DOCUMENT RESUME

ED 273 677 TM 860 533

AUTHOR	Ligon, Glynn; Ellis, John
TITLE	Adjusting for Rater Bias in Teacher Evaluations: Political and Technical Realities.
INSTITUTION	Austin Independent School District, Tex.
REPORT NO	AISD-85-55
PUB DATE	Apr 86
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Bias; Career Ladders; Elementary Secondary Education; *Equated Scores; Evaluation Criteria; Evaluation Methods; Evaluation Problems; *Interrater Reliability; *Merit Rating; *Politics of Education; Scores; Teacher Effectiveness; *Teacher Evaluation

ABSTRACT

For Texas's Career Ladder System of rewarding good teachers, teachers' performance evaluations from 1981 to 1984 were used to rank teachers in the Austin Independent School District. Significant biases were noted between raters, between years, and between elementary and secondary teacher ratings. To adjust for these biases, each teacher's raw score average was converted to a z-score, and adjusted for the factors associated with bias. The final z-score was used to determine selection for the Career Ladder; quota systems were not used. A small correlation was noted between teachers' z-scores and elementary school students' reading and mathematics achievement on the Iowa Tests of Basic Skills. Teachers' raw scores were less able to predict student achievement. However, z-scores were not well received. This technique was eventually abandoned due to: (1) dissatisfaction and distrust; (2) changes in legislation and confused communication; (3) the non-selection of some key teachers; (4) differing goals of the teachers' associations; and (5) the provision of additional funding for a greater number of career ladder teachers, which made z-scores unnecessary. (GDC)

ADJUSTING FOR RATER BIAS IN TEACHER EVALUATIONS: POLITICAL AND TECHNICAL REALITIES

Glynn Ligon
John Ellis
Austin Independent School District
Austin, Texas

Paper Presented at the Annual Meeting of
the American Educational Research Association

San Francisco, California April, 1986

Publication #85.55

ADJUSTING FOR RATER BIAS IN TEACHER EVALUATIONS: POLITICAL AND TECHNICAL REALITIES

What do you do when you have to give $1,500 to your "best" teachers but your evaluation system has a few biases in it? The law says you must give out the bonuses and all teachers cannot get one. Is there a valid method for adjusting for the biases to ensure that the money goes to the best teachers?

In the summer of 1984, the Texas Legislature approved a Career Ladder system to reward teachers. Incredibly, the system was implemented immediately using 1983-84 or earlier personnel evaluations completed without knowledge of the ultimate use of the ratings. This presented the Austin Independent School District, Austin, Texas with the dilemma of how to use these evaluations to select the "best" teachers. This paper describes and critiques how evaluation ratings were adjusted to address the issue of bias. The method used will be critiqued from two perspectives--those of the research and evaluation Director and of the Superintendent. These two perspectives provide assessments of both the technical and political successes achieved by the method employed. The research and evaluation Director critiqued the adjustment method based upon the technical and statistical effects of using standardized evaluation ratings on selections for the Career Ladder. The Superintendent assessed the political impact in terms of the success of the adjustment technique in meeting systemwide objectives and needs.

Overview

The annual performance evaluations across three school years (1981-82, 1982-83, and 1983-84) were used to rank teachers from highest to lowest for selection of the highest rated for the Career Ladder. Significant rater bias was evident across campuses. With unadjusted raw score averages across the 43 factors rated for each teacher, some campuses would have had all teachers rated above the District average, and other campuses would have had very few. There was also a mild effect for the year of rating with rating inflation apparent after the first year and a problematic difference between elementary and secondary teacher ratings.

To adjust for these biases, each teacher's raw score average was converted to a z-score within all evaluations from the year of

Figure 1: COMPARISON OF RATINGS BY PRINCIPALS

Senior High	Average Rating	Lowest Rating	Highest Rating
A(N= 99)	3.67	2.83	4.57
B(N=105)	3.64	2.61	4.58
C(N=102)	3.71	2.77	4.76
D(N= 84)	3.94(H)	2.82	4.97(H)
E(N= 90)	3.89	2.99(H)	4.73
F(N=120)	3.37(L)	2.01(L)	4.30(L)
G(N=164)	3.67	2.78	4.55
H(N= 97)	3.60	2.46	4.91
I(N= 79)	3.45	2.87	4.70

