DOCUMENT RESUME

ED 350 683 EA 024 371

AUTHOR Denoyer, Richard; White, Michael
TITLE Is This Really Fair: Education Reform and Ranking Districts by Test Scores.
PUB DATE [92]
NOTE 10p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Accountability; *Educational Equity (Finance); Elementary Education; *Fiscal Capacity; Income; Performance Tests; *School Districts; Scores; Socioeconomic Status; *Standardized Tests; State Standards; *Test Validity
IDENTIFIERS *Ohio; *Ranking

ABSTRACT

Findings of a study that investigated the relationship between the socioeconomic characteristics of Ohio school districts and district test performance are presented in this paper. Methodology involved ANOVA analyses of data derived from the 611 Ohio school districts that participated in the state's 1989-90 testing program. The independent variable was district wealth and the dependent variable was district performance based on students' scores on the state's reading, language, and mathematics tests. Wealth was operationally defined as the average family income per tax return in the first analysis, and as the percentage of district families receiving aid to dependent children (ADC) in the second analysis. Findings show that district performance rankings were largely a measure of family wealth, which did not reflect school effectiveness or instructional quality. School districts should not be ranked according to test scores that are highly influenced by wealth, but on the basis of three categories of educational quality: access to knowledge; district expectations; and the teaching environment. Two tables are included. (LMI)

IS THIS REALLY FAIR:

EDUCATION REFORM AND RANKING DISTRICTS BY TEST SCORES

Richard Denoyer and Michael White

Substitute House Bill 231 and Senate Bill 140 required all school districts in Ohio to administer standardized tests last year to fourth. sixth. eighth grade students. Each district reported the average test scores of these students in reading. language arts. and mathematics. Each district also reported the number of students in these grades scoring at or above the 75th. 50th. and 25th percentiles and the number of these students who achieved as well as. higher than. and lower than their ability.

The State Department of Education compiled the statewide test results in a 1.400-page report. This report contained state aggregate. district. and school building achievement test results and achievement as predicted by ability for the three grade levels and three subject areas.

Although all school districts used the same NCL score scale and the same national benchmarks. several State Board of Education-approved tests. each with its own normative data. were used.

As predicted. the release of this data led to newspaper's ranking districts in their county and across the state according to test scores.

Proponents of these ranking argued that the information enabled districts and their residents to gauge the quality of their school districts. Those of us who opposed such rankings seen them as statistically flawed and as providing a limited amount of information. We feared that school and district test score comparisons would mislead the public and send education reform efforts in the wrong direction.

Ranking schools and districts according to student performance on an achievement test is a bad idea. Test scores that fail to take into account the effects of background variables such as ability, poverty, or language proficiency are not comparable within or across districts. Rankings based on an average achievement test score or the proportion of students achieving at a specified percentile are unfair to schools and districts serving significant numbers of students from socio-economically deprived families. Research shows that the average income of a community from which a school district draws its students will impact test scores. On the average, students from districts whose residents have high incomes score better on standardized tests than students from districts whose residents are less affluent. There are various arguments about why this pattern arises, but most testing and evaluation experts agree that this association does not necessarily imply that school districts with more affluent families are doing a better job than school districts serving average or poor families.

The question this study addressed is whether there was a strong relationship between the socio-economic characteristics of school districts involved in Ohio's 1989-90 testing program and test performance.

Methods

The population of this study was the 611 Ohio school districts that participated in the state's testing program during the 1989-90 school year. The independent variable was district wealth and the dependent measure was district (grades 4, 6 and 8 combined) test performance in reading, language and mathematics.

Two statistical analyses were performed on the data. In the first analysis, wealth was operationally defined as the average income per tax return. Using this information, districts were divided into five income groups. Group 1 was composed of districts having an average income of 20,000 dollars or less. Group 2 consisted of districts having an average income of 21,000 to 25,000 dollars. Districts whose residents had average incomes between 26,000 and 30,000 composed group 3. Group 4 districts had average incomes between 31,000 and 35,000, while group 5 had average incomes at or above 36,000 dollars.

In the second analysis, wealth was operationally defined as the percent of a district's student population whose families were receiving aid to dependent children (ADC). Using this information, districts were divided into three ADC groups. Group 1 was composed of districts having 10 percent or less of their families receiving ADC. Group 2 consisted of districts having 11 to 20 percent of their families receiving ADC. Districts with 21 percent or more of their families receiving ADC composed group 3.

Analysis

Six separate ANOVAS were conducted. The first set of three analyses examined the relationship between income groups and average NCE(normal curve equivalents) scores in reading, language and mathematics. The second set of ANOVAS was conducted to determine if there was a relationship between aid to dependent children groups and NCE scores in reading, language and mathematics.

