra-rater reliability of the APT observations would be much lower if there were greater variability in teachers' performance on the APT. We can also project that the APT would be more reliable for high-scoring teachers and less reliable for low-scoring teachers. This trend has strong implications for use of the APT in employment decisions, since teachers scoring below the passing standard will be those considered for dismissal from the system. It would be wise, therefore, to exercise caution in using APT scores for summative decision-making without either demonstrating the reliability of APT scores for targeted teachers or accumulating additional

evidence of incompetency.

Data comparing different categories of observers (i.e., Principals vs. District Observers; In-Field vs. Out-of-Field Observers) show minor differences between ratings. However, the question of whether or not observers would be less reliable when assessing teachers on future occasions or without another observer present should not be ignored in this particular situation where observers, mostly principals, are forced into conflicting evaluative and supportive roles. One way for principals to reconcile their new evaluative function with their well-established and well-accepted supportive function is to be more lenient in their ratings, i.e., when in doubt give teachers credit for demonstrating a particular competency. In fact, preliminary data on CCSD's 1983-84 teacher evaluation program indicate that the distribution of APT scores is much more negatively skewed than last year's. There is also evidence that a substantial minority of observers have "favorite competencies" and are more likely to deny credit for them to a greater degree than other behaviors.

The new wave of accountability, coupled with the expanding literature on teacher and school effectiveness, will encourage more and more states and school districts to evaluate teacher performance through classroom observational techniques. These assessment procedures, though not new to educational researchers, are quite novel to school principals, the principle evaluators of teachers. Not only must these individuals deal with role conflicts, but they also must learn, practice and perfect a new method of teacher evaluation and use observation procedures in a consistent and reliable manner. When teacher evaluation is based upon a high-inference rating system, such as the APT, which requires a greater amount of interpretation compared with low-inference measures, it is critical that users of observation instruments mandate that

observers successfully participate in a reliability training program. In addition, users should also: (a) allow sufficient lead time before implementing the evaluation system so that observers can practice their observation skills; (b) continue to periodically collect data on rater reliability after preliminary studies have been completed; and (c) redefine and re-clarify descriptions of teacher behaviors and competencies contained on the observation instrument to reduce, as much as possible, subjectivity of the instrument, thereby increasing rater reliability.

References

Allen, K. H., and Jarvis, M. E. (1983). Analogizing teacher evaluation policies and procedures with case law. Paper presented at the meeting of the American Educational Research Association, Montreal.

Assessments of performance in teaching observation instrument. Columbia, S. C.: South Carolina State Department of Education.

Evertson, C. M. and Holley, F. M. (1981). Classroom observation. In J. Millman (Ed.), Handbook of teacher observation. Beverly Hills: Sage Publications, Inc.

Huynh, H. (1983). Summary results of APT reliability study. Columbia, S. C.: University of South Carolina, College of Education.

Table 1

APT Reliability Observation Schemes for

Experienced and Beginning Teachers

	Observation		
Scheme	First	Second	Third
1	Principal Other	Principal District Staff	District Staff Other
2	Principal District Staff	Principal Other	District Staff Other
3	Principal District Staff	District Staff Other	Principal Other
4	Principal Other	District Staff Other	Principal District Staff
5	District Staff Other	Principal District Staff	Principal Other
6	District Staff Other	Principal Other	Principal District Staff

Table 2

APT Reliability Observation Patterns for

Special Education Teachers

Observation Pattern	Teacher	Scheme	Observation		
			First	Second	Third
A	1	1	Principal	District	Principal/District
	2	2	District	Principal	Principal/District
	3	1	Principal	District	Principal/District
B	1	2	District	Principal	Principal/District
	2	1	Principal	District	Principal/District
	3	2	District	Principal	Principal/District

Table 3

APT Reliability Observation Schemes for

Chapter I Teachers

	Observation		
Scheme	First	Second	Third
1	Principal	District	Principal/District
2	District	Principal	Principal/District

Table 4

Sample Description

Group	Number of Teachers	Number of Observations Per Teacher	Total Number of Observations
Regular	89	6	534
Special Education	54	4	216
Chapter I	24	4	96
Total Experienced	167		846
Beginning	47	6	282
TOTAL	216		1,128

Table 5

Description of Participating Teachers: Number and Percentage (Within Teacher Group) According to Biographical Variables

