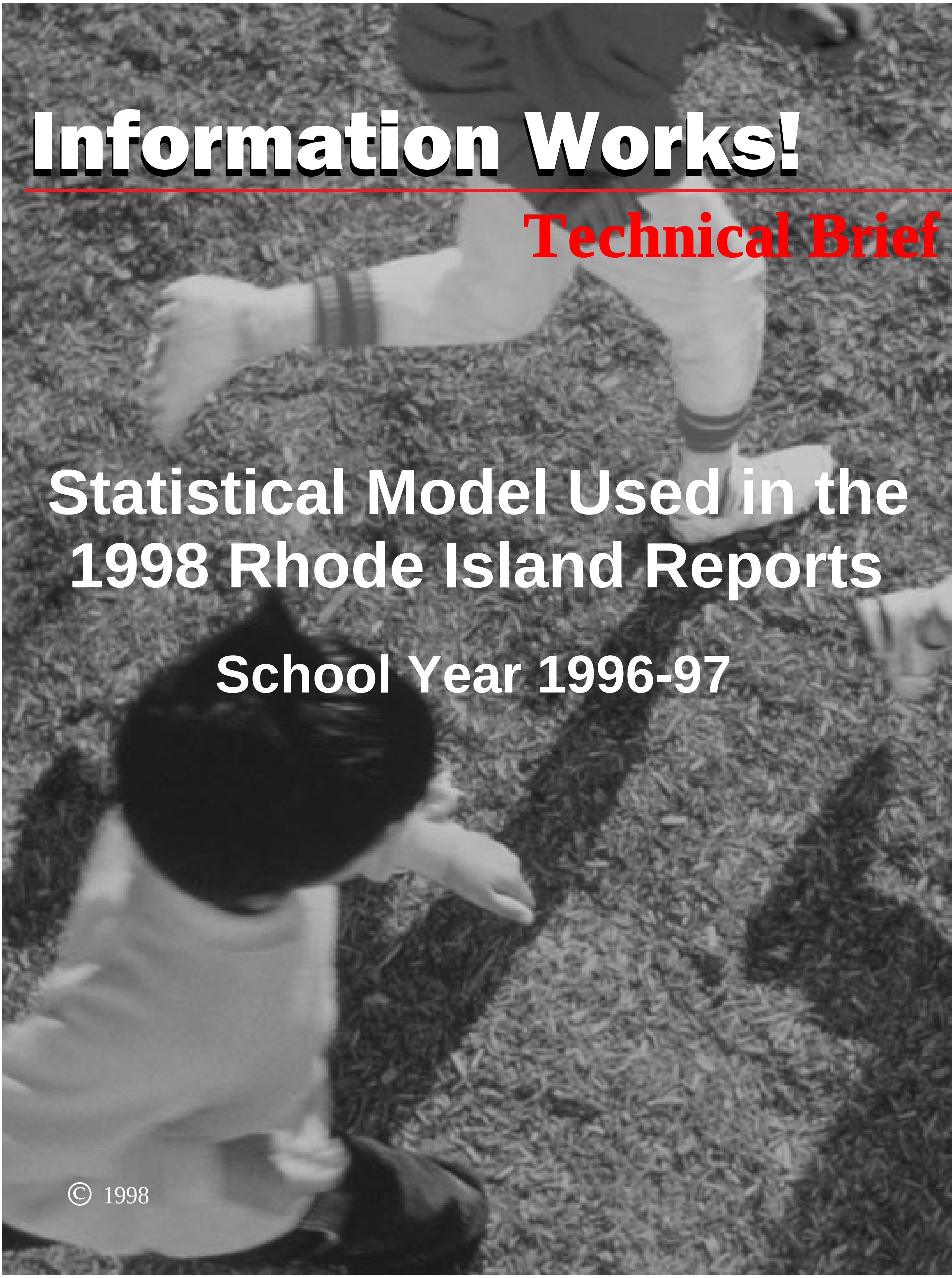

Information Works!

Technical Brief

Statistical Model Used in the 1998 Rhode Island Reports

School Year 1996-97

INFORMATION WORKS!