H=Highest among Principals L=Lowest among Principals

Junior High			
A(N=56)	3.98	3.04(H)	4.66
B(N=46)	3.53(L)	2.97	4.55
C(N=45)	3.88	2.94	4.63
D(N=49)	3.85	2.62	5.00(H)
E(N=49)	3.53(L)	3.01	4.08(L)
F(N=61)	3.65	2.32	5.00)H)
G(N=53)	3.69	2.99	4.24
H(N=41)	4.15(H)	2.80	4.92
I(N=56)	3.55	2.50(L)	4.46
J(N=39)	4.08	2.99	4.94

Elementary*			
A(N-29)	3.24(L)	2.49	3.92
B(N=15)	4.72(H)	3.51	5.00(H)
C(N=63)	3.92	1.52(L)	4.87
D(N=20)	4.35	3.94(H)	4.99
E(N=63)	3.39	2.40	3.90(L)

*Selected Schools

RATING SCALE:
5 = Outstanding
4 = Strong
3 = Good/expected
2 = Minimally acceptable
1 = Unacceptable

rating and the instructional level of the teaching assignment (i.e., elementary, secondary). Then the teacher's z-score was converted to another z-score within the group of all teachers evaluated by the rater. This second z-score was then ranked among all teachers to determine selection for the Career Ladder.

Is There Rater Bias?

Probably. Even without data to examine, almost everyone would predict confidently a wide variation among raters. Figure 1 supplies the data to show this is true in our schools. On a five-point scale, the principals ranged from 3.24 to 4.72 in their average ratings. Proving bias is tough because none of our teachers have been rated by more than one principal each year, and someone could argue that teachers vary that much across campuses. However, in the case study arena, we did have a high-rating principal trade schools with a low-rating principal. The next year the ratings of the teachers in the two schools shifted. Knowing that the two faculties remained relatively stable across the two years, one could conclude that true rater bias influenced the shift. Figure 2 provides the mean ratings for these principals.

Figure 2: CASE STUDY--MEAN RATINGS FOR TEACHERS IN SCHOOLS THAT TRADED PRINCIPALS

	School A	School B
Principal A	3.91 (1982)	3.46 (1983)
Principal B	4.37 (1983)	4.33 (1982)

Are There Other Biases?

Yes. Figure 3 shows that the mean ratings of our teachers have varied across years. In 1981-82, intense training and a challenge to rate teachers using the full range of the five-point scale appear to have moderated the ratings. The next two years, ratings rose, possibly because the impact of the earlier training faded. This bias based upon the year of rating is critical because teacher evaluations from all these years had to be used for 1984-85 Career Ladder decisions.

A major bias lies in the difference between elementary and secondary teacher ratings. Whether the evaluation system favors elementary teachers, or elementary teachers are better, or elementary principals are higher raters is unknown. Figure 3 also shows that the average elementary teacher is rated about a quarter of a point higher than the average secondary teacher.

What Options Were Considered for Adjusting?

The identified biases made using unadjusted raw scores unacceptable for selecting the top teachers. Among others, these were the adjustment options given serious consideration.

- Adding or subtracting an increment based upon the difference between the overall mean rating and a principal's mean rating. This method was not preferred because the variance of ratings given by very high or low raters would not be adjusted. Mainly, the increment approach was discounted as being a quota system.

- Selecting the same percentage of elementary and secondary teachers. This method was also discounted as being a quota method that did not adjust for biases within the elementary and secondary groups.

- Converting the raw scores to z-scores. This method was preferred because it transformed each principal's distribution of ratings to a common mean and standard deviation. This method is not a quota system because teachers rated highly by a high rater are still generally high within the distribution of z-scores.

Figure 3: MEAN RATINGS BY YEAR AND GRADE LEVEL

	Elementary	Secondary
1981-82	3.92 (SD .53)	3.68 (SD .39)
1982-83	4.00 (SD .53)	3.77 (SD .43)
1983-84	4.00 (SD .53)	3.73 (SD .44)

How Were Z-Scores Calculated?

Three types of biases had to be considered in calculating z-scores--rater, year, and level (elementary or secondary). Therefore, a two-step process was designed. First, a z-score was calculated for each teacher based upon year and level. Specifically, all the elementary teachers rated during the same year were used to calculate z-scores, and all secondary teachers similarly. This made six groups (three years X two levels) within which z-scores were calculated. Special calculations were made for teachers/groups with ratings made earlier than the three-year period discussed here or with a level not clearly elementary or secondary. Second, these z-scores were used to calculate a second z-score among all teachers with the same rater/principal.

