DOCUMENT RESUME

ED 305 395 TM 012 949

AUTHOR Guskey, Thomas R.; Kifer, Edward W.

TITLE Ranking School Districts on the Basis of Statewide Test Results: Is It Meaningful or Misleading?

PUB DATE Mar 89

NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 1989).

PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Academic Achievement; Demography; Educational Quality; Elementary School Students; Elementary Secondary Education; *School Districts; School District Spending; Scores; Secondary School Students; *State Programs; Student Characteristics; Testing Programs; *Test Results; *Validity

IDENTIFIERS *Kentucky; Kentucky Essential Skills Test; *Rank Order

ABSTRACT

The validity of various procedures for rank ordering school districts throughout a state on the basis of statewide test results was studied. Data were drawn from the 1987 administration of the Kentucky Essential Skills Test (KEST), which is given each spring to all children enrolled in public school grades kindergarten through 12 in 178 school districts in the Commonwealth of Kentucky. With the data, school districts were ranked, first using different aggregates of achievement scores and then using achievement scores adjusted for different combinations of demographic factors. The demographic factors and school-based indices included: percent economically deprived, percent local revenue, assessed property value, cost for instruction, current expenses, percent local education agency support, local financial index, teacher/student ratio, and attendance and dropout rates. Achievement variables included: total battery normal curve equivalents score (TBNCES), grade 5 TBNCES, grade 10 TBNCES, reading comprehension NCES, and mathematics score. These rankings were compared, and differences were noted. The various ranking procedures yielded widely varied ranks for most of the districts, regardless of whether non-adjusted or adjusted scores were used. Thus, any ranking procedure is likely to create inaccurate public and governmental perceptions of variation in educational programs, and may lead to misgiven conclusions regarding the programs' quality. Seven data tables are provided. (Author/TJH)

RANKING SCHOOL DISTRICTS ON THE BASIS OF STATEWIDE TEST RESULTS:

IS IT MEANINGFUL OR MISLEADING?

Thomas R. Guskey & Edward W. Kifer

College of Education
University of Kentucky
Lexington, Kentucky 40506
Phone: 606/257-8666

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, March, 1989.

Abstract

The purpose of this study was to investigate the validity of various procedures used to rank-order school districts throughout a state on the basis of statewide test results. Data for the study were drawn from the 1987 administration of the Kentucky Essential Skills Test (KEST), which is given each spring to all children enrolled in public school grades kindergarten through twelve in the Commonwealth of Kentucky. Using these data, school districts were ranked, first using different aggregates of achievement scores and then using achievement scores adjusted for different combinations of demographic factors. These rankings were compared and differences noted. Results showed that the various ranking procedures yielded widely varied ranks for the majority of districts, regardless of whether non-adjusted or adjusted scores were employed. It was thus concluded that any ranking procedures is likely to create inaccurate public and governmental perceptions of variation in educational programs, and may lead to misgiven conclusions regarding the quality of those programs.

RANKING SCHOOL DISTRICTS ON THE BASIS OF STATEWIDE TEST RESULTS: IS IT MEANINGFUL OR MISLEADING?

Over the past decade a growing number of states have initiated statewide testing programs as part of broad-based educational efforts. While some states administer well known, nationally-normed, standardized achievement tests, most employ tests that are developed internally, either by local testing experts or through contractual arrangements with a r putable test-development firm. The results from these tests are interpreted in a variety of different ways and used for a variety of different purposes (Pipho, 1988). In many states, however, they are used as a basis for making comparisons among school districts or schools in an effort to evaluate the quality of educational programs. Such comparisons are typically made by first rank-ordering districts or schools on the basis of some aggregate of test outcomes, and then using these ranks to determine who is doing "best" and who is doing "worst" relative to others within that state. Often such rankings are also employed in various policy initiatives directed toward improving the quality of educational programs.