Biographical Variable	Regular #	Regular %	Spec. Ed. #	Spec. Ed. %	Chapt. I #	Chapt. I %	TOTAL EXPERIENCED #	TOTAL EXPERIENCED %	BEGINNING #	BEGINNING %
Race										
Black	36	41%	12	22%	15	63%	63	38%	1	2%
White	51	59%	40	74%	4	17%	95	57%	43	91%
No Data	2	2%	2	4%	5	21%	9	5%	3	6%
Sex										
Male	14	16%	7	13%	1	4%	22	13%	7	15%
Female	75	84%	46	85%	22	92%	143	86%	38	81%
No Data	0	-	1	2%	1	4%	2	1%	2	4%
Age										
20-25	3	3%	9	17%	1	4%	12	7%	24	51%
26-30	14	16%	16	30%	1	4%	32	19%	10	21%
31-40	38	43%	21	39%	9	38%	68	41%	7	15%
41-50	23	26%	2	4%	1	4%	26	16%	3	6%
51 or more	11	12%	5	9%	9	38%	25	15%	0	-
No Data	0	-	1	2%	3	13%	4	2%	3	6%
Education										
Bachelor Degree	50	56%	26	48%	14	58%	90	54%	36	77%
Master Degree	30	34%	26	48%	6	25%	62	37%	8	17%
Master Deg. & 30 hrs	5	6%	0	-	3	13%	8	5%	0	-
Doctorate	1	1%	2	4%	0	-	3	2%	0	-
Bus., Cler., Voca.	2	2%	0	-	0	-	2	2%	1	2%
Other	1	1%	0	-	0	-	1	.5%	0	-
No Data	0	-	0	-	1	4%	1	.5%	2	4%

Table 6

Percentage of Total APT Scores Below Revised State Standard (44)

Teacher Group	ALL TEACHERS* No. Tchrs.	No. Obs.	%	BLACK TEACHERS No. Tchrs.	No. Obs.	%	WHITE TEACHERS No. Tchrs.	No. Obs.	%
Beginning Teachers	47	282	26.6	1	6	33.3**	43	258	27.5
Experienced Teachers									
Regular	89	534	14.2	36	216	21.3	51	306	9.2
Special Education	54	216	7.9	12	48	10.4	40	160	7.5
Chapter I	24	96	5.2	15	60	6.7	4	14	6.3
TOTAL	167	846	11.9	63	324	17.0	95	480	8.5

*Due to missing data on race, the number of white teachers and the number of black teachers do not add up to the total number of teachers.

**Based on only teacher (with six APT scores).

Table 7

Frequency Distributions for Total APT Scores:

Cumulative Percentages

Total Score	Regular Total	Regular Black	Regular White	Spec. Ed.	Chapt. I
27	0.2	-	0.3	-	-
28	0.2	-	0.3	-	-
29	0.2	-	0.3	-	-
30	0.4	-	0.7	-	-
31	0.4	-	0.7	0.5	-
32	0.4	-	0.7	0.9	-
33	0.4	-	0.7	0.9	-
34	0.6	-	1.0	0.9	-
35	0.9	0.9	1.0	0.9	-
36	1.5	2.3	1.0	1.4	-
37	1.9	3.2	1.0	2.3	-
38	2.2	3.7	1.3	2.8	-
39	3.2	6.0	1.3	2.8	-
40	4.1	6.9	2.0	3.3	-
41	6.6	10.6	3.6	3.3	2.1
42	10.7	16.2	6.5	5.1	4.2
43	14.2	21.3	9.1	7.9	5.2
44	19.5	28.2	12.7	11.6	12.5
45	27.5	38.0	19.9	14.4	22.9
46	37.5	51.4	27.8	27.4	39.6
47	47.2	63.0	35.9	41.4	63.5
48	59.9	77.8	47.7	55.8	81.3
49	74.9	85.6	68.0	75.3	96.9
50	90.8	95.4	87.3	92.1	99.0
51	100.0	100.0	100.0	100.0	100.0

Table 8

Frequency Distributions for PD Scores:

Cumulative Percentages

Performance Dimension	PD Score	Regular Total	Regular Black	Regular White	Spec. Ed.	Chapt. I
I. Planning	1	-	-	-	0.5	-
	2	0.7	1.4	0.3	0.5	-
	3	1.9	3.2	1.0	2.3	1.0
	4	4.3	5.6	3.3	6.0	2.1
	5	11.8	17.6	7.2	9.3	6.3
	6	27.5	37.5	19.9	26.0	31.3
	7	51.3	64.4	41.8	58.1	67.7
	8	100.0	100.0	100.0	100.0	100.0
II. Instruction	5	0.4	-	0.7	0.5	-
	6	0.4	-	0.7	0.9	-
	7	1.5	1.9	1.3	1.4	-
	8	4.5	8.3	2.0	2.8	-
	9	15.7	25.5	9.5	9.3	7.3
	10	41.4	53.7	33.7	37.7	34.4
	11	100.0	100.0	100.0	100.0	100.0
III. Management	1	0.4	-	0.7	-	-
	2	0.4	-	0.7	-	-
	3	0.4	-	0.7	-	-
	4	0.4	-	0.7	-	-
	5	0.4	-	0.7	-	-
	6	0.7	-	1.3	-	-
	7	0.9	0.5	1.3	-	-
	8	2.8	2.3	2.9	-	3.1
	9	6.1	6.9	5.6	2.3	3.1
	10	16.5	17.1	16.3	10.2	16.7
	11	100.0	100.0	100.0	100.0	100.0
IV. Communication	6	-	-	-	0.5	-
	7	0.7	0.9	0.7	2.3	-
	8	2.4	4.6	1.0	3.3	-
	9	15.5	21.3	11.4	11.6	10.4
	10	52.4	69.9	39.9	45.6	58.3
	11	100.0	100.0	100.0	100.0	100.0
V. Attitude	2	0.2	0.5	-	-	-
	3	0.4	0.9	-	-	-
	4	1.1	2.3	0.3	-	1.0
	5	3.9	6.9	1.6	3.3	2.1
	6	10.1	14.4	7.2	4.2	6.3
	7	19.3	21.8	17.6	15.8	28.1
	8	40.1	45.4	36.3	31.6	59.4
	9	68.5	75.9	62.7	65.1	88.5
	10	100.0	100.0	100.0	100.0	100.0

Table 9

Percentage of Observation Sheets on Which Teachers

Demonstrated Each Competency

APT Competency	Begin. Teachers	Regular Total	Regular Black	Regular White	Spec. Ed.	Chapt. I
PD 1: PLANNING						
a. outcomes stated	81.2	88.0	86.1	90.5	88.4	87.5
b. objectives compatible	97.2	97.8	97.7	98.0	98.1	97.9
c. procedures stated	89.0	93.1	89.8	95.1	90.3	87.5
d. students involved	99.6	99.6	99.1	100.0	98.6	97.9
e. materials stated	89.4	90.1	86.1	93.1	88.0	83.3
f. differences planned	68.1	72.5	62.5	79.7	82.4	84.4
g. objectives assessed	57.8	76.4	74.1	78.4	70.8	71.9
h. progress recorded	61.3	85.0	75.0	91.5	81.0	81.3
PD 2: INSTRUCTION						
a. began promptly	98.9	98.9	99.5	98.4	99.1	97.9
b. objectives addressed	98.2	98.5	99.1	98.0	99.1	100.0
c. needs accomodated	64.5	73.4	60.6	81.4	82.9	90.6
d. interest stimulated	77.7	89.0	84.7	91.5	83.8	82.3
e. approaches varied	98.9	100.0	100.0	100.0	98.6	100.0
f. sizes varied	85.1	87.6	82.4	91.5	90.3	93.8
g. active opportunities	98.6	99.4	99.1	99.7	100.0	99.0
h. application opport.	94.7	97.4	95.8	98.4	98.6	100.0
i. information obtained	97.2	96.8	96.3	97.1	96.3	97.9
j. progress provided	98.2	98.3	96.8	99.3	100.0	96.9
k. physical arrangement	97.2	96.8	96.3	97.1	99.1	100.0
PD 3: MANAGEMENT						
a. behavior established	98.6	98.5	99.1	98.0	99.5	99.0
b. firm enforcement	92.9	98.1	99.1	97.7	100.0	97.9
c. procedural confidence	95.7	99.1	99.1	99.0	99.1	99.0
d. instruction continued	87.9	95.1	94.0	95.8	97.7	91.7
e. disruptions addressed	93.3	96.6	96.3	96.7	99.5	97.9
f. codes enforced	93.3	96.8	98.6	95.4	97.2	97.9
g. inattentive involved	85.1	96.3	95.8	96.4	98.1	97.9
h. special assistance	97.2	96.6	96.8	96.4	99.1	100.0
i. strategies adjusted	95.0	97.8	98.6	97.1	98.6	97.9
j. patient, poised	98.9	98.3	97.7	99.0	99.1	100.0
k. fair, impartial	97.5	97.8	98.1	97.7	99.5	97.9