This sounds straightforward, but explaining it to the teachers was difficult. Attachment A contains both a copy of the worksheet developed to allow teachers to calculate their own z-scores and a copy of the step-by-step procedure documented for our use.

Do Z-Scores or Raw Scores Correlate with Student Learning?

Several factors argue against finding a correlation between a teacher's evaluation rating and that teacher's students' achievement. These include the following.

- The biases described (rater, year, level)
- The high correlation of achievement with other factors (i.e., previous achievement level, family income, etc.)
- The uncertainty as to which teacher actually delivers instruction to which students

However, a small, statistically significant correlation is reported in Figure 4 for both raw scores and z-scores with reading and mathematics achievement.

Achievement gain in this analysis is defined as gain on the Iowa Tests of Basic Skills (ITBS) Reading and Mathematics Total scores from spring, 1984, to spring, 1985. Using a linear regression model, 1985 scores were predicted from 1984 scores, student family income, school pupil/teacher ratio, sex, ethnicity, and status within the local reassignment patterns for integration. Each student's predicted ITBS grade equivalent score was subtracted from the actual score to calculate an "achievement discrepancy." A positive achievement discrepancy indicates a student who scored higher than the predicted score. Details of this model are contained in ORE Publication 83B, The Report on School Effectiveness--ROSE. Only elementary scores are used because of the complexity of matching courses, teachers, and test scores at the secondary level.

Figure 4: CORRELATIONS WITH ACHIEVEMENT DISCREPANCY FOR Z-SCORES AND RAW SCORES

	MATH		READING	
	Z	RS	Z	RS
	r	r	r	r
Grade	ROSE	ROSE	ROSE	ROSE
0	.278	.221	.212	.190
1	.162	.241	.169	.248
2	.035	.053	.101	.073
3	.338	.332	.294	.269
4	.415	.365	.173	.168
5	.090	.061	.119	.167
6	.221	.195	.172	.174
1-3	.159	.195	.176	.201
4-6	.249	.214	.155	.168
K-6	.212	.204	.172	.184

ROSE = Achievement Discrepancy

Do Z-Scores or Raw Scores Select Better Teachers?

Taking the theoretical situation of needing to divide teachers into two groups--above average and below average. Figure 5 compares the relative outcome using raw scores or z-scores for elementary teachers. Both statistics agreed on the status of 86% of the teachers. For raw scores, above average is defined as a rating at or above the mean for all teachers. For z-scores, above average is defined as a z-score of zero or above. Figure 6 gets at the heart of the question.

Figure 5: COMPARISON OF TEACHERS ABOVE/BELOW AVERAGE ON Z-SCORE AND RAW SCORE RATINGS

	Z < 0	Z ≥ 0
RS < Mean	412 (37.2%)	103 (9.3%)
RS ≥ Mean	54 (4.9%)	538 (48.6%)

N=1107

Figure 6. AVERAGE ACHIEVEMENT DISCREPANCY SCORES OF STUDENTS TAUGHT BY ABOVE/BELOW AVERAGE TEACHERS

	Z < 0	Z ≥ 0
RS < Mean	-.032(R) -.044(M)	.018(R) .007(M)
RS ≥ Mean	-.031(R) -.026(M)	.036(R) .046(M)

R = Reading Achievement Discrepancy
M = Mathematics Achievement Discrepancy

Average achievement discrepancy scores are positive for both reading and mathematics for students of teachers selected as above average on z-scores but below on raw scores. Thus, where z-scores and raw scores disagreed, the z-scores categorized teachers as above average who actually had positive achievement discrepancies even though their raw score ratings were below average. Also, those teachers designated above average by their raw score ratings, but below average by their z-scores, had negative achievement discrepancies. Thus, at least at the point where we might want to divide teachers into two groups, the z-scores select those teachers with positive achievement discrepancies more reliably than do raw scores.

Technically, Do Z-Scores Appropriately Adjust for Rater Bias?

