CITIZENS LEAGUE

RESEARCH INSTITUTE

Socioeconomic Disparities and the Comparison of District-Wide Scores on Standardized Tests

PATRICK JOSEPH RYAN

Education Policy Report
July 1999

Socioeconomic Disparities and the Comparison of District-Wide Scores on Standardized Tests

by

Patrick J. Ryan, Ph.D.

This spring the Ohio Department of Education (ODE) released 1999 report cards for all public schools in the state.[1] The news was not good. The state's fourth, sixth, and ninth graders scored a passing grade on about a third of all the proficiency tests that they took. Even when given multiple attempts, only about half of the students succeeded on the 4th grade reading test that soon will be required for promotion, and the 9th grade tests that already are tied to graduation. Large disparities in test scores between districts also are discouraging. In urban districts like Cleveland, Cincinnati, and Columbus only about a quarter of fourth graders were able to read at grade-level. Whereas in wealthy districts like Orange and Chagrin Falls over three-quarters of them do so. These gaps do not seem to narrow with older students. For example, less than six percent of East Cleveland's ninth-graders passed all five elements of the ninth-grade exams, while almost 90 percent of Bay Village's ninth-graders accomplished this task.

Youths from the poorest city districts are about half as likely as other Ohioans to graduate from high school. In Cleveland, the graduation rate runs at an abysmal 38 percent. A Cleveland ninth-grader has about a one in 15 chance of both graduating from high school within four years and passing all elements of the 12th grade exams. The dire social and economic consequences paid by those who face the world poorly educated or without a high school diploma are painfully obvious.

Given these facts, it is common to conclude that inner city schools are failing poor children. The Cleveland public school's new CEO, Barbara Byrd-Bennett, says that the current situation is intolerable, that the district has "failed our children," and that "we can not afford to lose another generation." The residents of Greater Cleveland agree with her and Citizen's League Research Institute polls show that they list education as their top social concern for the region. But, concern leads to change only when the sources of problems are adequately and widely understood. Of course, "the" answers to schooling problems are beyond the scope of this short report, but what follows will help one to more appropriately compare district-wide test scores. Certainly schools must always seek to improve their practices, but our analysis suggests that lower test scores in the biggest cities are primarily a product of urban economic and social decay.

Comparison is a key part of evaluating test scores. We used it throughout the beginning of this report. When we said only a minority of Ohio fourth graders read at grade-level, we were comparing student test scores to state standards. We compared the test scores of wealthy districts to poor ones to highlight the disparities in our educational systems. Though we use it frequently, comparative reasoning is often implicit and thus it can be deceivingly simplistic. For example, our comparison of Cleveland to Orange allowed us to assess the distance between the current educational situation and the ideal that public schools should provide all children with a fair chance in life. This comparison would be misleading, as we shall see, if it were employed to determine whether educators in Cleveland are doing their jobs as well as those in Orange. This is so because there are social and economic factors that greatly influence students that are

beyond the control of teachers and administrators. Controlling for these factors is the key to setting up useful comparisons of district-wide scores on tests. This is why the Ohio Department of Education included data in all its district report cards from so-called "similar districts." The department advises us that "comparing your district's performance to these 'statistically similar' districts is more valid - and more useful - than comparisons to other districts based simply on geography."

This is good advice. But, a short analogy will help us to specify precisely under what conditions comparisons are useful and fair. Take a person who is six feet, three inches in height. Is this person tall? Yes, in terms of the general population he or she would be, but under what circumstances does such a fact matter? Being tall would put one at an unfair disadvantage in, say, a Limbo contest. However, if one wanted to assess the significance of this height for competitiveness among NBA centers, the two operative questions would need to be asked again: How tall is 6'3" relative to the new group? What's the relationship of height to the task? All other things being equal, a NBA center who stands only 6'3" is at a decided disadvantage because he would be short relative to the group, and because height helps one succeed in the game of basketball. This is the unstated logic behind the common practice of creating separate basketball leagues or tournaments for players who are less than six feet tall.