Technical Brief on the Statistical Model Used in the 1998 Rhode Island School and District Reports (School Year 1996-97)

Across the nation, the public, law-makers and educators have become deeply concerned with finding ways to measure the effectiveness of schools and school systems. There is statewide agreement in Rhode Island that all students need to attain high standards that signal that they are proficient in mathematics, reading, writing, and health as well as other school subjects. The state will annually at a minimum measure student achievement in mathematics, reading, and writing and report out to the general public how many Rhode Island students are proficient in these subject areas.

Politicians and educators have been struggling with developing useful yardsticks for school effectiveness that are honest, accurate and easily comprehended. The initial stages of this effort have often resulted in evaluations that assess, judge and even mete out consequences to schools who've been measured by a small set of benchmarks that often do not reflect the context of the individual school. Many states, including Rhode Island, have published annual state achievement results. People have used these results to rank order districts and schools on the basis of those results as if the schools were competing on a level playing field. While valid for certain purposes, these methods encourage the public to draw sometimes unfortunate conclusions about the value of a school or school system because the yardstick itself is not sensitive to differences in school contexts, to other information about school practices, or to achievement results which are not part of the formal state assessment program.

INFORMATION WORKS!

The intention behind *Information Works!* is to offer a variety of school indicators to help the public and educators understand more completely the schools and the widely varying contexts within which schools, teachers and students operate. The ultimate goal of *Information Works!* -- and indeed Article 31 which created it -- is to assist each school in its efforts to improve student achievement by providing measures that focus on getting all students to the level of PROFICIENT as well as demonstrating less obvious progress toward these desired ends. As Article 31 recognizes, certain schools or districts will need more help and more time to produce optimal student achievement. *Information Works!* supplies information that can help schools by informing their strategic planning and school improvement initiatives. The Comprehensive Education Strategy (CES) states quite emphatically that many things will have to change in order to create high-performing schools for all RI children. Everyone from the schoolhouse to the statehouse will have to get involved if we are to succeed. Certainly, these changes must include the development of new methods for evaluating schools in ways that provide a rich, supportive context for school-based decision making and coordinated state policy changes.

STATISTICALLY GENERATED PERFORMANCE RANGE VERSUS ACTUAL PERFORMANCE

The second field of the Rhode Island school reports presents a school's achievement results on various state tests in a manner different from, but complementary to, the reporting of absolute scores on state achievement tests. The fourth field of the Rhode Island school reports takes the same achievement results and portrays them in yet a third manner – as disaggregated results reported by the different types of students found within a school. A combination of all three ways to view these achievement results leads to a much richer and deeper understanding of how an individual school is doing in helping all its students attain the required level of PROFICIENCY.

In brief, RIDE and researchers at URI's National Center for Public Education and Social Policy (NCPE) have used statewide student information to create within the second field of the report a virtual school, that is, data about groups of students whose characteristics are statistically the same as the ones in a single school report. This second field demonstrates how an individual school is performing while taking into account a number of student variables. In the past schools have been compared solely by raw achievement scores as if the challenges in each school were roughly the same. We know they are not. In the future, as schools improve and as our statistical model is refined with more variables, the "virtual schools" will also change.

All children in all schools can and must meet the Regents' high standards for proficiency. But schools and the wider public need to understand which schools are indeed making progress or are more effective than other schools which are resourced similarly and are facing similar challenges. Over time, this virtual school model will provide an additional measure that helps the state to target investments of both money and human resources to improve schools.

FACTORS ASSOCIATED WITH STUDENT ACHIEVEMENT

A wealth of studies show that family background characteristics are closely related to student achievement. Schools with less economically privileged students, for example, almost always have lower achievement scores. (When RI rank orders their state test results, the results closely mirror the socioeconomic status (SES) of the district; thus, high income districts have high scores and scores drop with a strong correspondence to the relative drop in income.) Changes in many other characteristics (variables) have also been shown by many research studies to correlate closely to student achievement. These include:

1. Prior achievement or aptitude
2. Participation in free and reduced lunch programs
3. Minority status
4. Educational level of the mother
5. Father's occupation

6. Family income
7. Number of siblings
8. Students receiving special services (e.g., special education, bilingual or LEP education).[1]