While the statistical methodology for analyzing the results from such tests is sound and, in most cases, appropriately applied, the interpretations of these analyses are frequently flawed. The purpose of our study is to illustrate how the use of rank-ordering procedures based on statewide test results is far more complicated than what might appear, and that often these complications lead to inappropriate or misgiven conclusions. We illustrate how such rankings can vary greatly depending upon what outcomes from the testing program are used, how those outcomes are aggregated, whether or not input variables are used to make statistical adjustments to the outcomes, and what particular input variables are employed in making such adjustments. The implications of these differences for policy makers are also discussed.

Background

Like most states, the Commonwealth of Kentucky has initiated a variety of educational reforms in recent years. Spurred by reports such as A Nation at Risk: The Imperative for Educational Reform, published by the National Commission on Excellence in Education (1983), the Kentucky General Assembly enacted a series of bills, all targeted at improving the quality of education provided to children in Kentucky public schools. In fact, in 1984 alone the General Assembly passed 43 pieces of legislation bearing on education. This legislation covered a variety of areas, including academic performance, accountability to the public, professional development, community involvement, school funding, and statewide unity of educational purpose.

One of the most prominent legislative initiatives of 1984 was Senate Bill 169, which called for the identification, development, and assessment of skills deemed essential for all of the Commonwealth's students:

> The 1984 General Assembly ... enacted a bill calling for the development of essential skills that should be taught and mastered by students in reading, writing, spelling, mathematics, and library use, including research and reference skills. The bill called for annual testing of these skills in every grade. For young children who do not master those

> skills, the General Assembly provided $16 million in new funds for remedial assistance beginning in the 1986 fiscal year. (Biennial Report, 1983-85, p. 9).

The passage of this bill led to the development of a new statewide test, called the Kentucky Essential Skills Test (KEST). The KEST was developed under contract with CTB/McGraw-Hill of California, and was intended to serve a dual purpose. It was to provide a basis for determining the mastery of a defined set of essential skills, and also produce estimates of student performance referenced to national norms.

The actual construction of the KEST took place in two phases. During the first phase, in 1985, essential skills tests were developed for the domains of reading and mathematics. In the second phase, in 1986, tests were developed for the other three essential skill domains of writing, spelling, and library use. These tests were organized into 13 levels, one for kindergarten and one for each of the grades 1 through 12.

State law required that the KEST be administered to all children enrolled in public schools within the Commonwealth between the 150th and 160th days of the school year. Results from the tests were then reported to the public in the form of a statewide profile, together with individual profiles summarizing results from each of the Commonwealth's 178 school districts. These results included both criterion-referenced and nationally norm-referenced indices.

The principal normative scores used in reporting the KEST results were Normal Curve Equivalents (NCE's). The NCE's are standard statistical scores, derived by dividing the normal distribution into 99 equal segments. Scores can range from 1-99, with a mean/median of 50 and a standard deviation of 21. They were used in this instance primarily because they lend themselves to mathematical manipulations while other more common normative scores, such as percentiles, are not appropriate for such manipulations.

With the report summarizing the results from the first administration of the KEST in 1985, the State Department of Education also published a ranking of all school districts in the Commonwealth. This ranking was based upon a Total Battery NCE Score, which was obtained by statistically combining the subtest scores for reading, writing, and mathematics, and then calculating a weighted average of this combined score across grade levels 3, 5, 7, and 10.

In 1986 the Center for the Study of Testing, Evaluation, and Educational Policy (CSTEEP) at Boston College was awarded the contract to conduct an independent evaluation of the KEST Program. One of the findings from this evaluation was that educators throughout the Commonwealth saw the ranking of school districts on the basis of KEST results as one of the most troublesome aspects of the program. Most regarded the ranks as "pernicious, misleading, fostering unhealthy competition between school districts and detrimental to morale." (CSTEEP, 1986, p. iii). For these reasons, the Department of Education did not rank the school districts when the results from the 1986 KEST administration were made public. But immediately upon release of the test results, newspapers throughout the Commonwealth published articles which included similar rankings (e.g. Roser, 1986).