This example illustrates that two questions must be considered when setting-up a meaningful comparison:

(1) What is the strength of a trait in a given case relative to the comparison group?

(2) What is the causal role of that trait for the tasks being compared?

Of course making fair (and thus useful) sets for comparisons of district test scores is more difficult than making height sets for basketball. The ODE uses district size, rural-urban location, tax capacity, socio-economic status (SES), and poverty to calculate similarity. Their method meets the first issue of relative strength in a sound manner, but the ODE does not address the second issue of causality.

Each of the five characteristics that they used are important, but only the last two are moderately to strongly correlated with test outcomes. For example, the number of students in a district may have important implications for management or governance, but in Ohio, district size has very weak and insignificant relationships to test scores. Whereas social and economic indicators are strong, reliable predictors. Half of all the variation in district test scores in Ohio is explained by disparaties in poverty and status. This is part of a larger connection between economics and education. Among 282 metropolitan areas across the country there is a strong correlation between high poverty and low educational attainment.[2]

In fairness to the ODE, it has become an accepted habit to put large urban school districts in a separate category, but this is done because poverty and low status typically run high among public school students in the cities. Since we have measures for these social and economic factors and since we have found no relationship between the size of the student population and test performance, it is difficult to justify including it as a criterion for test score comparisons. A similar argument applies to tax capacity and urban-rural location. Tax capacity is crucial for a district to strategically plan for future revenues, and there may be differences between urban and rural environment that require different support services. Yet, neither tax capacity nor urban-rural location is moderately related to test scores. We also examined indicators such as the percentage of minority students, teacher education, experience, and salaries. We found that poverty and status "explained" (r^2) between four to fifteen fold more of the variance in test scores than these other factors.[3]

To understand the implications of the ODE's failure to meet the test of possible causal relationships for comparisons, it is helpful to return to our basketball analogy. If one assumes that being short is generally related to being at an unfair disadvantage in the game of basketball,

then it makes sense to level the playing field by creating a league limited to those under six feet tall. In effect, the ODE has created leagues where players are not only grouped by height, but also by factors such as bald-headedness, the price of the player's shoes, and beverage preferences. Hairstyles, shoe quality, and drinking Gatorade may be associated with the success of basketball players, but the connection is probably weak and unreliable. Moreover, their possible causal role is less clear than is the relationship between the game and height. Adding weakly related factors to the selection criteria dampens the influence of the relevant ones and this increases the likelihood that unfair comparisons will be made.

Charts one and two show that in terms of social-economic status and poverty, school districts in the cities of Cleveland, East Cleveland, and Youngstown are in a class of their own. These three city districts have far higher proportions of students on public assistance and far lower levels of education, occupation, and income among their residents than other districts in the state. If one combines the two dimensions (status and poverty) into a composite SES-Poverty Index, these three districts are located within the range of what is referred to as "extreme negative values." This means they are more than three times the interquartile range (the length of the distance separating the middle 50 % of all values) below the lowest value within the interquartile range. Standard statistical manuals instruct researchers to check for errors in computation, data entry, and measurement when "extreme values" are encountered. Sadly, this is no research error. It is a true measure of the gross disparity of wealth in America that the children of these cities are forced to face poverty levels that are five to seven standard deviations higher than the state mean. Moreover, when compared with the 31 districts in the state who scored low enough on our SES-poverty composite to be considered "negative outliers" (1.5 times below the interquartile range), Cleveland Municipal School District is a negative outlier among even this poorest - lowest status group of districts.

Because Cleveland, East Cleveland, and Youngstown are "off the chart" in terms of the economic and social factors that influence test outcomes most, they have been placed into questionable comparison sets by the ODE's similarity method. For example, Cleveland has approximately double the proportion of students (59%) whose families are supported by Ohio Works First (formerly ADC) than does Columbus (31%), and Cleveland is a full standard deviation lower than Columbus on the state's scale of SES. Yet, because the ODE gave equal weight to district size, tax capacity, and urban-rural location in their definition of similarity, Columbus is categorized by the ODE as one of the six school districts most like Cleveland. Whereas, the more socially and economically similar (but much smaller) East Cleveland is not included in Cleveland's comparison group by the state.