We conclude that z-scores do adjust appropriately for rater bias--from a technical viewpoint. A bonus is the finding that z-scores actually can select the elementary teachers that are in the top half for producing student learning better than raw scores. The z-scores do accomplish what they were chosen to do--even out differences among raters without resorting to an arbitrary quota system.

What Were the Political Realities?

Although z-scores may have been a technically sound method for adjusting for rater bias, politically z-scores were a last resort that was summarily abandoned with little more than a "good riddance." Had we been forced by economics to reward only a percentage of the teachers eligible for the Career Ladder, z-scores would have been used reluctantly. However, once the Board of Trustees committed to raise property taxes sufficiently, they rewarded all eligible teachers and declared z-scores an anathema.

Political Context

Opposition to the Career Ladder or any form of merit pay surfaced early in the Legislative process in Texas and has continued to the present time. A Select Committee was appointed by Governor Mark White, headed by H. Ross Perot, to determine legislative remedies for the ills of Texas schools. It was commonly suggested to the Committee that something had to be done to improve the status of teachers. Two ideas frequently put forward by teachers' groups were higher pay and lower class size. Some sympathy existed for these ideas in the Select Committee and in the State Legislature, but there were caveats.

The Legislature, in particular, wanted some assurances of results. They did not want to put their political necks in the noose of a tax increase without some guarantee that improved quality would follow the additional dollars. From this basic concern flowed the concept that the better teachers should receive Career Ladder money and, additionally, that all teachers should be tested to determine their competence to remain in the classroom.

The teachers' associations met these ideas with strong objections. In the give and take of the legislative process, it was suggested that the Career Ladder might be acceptable if the standards were flexible and sufficient dollars were authorized. But the test was a professional indignity that was adamantly opposed. Numerous leaders warned of the impossibility of determining teaching competence based on a paper-and-pencil examination and further advised that such a test was likely to have a disproportionate impact on minorities.

The Legislative battle climaxed when the Texas State Teachers Association, the largest of the teachers' groups and the NEA affiliate, confronted Lieutenant Governor William Hobby, a long-time friend of education and commonly regarded as the single most influential person in the Legislature. The discussion became acrimonious as teachers failed to convince Lieutenant Governor Hobby of their strenuous objections to the Career Ladder and the test. The Lieutenant Governor became so angry at the attitude of the Texas State Teachers Association that, in a highly publicized incident, he ordered the teacher representatives out of his office and told them never to come back. As of this writing, the rupture continues.

As the Legislative session continued, the impending reality of the Career Ladder, and what came to be the TECAT, the Texas Examination of Current Administrators and Teachers, became unavoidable. When Wilhelmina Delco, Chairman of the House Higher Education Committee, and a Black, who is a formidable supporter of minority issues, acknowledged that the leadership had agreed that the test was the price to be paid for a tax increase, the issue was virtually decided.

It must be emphasized that not all teachers' groups and not all teachers opposed the Career Ladder and the test. Furthermore, they are separate, though related, issues and the support for each varied. Nevertheless, the overwhelming attitude of teachers throughout the State was vocally expressed as dissatisfaction with these changes. There is no question that in Austin, where we had to consult (negotiate) with the local

affiliate of the TSTA, our ability to implement the Career Ladder smoothly, and to gain support for the concept of a z-score which would not only eliminate rater bias but also potentially reduce the number of awardees, was seriously flawed by the negative teacher attitude toward these reforms.

Did People Understand and Accept Z-Scores?

The concept of adjusting for high and low raters was endorsed by all, but not without great disagreement as to the relative merits of using z-scores or merely "adjusting" by using the lowest possible, legal raw-score standard and avoiding most of the biased ratings. The process of converting a raw score to a z-score was mysterious to most people. Typically the unknown is feared and mistrusted.

The bottom line appeared to be that the Career Ladder itself was not generally accepted by the teachers' organization, and, therefore, anything associated with it--especially a method that might help the Career Ladder process work--was rejected. The teachers' organization wanted everyone who met the minimum state standards to be placed onto the Career Ladder. Our interpretation of that minimum level was an average raw score rating above 3.0 on our five-point scale.

Other people were more accepting of z-scores--if only a percentage of eligible teachers could be rewarded. The Career Ladder Committee, which legally had the authority to determine which teachers qualified, the administration, and the Board of Trustees all endorsed z-scores as a better alternative to raw scores with their incumbent bias.

What Factors Caused Z-Scores to be Abandoned in AISD?