Additionally, at higher levels of the system beyond groups of individual students, we know that a number of other factors are associated with student achievement such as school settings (urban, rural, suburban), per pupil expenditure, policies and practices within schools or school districts, and community characteristics (e.g., job market, tax support).[2]

Performance indicators of school effects have been systematically collected in a variety of places in the U.S. and elsewhere.[3] Attempts to identify effective schools have created many controversies over the kinds of data to be collected, the appropriate methodologies to be employed and the interpretation of specific results. The first few years of *Information Works!* will probably witness similar controversies. The researchers who constructed the 1998 RI model are not wedded to it. This year's model was based on various data sources that were already available and took into account the strengths and weaknesses associated with each available data set. As both the quality of the data improves and new research is accomplished, the model will evolve and become increasingly sophisticated.

The following few sections of this brief describe some general statistical principles that are important for understanding the statistical model used in this year's *Information Works!* These principles are then applied specifically to the model that was created to generate the second field of the Rhode Island school and district reports. **Readers already familiar with hierarchical regression analysis may wish to skip directly to the sections on MULTIPLE REGRESSION AND THE RI MODEL, starting on page 8.**

---

[1] Cf. Daniel Koretz, "Indicators of educational achievement," In Indicators of Children's Well-Being, Eds. Robert M. Hauser, Brett V. Brown, William R. Prosser. New York: Russell Sage Foundation, 1997, pp. 208-234; Stephen P. Klein et al., "Gender and racial/ethnic differences on performance assessments in science," Educational Evaluation and Policy Analysis, 19(2): 83-98, 1997; Bonnie L. Halpern-Felsher et al., "Neighborhood and family factors predicting educational risk and attainment in African American and white children and adolescents," in Neighborhood Poverty: Context and Consequences for Children, Eds. Jeanne Brooks-Gunn, Greg J. Duncan, J. Lawrence Aber. New York: Russell Sage Foundation, 1997, Volume 1, pp. 146-173; Jeanne-Brooks Gunn, Greg J. Duncan, "The effects of poverty on children," Children and Poverty, 7(2): 55-71, 1997.

[2] David Kaplan, Pamela R. Elliott, "A model-based approach to validating education indicators using multilevel structural equation modeling," Journal of Educational and Behavioral Statistics, 22(3): 323-347, 1997; Garrett K. Mandeville, "The South Carolina experience with incentives," In Midwest Approaches to School Reform, Eds. Thomas A. Downes, William A. Testa. Chicago: Federal Reserve Bank of Chicago, 1994, pp. 69-91; Robert D. Felner et al., "The impact of school reform for the middle years: Longitudinal study of a network engaged in Turning Points-based comprehensive school transformation," Phi Delta Kappan, 78(7): 528-532, 541-550, 1997.

[3] J. Douglas Willms, Monitoring School Performance: A Guide for Educators. Philadelphia: Falmer Press, 1992; Centre for Educational Research and Innovation, Measuring the Quality of Schools. Paris, France: Organization for Economic Co-operation and Development, 1995; Centre for Educational Research and Innovation, Measuring What Students Learn. Paris, France: Organization for Economic Co-operation and Development, 1995; Sandra Black, "Measuring the value of better schools," Economic Policy Review, 4(1): 87-94, 1998.

STATISTICAL SIGNIFICANCE

Researchers want to know and to be able to say with some degree of confidence whether any relationships they have found between various types of data are different from relationships they would find solely due to chance. A measure for the degree of confidence we have in a relationship is *statistical significance*. Most researchers are willing to declare that a relationship is statistically significant if the chances of observing the relationship in the sample are less than 5%, assuming no other factors are affecting the data set. (Statistical modeling is based only on the factors included in the model and by its artificial nature automatically excludes all other factors.) In other words, a relationship is considered to be statistically significant if it appears less frequently than 95% of the relationships among the selected variables we would expect to see just by chance.