To their credit, ODE did acknowledge that some districts were so unusual that they could not be easily compared to many other districts. For unusually large districts or ones that were unusual in terms of their overall formula, they lowered the number of districts allowed into the comparison set from the closest 20 to the closest six. Of course, this step does not address the problem of placing weakly associated factors within the similarity formula. Moreover, it is a somewhat arbitrary, if sensible step. Why six districts? It is less arbitrary to allow the distribution values to set the limit as to which districts are comparable. We have done this by including in each comparison set the 20 closest districts within a distance equal to the interquartile range (middle 50 % of all values).

Using the foregoing methods to compare districts substantially alters the assessment of school performance among many well-known districts in Northeast Ohio. The new appraisal holds for 4th, 9th, or 12th grade test results. Chart 2 displays comparison data using 9th grade results because they are the most important in terms of state policy. These new figures suggest that very poor districts like the ones in the cities

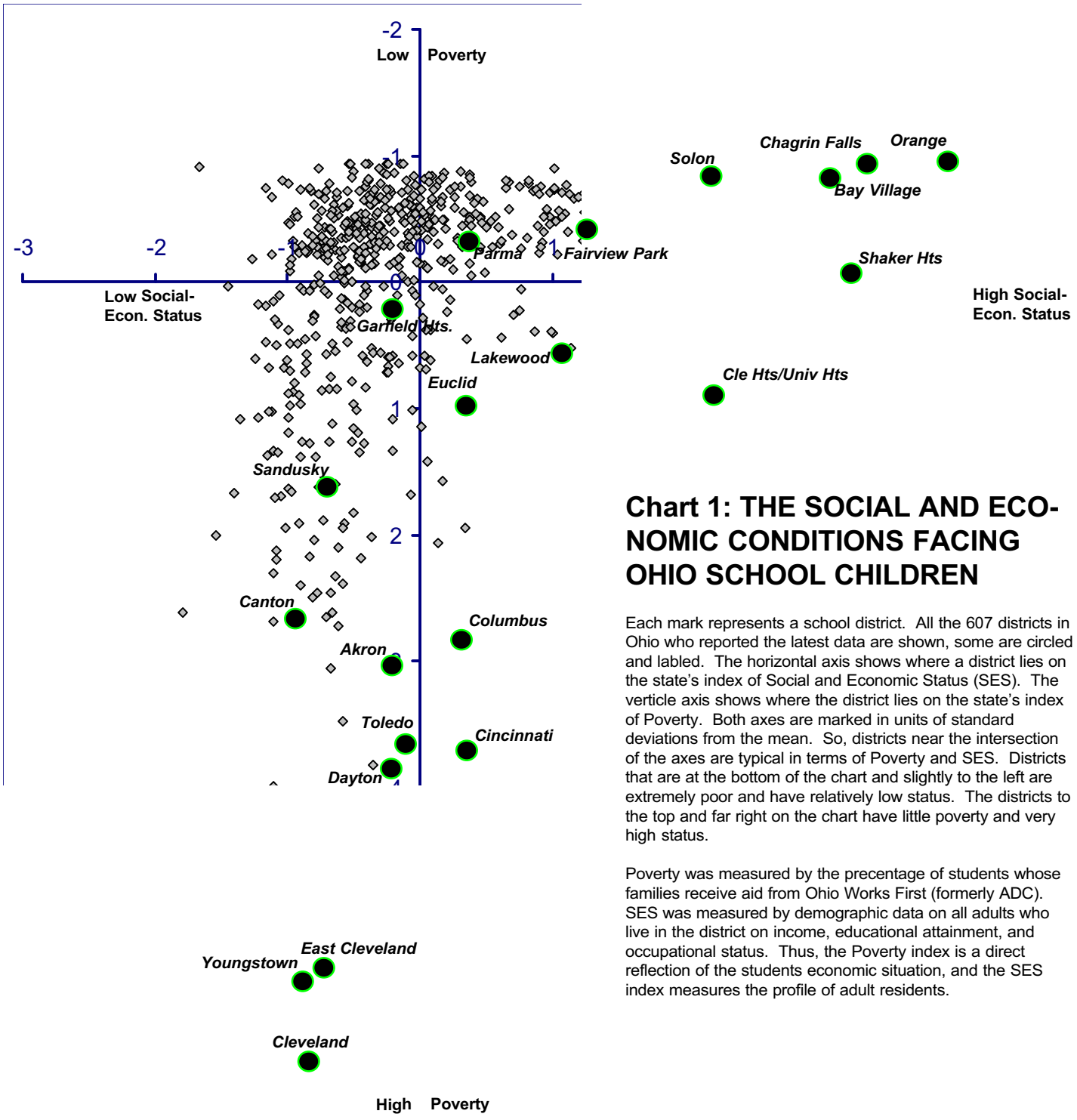


Chart 1: THE SOCIAL AND ECONOMIC CONDITIONS FACING OHIO SCHOOL CHILDREN

Each mark represents a school district. All the 607 districts in Ohio who reported the latest data are shown, some are circled and labled. The horizontal axis shows where a district lies on the state's index of Social and Economic Status (SES). The verticle axis shows where the district lies on the state's index of Poverty. Both axes are marked in units of standard deviations from the mean. So, districts near the intersection of the axes are typical in terms of Poverty and SES. Districts that are at the bottom of the chart and slightly to the left are extremely poor and have relatively low status. The districts to the top and far right on the chart have little poverty and very high status.

Poverty was measured by the precentage of students whose families receive aid from Ohio Works First (formerly ADC). SES was measured by demographic data on all adults who live in the district on income, educational attainment, and occupational status. Thus, the Poverty index is a direct reflection of the students economic situation, and the SES index measures the profile of adult residents.