Some of the strongest criticism of this ranking of districts came from poorer and rural school districts in the Commonwealth which serve economically disadvantaged populations (Wagar & Estep, 1986). School personnel in these districts emphasized that it was inherently unfair to compare their test results with those from districts which have greater financial resources and serve a more advantaged student population. They argued that demographic factors such as these have a very strong influence on school achievement, but lie outside the direct control of a school district's teachers or school administrators.

It was within this context that we undertook the present study. Our initial purpose was to explore the influence of selected demographic factors on school achievement as measured by the KEST, and to determine the effect of correcting for such factors upon the ranking of school districts within the state. In particular, we wanted to determine if correcting for certain demographic factors would lead to a more meaningful interpretation of the test results and a more equitable system of evaluating the educational programs in various school districts. We then turned to consider the effects of using different aggregates of outcome variables and how this might influence the relative standing of districts as well.

Method

The data for this study were drawn from the Kentucky Essential Skills Test: Statewide Testing Results (1987), an annual report published by the Kentucky Department of Education. This report lists the results from the KEST in profiles for each of the 178 school districts in the Commonwealth, as well as data on several demographic and school-based indices. While the test data in this report were from the spring 1987 administration of the KEST, the demographic information was based on school district profiles gathered in 1986. Together, these achievement and demographic measures composed the data set for the study.

The demographic and school-based indices contained in the report included the following:

A. Percent economically deprived (%DEP) - The percentage of children eligible for free school lunch benefits in proportion to the total number of school age children in the district.

B. Percent local revenue (%LOC) - The percentage of a district's total revenue that is received from local sources.

C. Assessed property value (PROP) - This is calculated by dividing the total assessed property value in the district by the average daily attendance for the district. For convenience, this index was further divided by 1000.

D. Cost for instruction (COST) - The cost per pupil for instruction, calculated by dividing the total amount spent for instruction by the average daily attendance. The total amount spent for instruction does not include federal programs.

E. Current expenses (EXPS) - The district's annual current expenses divided by the average daily attendance for the district. It includes costs for administration, instruction, attendance services, health services, pupil transportation, operation of plant, maintenance of plant, and fixed charges.

F. Percent LEA support (%LEA) - The amount of money provided by the Local Education Agency (LEA) in addition to state funds as a percentage of the total expenditures for instructional salaries.

G. Local financial index (LFIN) - This index is calculated by dividing the local revenue per pupil in average daily attendance by the assessed property value per pupil in average daily attendance. It is intended to be an index of the amount of effort a community puts into the support of its schools based upon its ability to pay.

H. Pupil-teacher ratio (P/T) - This index is calculated by dividing the enrollment at the school building level by the number of classroom teachers reported on federal and state salary schedules.

I. Attendance rate (ATTD) - The aggregated days attendance divided by the aggregated days membership, expressed as a percentage.

J. Dropout rate (DROP) - The percentage of pupils enrolled in grades 7-12 who dropout of school during the school year. It includes pupils who became 16 and dropped out, pupils excused from school because of mental or physical disability, pupils discharged, and pupils excused from school because of marriage.

In addition to these ten demographic and school-based indices, five achievement variables were drawn from the report to use in our analysis. These included:

K. Total battery NCE score (ACH) - The average NCE score obtained by statistically combining the subtest scores in reading comprehension, writing, and mathematics for pupils in grades 3, 5, 7, and 10, weighting each grade level NCE by the number of pupils tested at that grade. This is the measure that was used as a basis for originally ranking school districts by the State Department of Education and has been used since by the popular press throughout the Commonwealth.

L. Fifth grade total battery NCE score (ACH5) - The average NCE score obtained by combining the subtest scores for reading comprehension, writing, and mathematics for all fifth grade pupils tested. This measure was included in order to explore achievement differences reflected in the elementary program only.

M. Tenth grade total battery NCE score (ACH10) - The average NCE score obtained by combining the subtest scores for reading comprehension, writing, and mathematics for all tenth grade pupils tested. This measure was included in order to consider the influence of the secondary program in the district.