Thus, on the second field of the school report charts, the band (range) illustrating the statistical model's projection represents this 95% confidence level -- i.e., that there is no more than a 5% possibility that a school's actual scores would lie outside the band due solely to chance. This confidence level for the model (represented by the range of the band) includes not only actual numerical calculations for scores but also includes statistical errors that are part of the model. (Please note that all models have statistical errors and take them into account during calculations.) However, because this model is based on statistical probability, there is the possibility that a school could lie below or outside the band this year or even in subsequent years solely by chance. One goal of the SALT initiative is to shift the fundamental paradigm of school improvement in Rhode Island toward a blend of a rich variety of data sources for measuring, improving, and judging school performance. This means that the statistical model alone should not be used to assess a school but must be coupled with other data from independent sources that confirm that the results are not due to chance. For example, by using a combination of SALT survey data, observations of independent observers, analysis of selected samples of actual student work and other forms of local student assessment results, an observer could confirm that these "adjusted" assessment results are not due to chance. The RI Skills Commission is working on a similar paradigm shift at the level of the individual high school student in their efforts to design a Certificate of Initial Mastery (CIM) that credentials a student's achievement on the basis of having observed a rich array of data that demonstrates convincingly that the student has the requisite competencies for life, living and employment in the 21st century.

The meaning of statistical significance also holds true for any data results for individual students, groups of students or groups of schools. As a rule of thumb, for example, statisticians routinely conduct their studies accounting for the fact that 5% of *any* data set is likely flawed due to entry or other data errors. Statistical probability underscores why it is always inadvisable to judge schools solely, for example, on the results of state testing programs which by their very nature are limited in time, scope, complexity, and above all else, are themselves subject to the laws of probability and statistics.

Please note that *statistical significance* does not mean that two variables have a relationship that is necessarily more than *statistically* important. For example, a school may sit just outside the top of a band several years in a row. Quite possibly the school's staff and its parents might attach very different values to the three percentage points difference between being within or just outside the band. The school's staff may interpret the percentage points as evidence that they are doing better than other schools and thus do not need significant improvement. The parents might see these percentage differences as evidence of an even larger gap between current proficiency and proficiency on the part of ALL students in the school. Sometimes a very interesting relationship may be missed if it fails to achieve statistical significance and the lack of complimentary observations do not flesh out this subtle relationship. Without a very strong relationship, a sufficiently large sample, or complimentary observations, chance is hard to rule out.

In terms of Rhode Island achievement results, for example, certain schools show disaggregations of student achievement results among groups of students (whites versus other groups, LEP versus non-LEP) which are not statistically significant due solely to the fact that the sample does not include a sufficient numbers of individuals to achieve statistical significance. Conversely, other schools show “gaps” which are statistically significant, but the 2-3 points difference may be educationally unimportant given the possibility of variation in achievement test scores from one day to the next among the same (or similar) group of students. In other words, there is a natural fluctuation in individual (and sometimes whole class) performance due to other factors.

SIMPLE REGRESSION

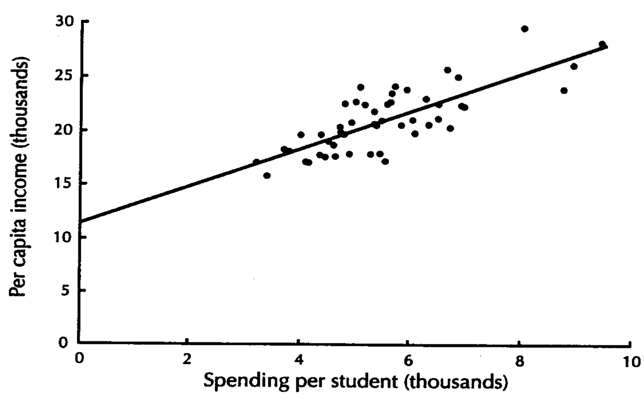

Diagram “A”

The simplest kind of visual description of a relationship between two variables is a straight line. Imagine, if you will, plotting (scatter plotting) a whole set of spending per student data and then drawing a straight line that comes as close as possible to all the points in the scatter plot. (See diagram “A”.)[4] We call this procedure “regression,” the resulting line the “regression line” and the formula that describes the line the “regression equation.” The word “regression” originated from Francis Galton’s work in the late 1800s when he realized that for many relationships there was “regression” (reversion) toward what he termed “mediocrity.” We now express this frequently seen statistical phenomenon as “regression toward the mean.” Human height data, for example, demonstrates that if two parents both have above average heights, their children are more likely than not to have average or below average heights.