of Cleveland, East Cleveland, and Youngstown are doing better relative to comparable districts than the state report card would lead one to believe. This is especially true of Youngstown, East Liverpool, and Steubenville. In difficult social and economic conditions these districts do relatively well. Conversely, our calculations suggest that several large city districts including Dayton, Cincinnati, and especially Columbus were significantly over-rated by the state's similarity method. The same was true of the suburban districts of Lakewood, Garfield Heights, and Parma. Among the wealthy-high status districts, Bay Village and Solon recorded strong relative scores, while Shaker Heights, Beachwood, and to some extent Orange, recorded low ones.

The strength of a grouping method is that it allows one to identify the specific districts that one should use for test comparsions. This will aid the dialogue among those interested in school improvement. But, the grouping method is suseptible to a few serious problems that can only be overcome by taking an entirely different approach. First, grouping districts opens the possibility that one unusually low or high scoring district in the comparison set will unduely skew the comparative group mean. This throws off the interpretation of a district's scores. So too, districts on the extremes have very small comparison groups. For example, by CLRI's comparative method only East Cleveland and Youngstown are comparable to Cleveland. The average of two districts can be easily skewed, but the alternative provided by the ODE of setting an arbitary set number at six, as we have discussed, creates its own problems. Fortunately, we can deal with these weaknesses by using statistical regression instead of grouping comparison sets of districts.

Chart 2 shows figures produced through this statistical regression method along side those for the ODE's and CLRI's grouping methods. In short, regression allows us to determine the general relationship between poverty-status indicators and tests scores in the state as a whole. We can then use this relationship to predict what a district should score given its social and economic setting. The difference between the predicted score and the actual score provides an estimate of how well the district scored after taking into account its social and economic conditions.

Chart 3 graphically shows how the regression figures are created and displays the locations of a few dozen districts that should be familiar to residents of Northeast Ohio. Districts located above the regression line scored better than expected, those located below it scored worse than expected.