N. Reading comprehension NCE score (READ) - The average NCE score obtained by combining the reading comprehension scores of pupils in grades 3, 5, 7, and 10, weighting each grade level NCE by the number of pupils tested at that grade. This measure was used to explore differences among districts based on the reading program only.

O. Mathematics NCE score (MATH) - The average score NCE obtained by combining the mathematics scores of pupils in grades 3, 5, 7, and 10, weighting each grade level NCE by the number of pupils tested at that grade. This measure was used to investigate differences among districts based on mathematics achievement alone.

Also included in the data set were the number of students in each district tested at grades 3, 5, 7, and 10, as well as the total number of students enrolled at each of these grade levels. These measures were included to determine the possible influence of district size, or if any irregularities in testing practices might have influenced test results.*

* Some districts were suspected of enhancing their test results by simply testing a smaller portion of the students enrolled.

Procedure

For our purposes we first replicated the ranking procedure that had been employed by the State Department of Education. This involved computing a Total Battery NCE Score for each school district in the state, and then using this score to rank-order all 178 districts. We then compared the ranks obtained by this procedure with several other alternative rankings, each employing a different aggregate of outcomes or outcomes adjusted statistically for various combinations of input variables. Specifically, we reranked the school districts based upon scores obtained from single grade levels (fifth and tenth), scores obtained from single subject areas or domains (reading comprehension and mathematics), scores adjusted for the entire set of demographic and socio-economic variables, and scores adjusted for the single most influential input variable (percent of the student population considered economically deprived). We then compared the results from these various rankings in terms of consistency and interpretive regularity.

Results

The first step in our analysis was to compute means and standard deviations for all of the measures. These are displayed in Table 1, and show there is great variation across districts in nearly all of the measures. The average percent economically deprived, for example, is 36%. But districts range from one with less than 3% to another with more than 82%. The only measures that are exceptions are the local financial index and attendance rate, where there is relatively little variation across districts.

Insert Table 1

Next we computed correlation coefficients among the measures. These are illustrated in Tables 2, 3, and 4. It can be seen in Table 2, as might be expected, that most of the demographic variables are significantly correlated with one another. Attendance rate and dropout rate are the only exceptions, perhaps because of the rather limited variation in these two measures. However, these two variables are also probably the most alterable through the efforts of educators, regardless of the demographic characteristics of the school district or the student population served by that district.

Table 3 shows the correlations between demographic variables and the selected student achievement indices. These figures suppor the contention that such demographic factors are indeed related to variation in student achievement scores. The one exception is pupil-teacher ratio, which is uncorrelated to any achievement measure. Again, however, this is probably due to the rather small variation found in this index across the majority of school districts. The size of the district and the proportion of students tested were also found to be unrelated to achievement scores, indicating that while irregularities in testing procedures may exist, there is no evidence to suggest such practices are widespread.

Correlations among the selected achievement indices are shown in Table 4. These figures illustrate that although all correlations are statistically significant, many are of relatively modest magnitude. For example, the correl-- on between fifth grade achievement (ACH5) and tenth grade achievement (ACH10) is only .44.

Insert Tables 2, 3, & 4

Our third step was to conduct a regression analysis on the total achievement scores. This technique allowed us to eliminate or partial out that portion of the variation in the achievement scores that is attributable to demographic and school-based indices. The residuals resulting from this analysis represent achievement measures that a e adjusted, or "corrected," for demographic and socio-economic differences among districts. In other words, these adjusted achievement measures are unrelated to any of the demographic factors. T y are the achievement scores that would be anticipated if all districts had the same demographic and school-based characteristics.

In our initial regression analysis we used eight of the ten demographic and school-based indices included in the State Department of Education report as independent variables. Attendance rate and dropout rate were excluded because we believed these were more appropriately considered alterable outcome measures that, as mentioned earlier, are under the direct or indirect influence of educators. The results from this multiple regression analysis are illustrated in Table 5. In total, the eight variable model accounted for 45.7% of the variance in Total Battery Achievement scores. That is, nearly 46% of the variance in the total achievement scores is attributable to differences in the demographic and socio-economic characteristics of the districts.