[4] Reproduced from Statistics: The Easy Way, Douglas Downing, Jeffrey Clark, Barron’s Educational Series Inc., Hauppauge, NY, 1997, p.256.

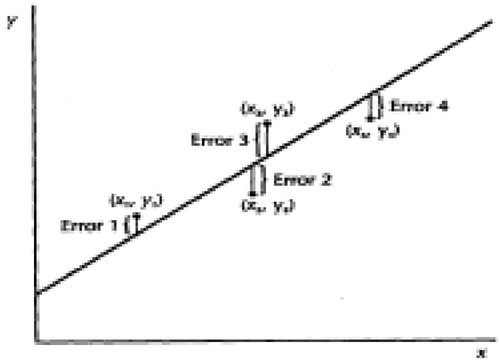

Diagram "B"

Imagine, if you will, plotting achievement scores for grade eight students on a particular achievement test. The vertical axis can be achievement scores recorded as a number. The horizontal axis can be the education level of the child's mother as reported by the child, also expressed as numbers assigned to each level. (Of course, this axis could be any other variable for which you have consistent data which you believe to be reliable). The question is, then, where is the best straight line that relates these two variables (achievement score and mother's education level) to each other? You could take a ruler and try to fit a line through the scatter plot. However, different people would draw different lines, based on their best visual guess as to which line is closest to most of the points. To find the one line out of the infinite possibilities that is as close as mathematically possible to all of the points, statisticians commonly use a procedure called the "least squares line." (See diagram "B".)[5] To determine the least squares line, priority is given to the vertical axis (in this case achievement scores) to calculate how close the points fall to the line. Those distances are then squared and added up for all of the points in the sample. For the least squares line, that sum is smaller than it would be for any other line. The vertical distances are chosen because the equation is often used to predict that variable when the one on the horizontal axis (mother's education level) is known.

All straight lines can be expressed by this formula for the least squares line. The standard mathematical convention is to write an equation for the line relating the two variables as: $y = a + bx$, where y represents the vertical axis (achievement scores in our example); x represents the horizontal axis (education level of the mother in this example), and a and b are replaced by numbers, i.e., two unique constants derived from this particular regression line. The number represented by a is called the intercept and the number represented by b is called the slope. The intercept describes one particular point on the line that falls where the line crosses the vertical axis, when the horizontal axis is at zero. A positive slope describes how much of an increase there is for the variable on the

[5] Reproduced from Statistics: The Easy Way, Douglas Downing, Jeffrey Clark, Barron's Educational Series Inc., Hauppauge, NY, 1997, p.259.

vertical axis (here achievement scores) when the other variable, on the horizontal axis (education level of the mother), increases by one unit. A negative slope indicates a decrease in one variable as the other one increases. Thus, for example, as a school's population becomes poorer in the overall data set of all RI schools (e.g., an increase in the numbers of students eligible for free and reduced lunch), achievement tends to decline (decrease in scores).

CORRELATION

Researchers find it convenient to have a single number to measure the strength of the relationship between two variables and to have that number be independent of the units used to make the measurement. The correlation between two measurement variables is an indicator of how closely their values fall to a straight line. Sometimes this measure is called the "Pearson product moment correlation" or the "correlation coefficient"; sometimes it is simply represented by the letter "r." A correlation of zero could indicate that there is no linear relationship between the two variables. It could also indicate that the best straight line through the data on a scatter plot is exactly horizontal. A positive correlation indicates that the variables increase together. A correlation of +1 (or 100%) indicates that there is a perfect linear relationship between the two variables. As one increases so does the other, proportionally. A negative correlation indicates that as one variable increases the other decreases. A correlation of –1 indicates that there is a perfect linear relationship between the two variables, but as one increases the other decreases.