According to the regression method, three percent more of Cleveland's ninth graders passed all elements of the 9th-grade exams than one should expect given the relationship between poverty-SES and test scores in Ohio at large. Of the major urban districts only Youngstown bested Cleveland with a score of plus 9%. Columbus scored the poorest among large districts with 17% fewer ninth-graders passing all tests than one should expect. Among the districts that score furthest below expectations in Northeast Ohio were Beachwood, Cleveland Hts-University Hts, Euclid, Garfield Hts, Orange, and Shaker Heights. Among those scoring highest above expectations were Steubenville, Independence, Youngstown, Solon, Chardon, New Philadelphia, and Brecksvl-Broadvw. Hts.

That some school districts out-performed other districts facing similar conditions does not mean that residents of those districts should be satisfied with the latest scores. But, it is important to distinguish between testing disparities that result from external factors such as massive social and economic inequalities and ones that may have internal causes such as wasteful management or apathetic teachers. The strong correlation between poverty, status, and scores, suggests to us that the troubles facing urban districts have more to do with a general decay of urban centers than they do with a lack of skill by

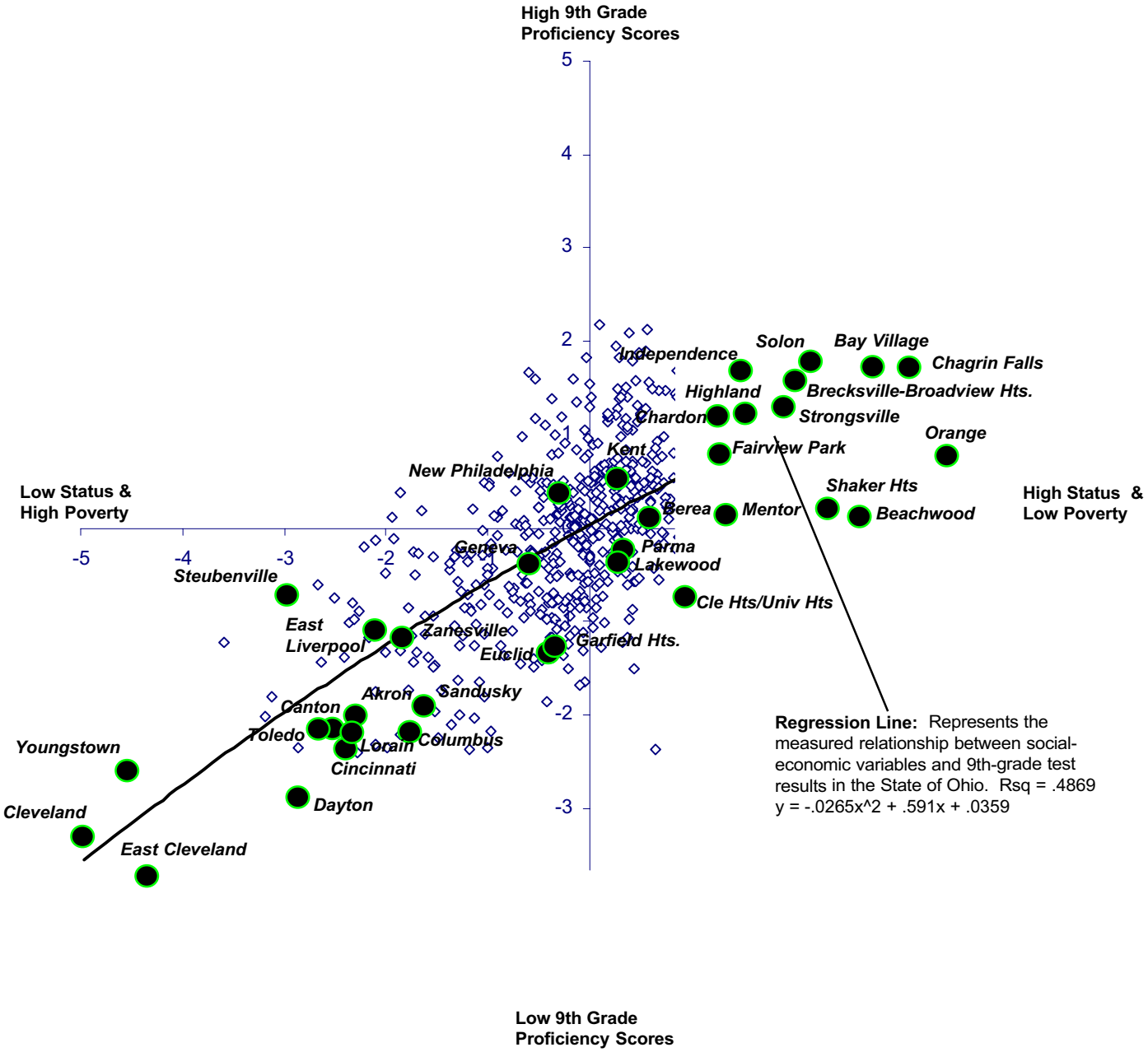
In conclusion, we should return to the basketball analogy and address one of its limitations. In the game of basketball some players are just shorter. We know that we can not make