To develop the most efficient model, we next reduced the number of independent variables included in the analysis in stepdown fashion, checking the statistical significance of the reduction in the R-square value with the deletion of each variable. A single variable, percent economically deprived (%DEP) was found to be the most powerful. In fact, the addition of other variables in the model made no significant contribution. As the results in Table 6 illustrate, nearly 42% of the variance in the total achievement scores can be explained with this single variable.

Insert Tables 5 & 6

Our final step was to rank order all school districts in the Commonwealth using various criteria, and then to compare these rankings. We first ranked all districts using the Total Battery Achievement scores (ACH), the criterion initially used by the State Department of Education. Then we ranked the districts on the basis of fifth grade achievement alone (ACH5), tenth grade achievement alone (ACH10), reading comprehension scores (READ), and mathematics scores (MATH). Next, using the residuals from the regression analysis of the full 8-variable model, we ranked the districts according to adjusted or "corrected" total achievement scores (R8ACH), fifth grade achievement scores (R8ACH5), tenth grade achievement scores (R8ACH10), reading comprehension scores (R8READ), and mathematics scores (R8MATH). Lastly, we did the same with the residuals obtained from the regression analysis using only the single independent variable, percent economically deprived, which yielded five additional rankings. Thus, we had 15 different rankings of the 178 school districts in the Commonwealth, each based on a different criterion.

To compare these different rankings we calculated Spearman rank-order correlation coefficients among them. These are shown in Table 7. If rankings were comparable across the various criteria, we would expect these correlations all to be quite high. However, this was not the case. Although the correlation coefficients are all positive, they range in value from .14 to .95, the highest being restrict to those based on residuals calculated from the same criterion variable. The median correlation coefficient among the various rankings is only .56.

Insert Table 7

To estimate the magnitude of the differences in these various rankings, we next calculated difference scores. That is, we simply computed the differences between each pair of ranks. The mean of these differences, of course, is equal to zero. The standard deviation of these differences, however, offers an appropriate descriptive statistic of the magnitude of rank differences.

In Table 8 are displayed the standard deviations of the differences in ranks based on various criteria. These figures show that using different criteria result in ranks that are, in most instances, dramatically different. The standard deviations of the rank differences ranged from 15.3 to 67.0, the smallest differences again resulting from comparisons between ranks based on residuals calculated from the same criterion variable. The median difference standard deviation is 47.0, indicating that overall, a change in ranking criterion would result in a change in rank of nearly 50 places for one-third of the school districts in the Commonwealth. Given there are only 178 school districts in the Commonwealth, a difference of this magnitude is substantial.

Insert Table 8

Cautionary Notes

Before discussing the conclusions drawn from this analysis, a few cautionary notes are necessary. First, it is important to keep in mind that the criteria used in ranking the school districts in all phases of our analysis were achievement scores derived from a test of essential or basic skills. Clearly, using such a test as the sole criterion of the quality of a school district's educational program is inappropriate. Although we would agree that these skills should be taught and that all students should master them, we would hope that the curricula of all districts in the Commonwealth go far beyond the narrow range of learning outcomes assessed by this test.

Second, although correcting for the influence of demographic factors and then ranking schools and school districts accordingly is a widely used technique for identifying those considered to be "effective," we believe such a procedure to be inappropriate. Especially in terms of the skills measured by a test of this type, we believe that students from poor families should achieve at essentially the same high level as students from more affluent families (see Brookover, 1987). We employed the technique in this study only to illustrate some of the striking discrepancies that result when districts are ranked, regardless of the criteria employed.