Five major factors appear to have defeated z-scores.

1. Dissatisfaction and Distrust

   The predisposition to be negative about the Career Ladder carried over into discussions about eliminating rater bias. It became clear that no system that was complicated, hard to understand, and dependent upon mathematical assumptions would be accepted in the climate of distrust, even when the concept offered the promise of eliminating a basic unfairness.

2. Changing Rules and Confused Communication

The State Board of Education was responsible for issuing rules governing the implementation of the Career Ladder. The conflicts that were present in our local District were also played out at the State level. Decisions about the Career Ladder were changed substantially as the process unfolded, which created an air of uncertainty, postponed final action by local appraisals, and added considerably to the problem of communicating what action would actually be taken.

3. Some Key Teachers Not Selected

Some key teachers fell below the original zero z-score criterion for placement on the Career Ladder. The Career Ladder Committee's first charge was to award about $1.6 million to the top 1067 teachers. The Committee's response was a recommendation to reward 1207 teachers who met the basic eligibility criteria and whose z-scores were zero or greater. Figure 7 details the breakdown of teachers as the criterion for the Career Ladder evolved.

After the administration and Board of Trustees endorsed the Committee's recommendation, the teachers above and below the criterion were notified. However, notices were not sent to all 1207 teachers, but only to 1067 because some of the teachers on the initial list were found not to have the basic eligibility standards for service and education. Unfortunately, outstanding teachers were in the group that was not notified and the publicity about these teachers made the system suspect. At that point, phone calls to Board members and administrators began flooding in from teachers who would eventually be pleading their individual cases before the Board in open, televised meetings. The pleas were well founded and convincing. For example:

- I am the Teacher of the Year in my school, and you are telling me that I am not good enough to be on the Career Ladder?

- I have been on numerous curriculum writing teams and textbook committees, but now I am being told that I am a below average teacher!

4. Different Objectives

Even though the teachers' associations had agreed to the z-score methodology, the association never really accepted the concept that the best teachers should be selected for the Career Ladder. They wanted all teachers to receive the stipend. They did not perceive it to be in their ultimate interest to accept a system such as a z-score method that would eliminate one of their major criticisms--rater bias. They were likely to have a larger number of teachers selected if the standards were lowered as a precaution against bias than if the system corrected for bias. When the public criticism occurred by teachers not selected, the reluctant associations' support of the z-score changed to genuine resistance.

5. Full Funding Provided

Teachers wanted everyone who had the required years experience and the advanced academic training to be on the Career Ladder. When the Board of Trustees decided to "fully fund" the Career Ladder, then z-scores were no longer needed. None of the 29 teachers with a raw score rating at or below 3.0 was willing to argue in public for placement on the Career Ladder. A few did appeal on the grounds that their rater gave "only 3's"; however, the rater-bias problem was functionally laid to rest.

Political Reality

Z-scores rose to prominence on their own merit--the promise of making the selection of the best teachers fair. Z-scores fell from grace because they no longer served a purpose after only 29 teachers needed to be separated out from the other 1921 who were placed on the Career Ladder. The teachers' organization, the administration, and the Board of Trustees began the 1985-86 school year with an agreement to use a -1.0 z-score as the criterion for Career Ladder placement. However, we all quickly agreed on a 3.2 raw score average as a substitute criterion. Indeed, a 3.2 raw score reintroduces some biases avoided at the 3.0 level; however, no one was even tempted to argue for z-scores.

Summary

The use of z-scores to adjust for rater bias met with mixed success and eventually was abandoned. However, z-scores were abandoned for political reasons/realities rather than for technical shortcomings. If our teachers' organization, the staff and the Board of Trustees were ever again faced with a similar situation, would z-scores or raw scores be used? Actually, this question was answered when two months ago a raw score average of 3.2 was embraced by all for entry onto the Career Ladder in the current school year. Quiet consensus was achieved that z-scores were not desired. In a real sense, we chose to accept the rater bias present at this low level of rating rather than to deal with z-scores again.