Correlations of +1 or -1 would be extremely strong relationships. They are rarely observed when exploring relationships between different variables. Even when a perfect correlation is observed, things may not be as simple as they seem. Correlations can be affected by a number of factors. For example, data points that sit significantly outside the rest of the points in a data sample (outliers) will inflate the correlation when it is consistent with the trend (direction) of the data set as a whole. An outlier that is not consistent with the rest of the sample will likewise substantially decrease the correlation. For example, if one or more very low income students in a class submits perfect writing samples on a state assessment, their high scores will substantially suppress the overall correlation between SES and achievement seen in the data set as a whole.

Sometimes outliers occur simply because the data were erroneously recorded. The URI research team used a variety of methods to detect such errors in the data sets that make up the statistical model. They also used these methods to check (or "clean") the other data sets that make up this year's school and district reports. Changing the units of measurement does not affect correlations. For example, the correlation between weight and height remains the same regardless of whether height is expressed in inches, feet, or millimeters. Similarly, the kinds of correlations seen in the RI model hold true whether we are expressing the results as actual numerical test results (raw scores) or as percentage of items correct.

CORRELATION DOES NOT IMPLY CAUSATION

Even if two variables are legitimately related or correlated, there is not necessarily any causal relationship between them. In other words, changes in the one variable may not be directly caused by the independent operation of the other variable. The one may fluctuate in relation to the other due solely to chance (coincidence) or, as is often the case, each is strongly affected by one or more other (confounding) variables that were not considered by the researcher. Other possible reasons include both variables changing over time, one (response) variable causing a change in the other (explanatory) variable or one being the direct cause of the other, and one being a contributor but not the sole cause of the other. In the well-known expression “correlation does not imply causation,” statisticians summarize this understanding of the legitimate use of statistical relationships. In the absence of any other evidence, data from an observational study cannot be used to establish causation.

However, a causal connection probably does exist if we can establish that: 1) there is a reasonable explanation of cause and effect, 2) the connection happens under varying conditions, and 3) potential confounding variables are ruled out. The best way to determine these factors is through a designed experiment in which groups which are strongly similar to one another in terms of certain important variables are exposed to different approaches (treatments) and analyzed to see whether the variable of interest performs differently among the treated groups. One or more groups is also held constant and not subjected to treatment(s) as a “control” group(s).

In the RI model, the relationships between achievement on selected state tests, socioeconomic status (SES), limited English proficiency (LEP) and special needs do not lend themselves to a controlled experiment. Therefore, we should use the results of the model solely to look at aggregate differences in schools educating similar types of students rather than as predictive of actual individual student achievement or even aggregate student achievement for a single school. If careful local study of conditions surrounding student achievement can rule out some of the other factors which might explain the results, we can increase our confidence that any results seen across multiple years <u>may be</u> attributable to the factors which make up the RI model. Taken only by itself, this model, like each of the other data fields in this year’s Information Works!, does not provide sufficient information by itself and should not be the sole means for understanding and judging the complexity of a school, its students, and its results. When the fields are taken together and coupled with other local sources of information about a school, we begin to move out of the realm of speculation about a school and into a data-informed conversation about school contexts and school improvement efforts.

MULTIPLE REGRESSION AND THE RI MODEL

Up to this point, we have discussed the computer-modeling of student achievement in terms of the relationships between only two variables. Research shows that student achievement results are the result of a whole variety of factors ranging from things which are clearly definable and collectable -- like eligibility for free and reduced lunch and receiving certain kinds of special services -- to intangible but important factors such as an individual student’s motivation to perform well on a state test or the general climate of a school. (For example, studies show that most students perform better on state tests when there are local consequences attached to their performance.)

Statisticians over the years have built increasingly sophisticated models to relate various factors simultaneously. One of the most powerful of these methods is Hierarchical Linear Modeling (HLM). When applied to schools, HLM would consider several characteristics of a school as well as several characteristics for individual students. Researchers at URI attempted to use HLM but ultimately rejected it as an approach because our sets of schools which look similar to each other are too small to yield reliable results. HLM can predict scores for individual students with certain characteristics, but

Information Works! has no intention of focusing on individual students (teachers, or administrators). The school as a whole is the important unit of analysis and improvement for state accountability purposes.