Third, our analysis considered only the variation that exists between school districts. Other studies have shown, however, that greater variation can exist among schools within a district than does between districts, and even greater variation still among classrooms within a school (Madaus, Kellaghan, Rakow, & King, 1979; Rakow, Airasian, & Madaus, 1978). Thus, had the unit of analysis been schools or classrooms, undoubtedly there would have been greater variation in the measures of each criterion. Nevertheless, it seems unlikely that this increased variation would result in any greater consistency in ranks determined by different criteria.

Finally, in all of our analyses we used weighted averages of test scores that were aggregates either over content areas (eg. ACH5 and ACH10), grade levels (eg. READ and MATH), or both (eg. ACH). This was done to keep our analyses parallel to those which were used to arrive at the rankings originally developed by the State Department of Education. However, we do not advocate the use of such aggregates.

Super-aggregates such as these tend to mask more than they reveal. School districts with very different patterns of achievement can have average total achievement scores that are the same, but results which suggest different paths to take if they are interested in improvement. For instance, one district could be high in mathematics but low in reading, while another district is high in reading but low in mathematics. Although their averages might be similar, the changes they might wish to make should be very different. Even a score in a specific content area such as mathematics may be, in fact, too broad to provide the kind of information needed to implement new and presumably better instruction. Patterns of results within a content area -- we are doing well in metric measurement but not as well as we would like in solving word problems -- provide a better picture of a district's performance and, therefore, better notions of the changes that should be made.

Conclusions

Our analyses show clearly that the use of aggregate test scores to rank school districts can be problematic and potentially misleading. The ranks change, in some cases drastically, as the criterion changes, and whether or not adjustments a made to the criterion. Different approaches lead to different rankings; different rankings lead to different interpretations.

In order to rank schools or school districts, one must assume that both the criteria used and the technique employed are defensible. But since both the criteria and the technique are arbitrary, any one ranking will reflect poorly and unfairly on a significant portion of the units being ranked. Therefore, it follows that interpretations based on any such rankings will be unfair, inaccurate, or invalid.

There is, however, a more fundamental problem associated with rankings such as these. In essence, they do not even serve their intended purpose, which is to tell various interested audiences how well or how poorly schools or school districts are doing. State department a: d school district personnel, educational leaders, the press, and the public need good and accurate information if discussions about education are to produce positive changes in schools. Information that is potentially biased and likely to be misinterpreted will be counterproductive to the any educational improvement effort.

Regardless of the overall level of performance of a state's schools or school districts, a ranking procedure always produces a top and a bottom, with no way of knowing what either really means. A bottom ranked school district may be doing quite well on certain criteria and a top ranked district may be doing quite poorly. The ten shortest players in the National Basketball Association, for instance, are on the average pretty tall. Good information, something much more than a ranking, is needed if good decisions are to be made.

Assessing the performance of schools is, as the literature makes clear, a complex task. However, that should not keep us from working hard to accomplish it. Most importantly, we should not be satisfied with weak attempts based on misleading information when dealing with issues so crucial as the educational experiences of the young.

References

Brookover, W. B. (1987). Distortion and overgeneralization are no substitutes for sound research. Phi Delta Kappan, 69, 225-227.

Center for the Study of Testing, Evaluation, and Educational Policy (1986). An evaluation of the Kentucky Essential Skills Tests in mathematics and reading. Chestnut Hill, MA: Boston College.

Kentucky Department of Education (1985). Biennial report 1983-85. Frankfort, KY: Kentucky Department of Education.

Kentucky Department of Education (1987). Kentucky Essential Skills Test: Statewide testing results. Frankfort, KY: Kentucky Department of Education.

Madaus, G. F., Kellaghan, T., Rakow, E. A., & King, D. (1979). The sensitivity of measures of school effectiveness. Harvard Educational Review, 49, 207-230.

National Commission on Excellence in Education (1983). A nation at risk: The imperative for educational reform. Washington, DC: U.S. Government Printing Office.

Pipho, C. (1988). The uses of test scores. Phi Delta Kappan, 70, 278-279.