Results

Income: Significant main effects were uncovered from the ANOVAS for Reading $F(4,1828)=194.01$, $p=.0001$; Language $F(4,1828)=170.39$, $p=.0001$ and Mathematics $F(4,1828)=166.97$, $p=.0001$.

The Scheffe procedure indicated significant mean differences (p=.05) in reading, language and mathematics between group 5 and group 4, group 4 and group 3, group 3 and group 2, and group 2 and group 1 (see table 1).

ADC: Significant main effects were also uncovered from these ANOVAS for Reading F(2,1830)=190.27, p=.0001; Language F(2,1830)=117.65, p=.0001 and Mathematics F(2,1830)=133.36, p=.0001.

The Scheffe procedure indicated significant mean differences (p=.05) in reading, language and mathematics between group 3 and group 2, and group 2 and group 1 (see table 2).

Discussion

District rankings, which were hailed as measure of educational quality to a large degree, were merely a measure of family wealth. Districts whose residents had higher average income reported higher NCE scores than districts whose residents had lower average incomes. Districts with a lower percent of ADC families reported higher NCE scores than districts who had a higher percentage of ADC families.

Assuming that a high test-score rank reflects effective education is naive and dangerous. The danger is that people will be convinced that something is being done to identify and correct educational problems in Ohio.

Attention will be diverted from real problems and real solutions, while districts compete and are judged on a biased index of educational quality.

Statewide testing and the subsequent ranking of districts are powerful symbols and strike a responsive chord in politicians and the public. In the 1989 Gallop Poll on Education, 70% of American adults surveyed thought there should be a national testing program to measure the achievement of students. Clearly, test scores are seen as symbols of order, objectivity and important educational outcomes. They are thought to be scientific because they produce a score. The terminology used to describe tests and testing programs is seductive: "competency," "mathematical achievement," and "functional literacy." Words like these demand attention and support. Unfortunately, there is a wide gap between the richness of these terms and what such tests actually measure.

Accountability is an important component of educational reform. Districts should be expected to demonstrate to their residents that their tax dollars are being spent wisely and their children well trained. The key to accountability is to collect relevant information and present it in a useful format.

Ideally, school districts should be ranked not on test scores which are highly influenced by wealth, but on three broad categories of educational quality: access to knowledge, district expectations, and teaching

environment. Access to knowledge is the degree to which a district provides students with opportunities to learn, the number of courses offered, amount of real instructional time, availability of materials, computers, and laboratories; and the breadth of academic support and enrichment. District expectations are the degree to which the district encourages students and staff members to work hard and achieve, graduation requirements, and the ways a district recognizes accomplishments and effort. An effective teaching environment helps teachers and administrators as they try to develop and implement instruction. It is reflected in such things as class size, amount of time available for professional development, staff turnover, the number of staff members continuing their educations, and the degree to which staff members are involved in decision making in their buildings and program development

Educational quality is a complex concept. This year, all of us should temper our reading of test-score rankings with two warnings. First, without procedures to separate the effects of background variables such as ability, poverty, or language proficiency, test scores are not comparable within or across districts. This year's rankings reflect wealth, not effectiveness or quality of instruction. Second, while achievement test scores are easily obtained and reported measures, they may not provide the most useful insights.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR INCOME GROUPS ON READING, LANGUAGE AND MATHEMATICS

READING

Group:	Mean NCE	Std. Dev.:
Group 1	49.854	5.067
Group 2	53.141	4.728
Group 3	55.406	5.954
Group 4	59.194	4.041
Group 5	63.178	4.063

LANGUAGE

Group:	Mean NCE	Std. Dev.:
Group 1	49.281	5.35
Group 2	51.617	5.368
Group 3	54.307	6.649
Group 4	58.527	4.64
Group 5	62.643	5.066

MATHEMATICS

Group:	Mean NCE	Std. Dev.:
Group 1	48.322	6.807
Group 2	51.375	6.139
Group 3	53.489	7.376
Group 4	58.279	5.328
Group 5	64.457	6.232

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR ADC GROUPS ON READING, LANGUAGE AND MATHEMATICS

READING

Group:	Mean NCE	Std. Dev.:
Group 1	55.806	5.672
Group 2	51.938	4.564
Group 3	48.525	4.426

LANGUAGE

Group:	Mean NCE	Std. Dev.:
Group 1	54.48	6.549
Group 2	50.767	4.978
Group 3	48.109	5.026

MATHEMATICS

Group:	Mean NCE	Std. Dev.:
Group 1	54.456	7.391
Group 2	50.127	3.96
Group 3	46.563	5.984