Chart 2: EVALUATING DISTRICT SCORES

District	SES-POV Composite Index	% Passing All 9th Grade Tests	Comparing 9th Grade Test Results to the Pace Set By Similar Districts			Average Daily Membership
			ODE's Grouping Method	CLRI's Grouping Method	CLRI's Regression Method	
State Mean	0	56%	X	X	X	X
Cleveland	**-4.96**	12%	-11%	-5%	+3%	74,282
The Six Districts "Similar" to Cleveland According to the ODE						
Akron	-2.31	32%	+3%	EVEN	-8%	32,790
Cincinnati	-2.39	27%	+4%	-7%	-13%	48,075
Columbus	-1.78	29%	+1%	-13%	-17%	64,392
Dayton	-2.88	19%	-5%	-12%	-15%	25,790
Toledo	-2.67	30%	+2%	EVEN	-7%	39,010
Youngst'wn	**-4.55**	23%	-3%	+12%	+9%	12,599
Other Districts in N.E. Ohio						
Bay Village	+2.76	89%	+7%	+8%	+3%	2,480
Beachwood	+2.65	65%	-17%	-16%	-20%	1,564
Berea	+0.63	65%	+2%	-1%	-5%	7,965
Brecksvl-Broadvw Hts	+2.04	87%	+8%	+7%	+7%	3,891
Canton	-2.55	30%	-3%	EVEN	-8%	13,117
Chagrin Falls	+3.04	88%	+4%	+7%	+1%	1,885
Chardon	+1.24	81%	+8%	+11%	+7%	3,026
Clev. Hts.-Univ. Hts.	+0.92	52%	-16%	-16%	-20%	7,117
East Cleveland	**-4.35**	6%	-13%	-8%	-11%	6,593
East Liverpool	-2.12	45%	+3%	+14%	+3%	3,449
Euclid	-0.47	41%	-10%	+6%	-18%	6,000
Fairview Park	+1.28	75%	+6%	+5%	+1%	2,093
Garfield Hts.	-0.34	43%	-8%	-16%	-17%	3,364
Geneva	-0.50	58%	-2%	EVEN	-1%	3,269
Highland	+1.51	83%	+7%	+8%	+6%	2,316
Independence	+1.35	89%	+20%	+20%	+14%	952
Kent	+0.42	70%	+6%	+8%	+2%	4,055
Lakewood	+0.33	59%	+2%	-7%	-8%	7,336
Lorain	-2.31	29%	-3%	-2%	-11%	10,710
Mentor	+1.35	65%	-6%	-7%	-10%	10,836
New Philadelphia	-0.20	69%	+7%	+15%	+8%	3,325
Orange	+3.47	74%	-9%	-7%	-16%	2,232
Parma	+0.41	59%	+1%	-7%	-8%	13,071
Sandusky	-1.67	34%	-11%	-2%	-14%	4,582
Shaker Hts.	+2.34	67%	-10%	-15%	-16%	5,574
Solon	+2.15	89%	+9%	+9%	+8%	4,699
Steubenville	-2.96	52%	+10%	+24%	+19%	2,467
Strongsville	+1.89	83%	+7%	+3%	+3%	6,727
Zanesville	-1.85	45%	-1%	+11%	EVEN	4,753