Rakow, E. A., Airasian, P. W., & Madaus, G. F. (1978). Assessing school and program effectiveness: Estimating teacher level effects. Journal of Educational Measurement, 15, 15-22.

Roser, M. A. (1986, August 21). Test scores improve in KY schools. Lexington Herald-Leader, pp. A1 & A9-A10.

Wagar, K., & Estep, B. (1986, August 21). Wealthy districts fared best on skills test. Lexington Herald-Leader, pp. A1 & A8.

Table 1

Means and Standard Deviations of Various School District Measures (n=178)

Measure	Mean	Standard Deviation	Range
% Economically Deprived (%DEP)	36.01	15.96	2.7 - 82.1
% Local Revenue (%LOC)	17.81	10.07	2.7 - 66.8
Assessed Property Value (PROP)	101.10	41.12	33.9 - 268.3
Cost for Instruction (COST)	1477.03	230.00	1071.6 - 2998.4
Current Expenses (EXPS)	2040.70	309.40	1630.0 - 4522.0
% LEA Support (%LEA)	22.01	6.95	8.2 - 53.6
Local Financial Index (LFIN)	.43	.21	.14 - 1.16
Pupil-Teacher Ratio (P/T)	19.39	1.81	12.1 - 22.7
Attendance Rate (ATTN)	94.76	1.01	91.3 - 96.6
Dropout Rate (DROP)	3.34	1.90	0.0 - 13.8
Total Battery NCE Score - Combined Grades (ACH)	57.51	3.61	50.3 - 70.8
Total Battery NCE Score - 5th Grade (ACH5)	56.15	4.53	45.9 - 71.1
Total Battery NCE Score - 10th Grade (ACH10)	54.46	4.76	43.6 - 71.1
Reading Comprehension NCE Score (READ)	54.13	3.19	46.8 - 65.9
Mathematics NCE Score (MATH)	58.76	4.07	49.6 - 72.1

Table 2

Intercorrelations Among Selected Demographic Variables
(n = 178)

Meas	%DEP	%LOC	PROP	COST	EXPS	%LEA	LFIN	P/T	ATTD	DROP
%DEP	1.00									
%LOC	-.64*	1.00								
PROP	-.63*	.77*	1.00							
COST	-.33*	.75*	.61*	1.00						
EXPS	-.28*	.72*	.61*	.95*	1.00					
%LEA	-.40*	.79*	.63*	.78*	.73*	1.00				
LFIN	-.32*	.74*	.21	.63*	.58*	.61*	1.00			
P/T	.06	-.41*	-.33*	-.73*	-.67*	-.60*	-.40*	1.00		
ATTN	-.43*	.26	.22	.13	.13	.15	.16	-.10	1.00	
DROP	.45*	-.34*	-.28*	-.07	-.07	-.22	-.20	.04	-.51*	1.00

* $p < .001$

Table 3

Intercorrelations Among Selected Demographic Variables
and Achievement Indices
(n = 178)

Meas	%DEP	%LOC	PROP	COST	EXPS	%LEA	LFIN	P/T	ATTN	DROP
ACH	-.65*	.52*	.46*	.36*	.31*	.30*	.28*	-.07	.49*	-.42*
ACH5	-.50*	.49*	.38*	.38*	.36*	.32*	.32*	-.18	.37*	-.39*
ACH10	-.60*	.49*	.43*	.31*	.26	.25	.29*	-.05	.37*	-.30*
READ	-.69*	.52*	.47*	.37*	.31*	.31*	.29*	-.10	.41*	-.44*
MATH	-.49*	.44*	.39*	.31*	.24	.27	.24	-.04	.48*	-.41*

* $p < .001$

Table 4

Intercorrelations Among Selected Achievement Indices*
(n = 178)

Meas	ACH	ACH5	ACH10	READ	MATH
ACH	1.00				
ACH5	.81	1.00			
ACH10	.76	.44	1.00		
READ	.92	.75	.70	1.00	
MATH	.91	.76	.64	.79	1.00