(continued)

Percentage of Observation Sheets on Which Teachers Demonstrated Each Competency

Table 9 (continued)

APT Competency	Begin. Teachers	Regular Total	Regular Black	Regular White	Spec. Ed.	Chapt. I
PD 4: COMMUNICATION						
a. instructional plan	83.0	90.3	85.6	93.5	86.1	86.4
b. logical sequence	98.2	99.8	100.0	99.7	100.0	99.0
c. understandable level	98.9	99.4	100.0	99.0	98.1	99.0
d. explanations restate	98.6	97.4	98.1	96.7	99.5	100.0
e. illust. demonstrated	96.4	97.4	96.8	97.7	94.9	100.0
f. knowledgeable auth.	98.6	99.1	98.6	99.3	99.1	100.0
g. information accurate	98.6	96.4	96.8	96.1	97.2	96.9
h. legible writing	90.4	94.8	93.5	96.1	96.3	95.8
i. written communicat.	48.9	58.4	40.3	71.6	66.2	54.2
j. oral communication	99.6	97.0	94.4	98.7	99.1	100.0
k. speech quality	99.3	98.9	99.1	98.7	99.5	100.0
PD 5: ATTITUDE						
a. courtesy modeled	98.2	97.9	99.1	97.7	99.1	100.0
b. positive reinforce.	94.3	94.2	89.4	97.4	99.1	92.7
c. expression encouraged	83.3	90.1	89.8	89.9	89.4	86.5
d. learning personalized	68.8	74.7	69.0	79.4	78.7	66.7
e. supportive correction	94.7	95.7	95.4	95.8	99.1	90.6
f. reasons given	81.2	86.1	82.9	88.9	88.9	89.6
g. value communicated	59.9	64.8	57.4	70.6	72.7	70.8
h. enthusiasm communic.	75.2	83.3	82.9	83.3	88.9	75.0
i. open-mindedness	99.3	99.6	99.5	99.7	100.0	100.0
j. humor acknowledged	61.7	69.9	66.7	71.6	64.8	42.7

Table 10

Inter-Observation and Inter-Rater Reliability

Teacher Group	Number of Teachers	Inter-Observation Reliability			Average Inter-Rater Reliability
		Principals	District	Other	
Beginning teachers	47	.506	.664	.622	.574
Regular teachers	89	.713	.590	.684	.572
Special education teachers	54	.689	.777		.687
Chapter I teachers	24	.542	.389		.298
Beginning & Regular teachers					
Black	37	.787	.672	.734	.557
White	94	.614	.634	.650	.604
Special & Title 1 teachers					
Black	27	.636	.714		.541
White	44	.703	.717		.671

Table 11

Overall Standard Deviation, Reliability and Standard Error of Measurement

Teacher Group	Number of Teachers	SD	Reliability	SEM
Beginning teachers	47	3.646	.578	2.37
Regular teachers	89	3.402	.590	2.18
Special education teachers	54	3.054	.702	1.67
Chapter I teachers	24	1.949	.354	1.57
Black teachers	64	3.263	.589	2.09
White teachers	138	3.511	.633	2.13
All teachers	214	3.410	.589	2.19