Figure 7: EVOLUTION OF CAREER LADDER SELECTION CRITERION

3,592 Potential Career Ladder Teachers

1950 Met Basic Eligibility (Years Experience and Advanced Academic Training)

1067 Minimum Number of Career Ladder Positions Required to be Funded by State Law and Originally Funded by Board

140 Additional Teachers Recommended by the Career Ladder Committee to Include All with Z 0.0

131 Additional Positions Funded by Extra Dollars Committed by Board When Committee's Recommendation Adopted* (Z -0.17)

406 Teachers Added When Board Adopted a Criterion of Z -1.0)

177 Teachers Added When Board Adopted a Criterion of Raw Score 3.0

---

29 Teachers with Basic Eligibility but a Raw Score Equal to or Less Than 3.0

1921 Met Evaluation and Basic Eligibility Criteria for the Career Ladder

* Committee thought an additional 271 eligible teachers were at or above a z-score of 0.0; however, only 140 were. The Board funded 271 more positions allowing 131 teachers with z-scores above 0.0 to be placed onto the Career Ladder.

How Can I Calculate My Z-Score?

Your most recent personnel evaluation was used to calculate the score that determined Career Ladder placement. The ratings were adjusted for differences in the average ratings across years and across raters, as follows.

To calculate your score, add together the item ratings (with items 13-37, the "teacher effectiveness" items, each counted as two items). Divide the sum by the number of items on which you were rated, to get an average. Next, a z-score must be calculated, adjusting for differences in average ratings across years.

_._ _	Take **your average rating.**
- _._ _	Subtract the **average rating for all teachers for that year in the same grade span** (elementary, secondary, or other).
= _._ _	Then take this **difference**
÷ _._ _	and divide it by the **standard deviation of all scores for that year in the same grade span.**
= _._ _ _	This is your **Z1 score.** (The average is 0.000.)

This z-score must now be adjusted for rater differences, as follows.

_._ _ _	Take your **Z1 score.**
- _._ _	Subtract the **average Z1 score of all teachers rated by the rater who rated you.**
= _._ _ _	Then take this **difference**
÷ _._ _	and divide it by the **standard deviation of all Z1 scores for your rater.**
= _._ _ _	This is your **Z2 score.** (The average is 0.000.)

A Z2 score above 0.000 was required for placement on Level II of the Career Ladder.

An explanation of z-scores and the formula for calculating them can be found on page 251 of Statistics for the Social Sciences (2nd ed.), by William L. Hays.

NOTE: Round at each step to the number of decimal places indicated. A .005 rounds upward to .01; .0005 rounds upward to .001.

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

CALCULATION OF Z-SCORES FOR CAREER LADDER

Data Sources

1. Files of teacher evaluation results from previous evaluations of the Professional Personnel Evaluation System.

2. Additional evaluation forms not in the original data files. Provided by the Personnel Office.

   a. Forms completed in 1981-82 or subsequent years.

   b. Forms completed prior to 1981-82.

Creation of Basic File for Career Ladder Analyses

Step	Program/Files
Part A: Repeat steps 1-4 three times-- once for elementary teachers, once for secondary teachers, and once for teachers at other campuses.	
1. Identify professionals with one or more evaluations (special education or regular) in 1982-1984 who were active on the EMR at the time of analysis.	Program: DP-PEVFL 0104 Input Files: EMR EDPTEVYR
2. Create Z1, the z-score to correct for the year-to-year variation in evaluation ratings.	
3. Create Z2, the z-score to correct for the variation in ratings across principals.	
4. Write out a career ladder record with Z1, Z2, etc.	

Step	Program/Files
Part B: Sort the three output files from Part A together.	
Part C: Repeat three times as was done in Part A.	
1. Identify professionals with one or more evaluations (special education or regular) in 1982-1984 who were active on the EMR at the time of analysis. In addition pick up teachers who had evaluations prior to 1982, but treat them as if their evaluation was done in 1982.	
2. Create Z1, the z-score to correct for the year-to-year variation in evaluation ratings.	
3. Create Z2, the z-score to correct for the variation in ratings across principals.	
4. Write out a career ladder record with Z1, Z2, etc. only for teachers with evaluations prior to 1982.	
Part D: Sort files from Part C together.	
Part E: Sort output files from Parts B and D together as one file.	Output File: EDPDOSS4

Notes: Weighted averages were calculated using all competencies for which the teacher had a rating of 1 to 5. Competencies in section II, Teacher Effectiveness, were entered twice. The sum was divided by the number of items included. Because the form for special education teachers is longer than the regular form, those teachers were generally rated on more items.