* All correlation coefficients are statistically significant at $p < .001$

Table 5

Summary of Regression Analysis: Achievement Scores on Total Model
(Y = ACH)

Source	df	Sum of Squares	F-value	Prob	R-square
Model	8	1082.15	19.08	.0001	.457
Error	164	1162.58			
Total	172	2244.73			

Table 6

Summary of Regression Analysis: Achievement Scores on Percent of Students Economically Deprived
(Y = ACH)

Source	df	Sum of Squares	F-value	Prob	R-square
% Econ Deprived	1	943.08	123.93	.0001	.417
Error	171	1301.64			
Total	172	2244.73			

Table 7

Intercorrelations Among Rankings Employing Different Criteria
(n = 178)

Criteria	ACH	ACH5	ACH10	READ	MATH	R8 ACH	R8 ACH5	R8 ACH10	R8 READ	R8 MATH	R1 ACH	R1 ACH5	R1 ACH10	R1 READ	R1 MATH
ACH	1.00														
ACH5	.91	1.00													
ACH10	.91	.78	1.00												
READ	.80	.72	.77	1.00											
MATH	.76	.70	.64	.46	1.00										
R8ACH	.70	.57	.64	.53	.44	1.00									
R8ACH5	.55	.45	.56	.82	.19	.66	1.00								
R8ACH10	.40	.32	.33	.14	.73	.56	.15	1.00							
R8READ	.56	.65	.47	.44	.36	.86	.55	.48	1.00						
R8MATH	.58	.40	.75	.47	.30	.81	.59	.42	.60	1.00					
R1ACH	.74	.61	.66	.58	.47	.92	.61	.53	.80	.76	1.00				
R1ACH5	.56	.45	.56	.85	.21	.59	.91	.14	.51	.54	.68	1.00			
R1ACH10	.44	.34	.34	.17	.75	.54	.14	.95	.46	.38	.59	.20	1.00		
R1READ	.62	.68	.50	.49	.38	.80	.53	.45	.95	.57	.87	.58	.50	1.00	
R1MATH	.60	.42	.77	.49	.33	.77	.56	.41	.58	.95	.82	.60	.44	.63	1.00

Table 8

Standard Deviations of Differences in Rankings
Employing Different Criteria
(n = 178)

Criteria	ACH	ACH5	ACH10	READ	MATH	R8 ACH	R8 ACH5	R8 ACH10	R8 READ	R8 MATH	R1 ACH	R1 ACH5	R1 ACH10	R1 READ	R1 MATH
ACH	0.0														
ACH5	21.7	0.0													
ACH10	21.5	33.4	0.0												
READ	32.0	37.9	34.7	0.0											
MATH	35.0	38.9	42.6	53.0	0.0										
R8ACH	38.9	46.4	42.7	49.3	52.9	0.0									
R8ACH5	48.1	53.5	47.8	31.1	64.7	42.3	0.0								
R8ACH10	54.7	58.6	58.1	67.0	37.1	46.8	66.4	0.0							
R8READ	46.0	41.8	51.5	54.0	56.7	27.0	48.2	51.1	0.0						
R8MATH	46.0	54.8	35.5	52.5	59.2	30.7	46.2	54.2	44.7	0.0					
R1ACH	36.2	44.4	41.1	46.7	51.8	19.6	45.0	48.7	31.3	35.0	0.0				
R1ACH5	47.4	53.4	47.8	27.9	63.9	45.8	21.4	66.7	50.4	48.6	40.8	0.0			
R1ACH10	53.2	57.7	57.5	65.4	35.7	48.0	66.6	15.6	52.0	55.7	45.4	64.0	0.0		
R1READ	43.9	40.0	49.9	51.6	55.7	31.6	49.5	52.4	16.2	46.3	25.7	46.3	49.8	0.0	
R1MATH	44.6	53.8	34.0	51.1	58.0	33.6	47.7	54.4	45.8	15.3	30.4	45.6	52.9	43.1	0.0