Table 12

Average Raw Agreement Indices* for

Each Competency

APT Competency	Begin. Teachers	Regular Total	Regular Black	Regular White	Spec. Ed.	Chapt. I
PD 1: PLANNING						
a. outcomes stated	.84	.86	.82	.90	.84	.83
b. objectives compatible	.96	.97	.97	.98	.96	.96
c. procedures stated	.84	.90	.86	.93	.89	.82
d. students involved	.99	.99	.99	1.00	.98	.96
e. materials stated	.83	.86	.82	.88	.83	.86
f. differences planned	.73	.80	.75	.84	.83	.80
g. objectives assessed	.77	.76	.72	.80	.73	.73
h. progress recorded	.72	.87	.81	.90	.81	.75
PD 2: INSTRUCTION						
a. began promptly	.99	.98	1.00	.97	.98	.97
b. objectives addressed	.99	.97	.99	.96	.99	1.00
c. needs accomodated	.72	.75	.70	.78	.84	.85
d. interest stimulated	.69	.88	.87	.89	.77	.76
e. approaches varied	.98	1.00	1.00	1.00	.98	1.00
f. size varied	.79	.84	.80	.88	.87	.89
g. active opportunities	.97	.98	.97	.99	1.00	.98
h. application opport.	.91	.96	.94	.97	.97	1.00
i. information obtained	.95	.95	.94	.95	.94	.96
j. progress provided	.98	.97	.95	.98	1.00	.95
k. physical arrangement	.95	.95	.94	.96	.98	1.00
PD 3: MANAGEMENT						
a. behavior established	.97	.98	.99	.98	.99	.98
b. firm enforcement	.92	.99	.99	1.00	1.00	.97
c. procedural confidence	.96	.99	1.00	.98	.98	.98
d. instruction continued	.88	.92	.90	.93	.96	.85
e. disruptions addressed	.89	.96	.96	.96	.99	.97
f. codes enforced	.89	.96	.99	.94	.95	.96
g. inattentive involved	.85	.95	.97	.93	.97	.96
h. special assistance	.94	.95	.94	.96	.98	1.00
i. strategies adjusted	.96	.96	.97	.96	.97	.96
j. patient, poised	.98	.98	.97	.99	.98	1.00
k. fair, impartial	.96	.96	.97	.96	.99	.97

*The raw agreement index is the percentage of observations for which raters agreed on their ratings of an individual item. The index can range from 0 to 1.00, with 0 indicating no agreement on any observation and 1 indicating agreement on all observations. Since chance agreement is .50, the index actually ranging from .50 to 1.00.

Average Raw Agreement Indices
for Each Competency
Table I2 (continued)

APT Competency	Begin. Teachers	Regular Total	Regular Black	Regular White	Spec. Ed.	Chapt. I
PD 4: COMMUNICATION						
a. instructional plan	.81	.96	.81	.89	.80	.83
b. logical sequence	.98	1.00	1.00	1.00	1.00	.98
c. understandable level	.98	.99	1.00	.98	.98	.98
d. explanations restated	.97	.97	.99	.96	.99	1.00
e. illust. demonstrated	.97	.95	.92	.97	.92	1.00
f. knowledgeable auth.	.98	.98	.97	.99	.99	1.00
g. information accurate	.98	.93	.94	.92	.95	.95
h. legible writing	.86	.91	.89	.93	.94	.92
i. written communication	.61	.66	.67	.65	.76	.58
j. oral communication	1.00	.96	.95	.97	.99	1.00
k. speech quality	.99	.98	.99	.98	.99	1.00
PD 5: ATTITUDE						
a. courtesy modeled	.99	.97	1.00	.96	.98	1.00
b. positive reinforce.	.94	.92	.87	.96	.98	.91
c. expression encouraged	.76	.85	.88	.82	.81	.74
d. learning personalized	.63	.69	.62	.74	.75	.60
e. supportive correction	.96	.93	.89	.95	.99	.85
f. reasons given	.73	.82	.79	.86	.82	.82
g. value communicated	.63	.65	.62	.69	.68	.67
h. enthusiasm communic.	.69	.80	.79	.81	.88	.67
i. open-mindedness	.99	.99	.99	.99	1.00	1.00
j. humor acknowledged	.64	.71	.66	.75	.73	.58

Table 13

Differences in Mean, Standard Deviation

and Percent Below 44 for Principals and District Observers

Teacher	Number of Observations	Observer	Mean	SD	Percent below 44
Beginning teachers	94	District	45.14	3.61	22.3
		Principal	45.48	3.44	25.3
Regular teachers	178	District	46.74	3.29	15.2
		Principal	47.23	3.56	12.9
Special education teachers	108	District	47.26	3.18	9.3
		Principal	47.72	2.91	7.4
Chapter I teachers	48	District	46.90	2.02	6.3
		Principal	46.56	1.88	4.2
All teachers	428	District	46.54	3.30	14.3
		Principal	46.89	3.32	13.3