Chart 3: EVALUATING DISTRICT SCORES BY REGRESSION ANALYSIS

This graph allows one to see at a glance how a district is performing while taking into account social economic status and poverty. Districts below the regression line performed lower than expected on the 9th grade exams, those that appear above the line performed better than one should expect given the conditions of the district.



everyone tall, and this seems O.K. because basketball has no monopoly on social life. Education, poverty, and society are infinitely more complex than the simple issue of height in the game of basketball. No child is born with poverty or low status as an inherent trait. The consequences of poverty and low status are social, not natural, facts. We can not eliminate them easily in the way vitamins can prevent Rickets, but neither should we accept them as the natural order of things.

Those who value meaningful freedom through equal opportunity must work to rectify the inequalities faced by our children. It was in this spirit that Ohio Supreme Court Justice Andrew Douglas wrote in his concurring opinion in DeRolph v. Ohio that educational opportunity is a "fundamental constitutional right." Douglas' sentiments are not new. Over the past century and a half public schools have taken a leading role in the establishment of American democracy. But this history should teach us that public schools alone can not level the playing field. It is unreasonable to think that school districts, no matter how well operated or funded, can overcome those enormous social and economic disparities that exist within our urban regions unless school improvement becomes part of a more general effort to renew our cities and to create sustainable growth in the suburbs.

In May, the Citizens League Research Institute released the latest of a series of reports examining the socio-economic dynamics of northeast Ohio. This report showed that across the nation economically and socially weaker regions are ones where larger disparities exist between core areas and suburbs. Citizens who are concerned about Cleveland's schools because they want a society based on equal opportunity should not only join the effort to improve the public schools, but they should also work for balanced economic development, effective transportation, fair labor practices, and accessible health and child care services. All these policy areas influence the strength of families in our region and in turn they influence the degree to which students come to school ready to learn.

NOTES

[1] These were the latest data in a valuable series of statistical reports on schools that the ODE makes available annually to the public through their web site at www.ohio.ode.gov. It should be noted that CLRI reached conclusions similar to the ones offered here in a previous study "Public Schools, 1991" of Cuyahoga County districts.

[2] The 282 Metropolitan Areas referred to are defined by the U.S. Census. For these cities the correlation coefficient between the percentage of the population below the poverty line and percentage of adults without a high school diploma was .6846, p=.0000. Thus, nationwide poverty and educational attainment are strongly related.

[3] Among 607 Ohio school districts, the correlation between average daily membership and the latest available data on 4th, 9th, and 12th grade tests is very weak, sometimes changed signs depending on the test, and probably resulted from chance. (.0057, p=.889 for 4th grd.; -.0959, p=018 for 9th grd.; -.1648, p=.000 for 9th grd. cum.; and .0781, p=.054 for 12th grd.) For the other variables correlations were also weak and an often insignificant. Between 9th grade cum. and tax capacity the correlation was .0729; for urban-rural location it was .0613 ; for % minority students was -.0821; for teacher pay it was .2396; for teacher experience it was .0413; for teacher education it was .1751. In contrast, CLRI's combined index of Poverty and SES showed correlations with the test outcomes that were strong and reliable. (.6845, p=.000 for 4th grd.; .5115, p=.000 for 9th grd.; .6944, p=.000 for 9th grd. cum.; and .7188, p=.000 for 12th grd.)

A strong correlation does not prove conclusively that a causal relationship exists, but there are many studies that detail how and why poverty influences academic performance. We will not enter this complex discussion here because we are satisfied that, although the details can be debated, a strong causal relationship does exist. The weak and statistically insignificant relationships we found for other factors such as minority %, teacher qualities, tax capacity, district size, and urban-rural location do demonstrate the lack of strong causal relationships to test outcomes.

The fact that test data is reported by district, rather than student, introduces what is called the "ecological fallacy." We also only have failure-rate data, rather than actual scores on tests. Until the state approves a method for reporting individual level data, citizens will not be able to benefit from a complete use of analytic techniques.