Instead, the URI research team used hierarchical regression analysis, which is a specialized form of multivariate analysis.[6] Multivariate, like its name implies, looks at how multiple variables acting separately or combined in various ways, impact on the variable of interest (in this case student achievement on selected state tests). Broadly speaking, multiple regression analysis is a method of analyzing the variability of a dependent variable by using information available on a set of independent variables. Unfortunately, no drawing can illustrate the relationship the way the simple regression model was illustrated earlier in this paper.

Five independent variables, known in advance likely to relate statistically to student achievement on state tests, are collected annually from Rhode Island schools and students. These variables are:

1. The percentage of students within a school eligible for free or reduced lunch
2. The percentage of minority students (i.e., non-white) within a school
3. The highest education level for the child's mother as reported by the student
4. The percentage of students in a school enrolled in LEP or bilingual education programs
5. The percentage of students within a school receiving services under special education law

Because of the small number of schools in Rhode Island, comparisons were made among all schools in the state rather than just among groups of schools with similar demographic characteristics. The overall results of each of these variables viewed individually is shown in Table 1 which demonstrate overall a strong relationship between these five student characteristics and academic performance across all grade levels of the state assessments.

The first three variables listed above were found to have "multicollinearity." In other words, when one variable shifted, the others also shifted in similar manner. The correlations between the percentage of students eligible for free and reduced lunch and percentage of non-white students in a school, for example, was greater than .9 across all three tested grade levels. (Recall the discussion above about correlations approaching 1.00). One of the remedies for multicollinearity is to group variables in blocks. (We elected not to use factor analysis since it does not lead to a numerical result that can be easily understood by non-statisticians.) Therefore, equally weighted averages of the variables of eligibility for free and reduced lunch, mother's level of education, and minority status were used as a single block. For mother's education, the percentage of mothers whose education was reported as beyond high school was recoded to run in the

[6] For further information consult George A. Marcoulides, Scott L. Hershberger, Multivariate Statistical Methods: A First Course. Mahwah, NJ: Lawrence Erlbaum Associates, 1997; J. Scott Long, Regression Models for Categorical and Limited Dependent Variables, Thousand Oaks, CA: SAGE Publications, 1997; and one of the classics in the field, Jacob Cohen, Patricia Cohen, Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 2nd ed., 1983. A good general discussion of modeling is David W. Britt, A Conceptual Introduction to Modeling: Qualitative and Quantitative Perspectives. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

same direction as the other two variables, which is to say that the lower the education level of groups of mothers, the lower the educational achievement of corresponding groups of students. The combined equally weighted variable that results can be thought of as a poverty index (low SES) which is more stable across grade levels than any of the three variables viewed individually. As the index number increases in size, the more poor students there are within the school.

A second block was created using the percentage of students receiving bilingual or LEP services and the percentage of students within a school receiving special education services These variables were each introduced separately into the model but after the application of the SES variable described above. The researchers are aware that some students within a school receive both types of services and would be counted in this model twice. Students with multiple learning needs require more support, which in turn factors in the cumulative effects of multiple challenges.

Due to the limitations inherent in the relatively small data sets (the small number of schools in RI and the small number of students per tested grade level in some RI schools), the researchers chose to use overall building level variable data rather than data associated solely with the grade tested. So, for example, we used eligibility for free and reduced lunch data for the entire school rather than just the grade tested. Empirically, while these data sets were highly correlated, the researchers were more interested in overall school context rather than just the context specific to particular grades, consistent with RI's focus on school (not grade-level) accountability. This signals, for example, that grade four student achievement is not only the responsibility of the teachers and administrators specific to that grade, but is also conceptually and educationally linked to learning experiences in the prior grades. Table 2 shows the basic descriptive statistics for RI schools taken as a whole with the number of schools, range (how wide the scores were), mean (arithmetic average), and standard deviation (a measure of how spread out the scores are).

The dependent variables in the model are student academic achievement in each grade level tested across different subject areas. Researchers employed a separate regression model for each subject area (mathematics, writing, reading, and health) in each tested grade level (grades 4, 8, and 10). For both the New Standards Reference Examinations and the Rhode Island Writing Performance Assessment, the percentages of students who achieved proficiency or above was computed and used as the dependent variables of interest. Mainly because of space limitations, the school reports show only the results of New Standards Mathematics Skills and Problem Solving, rather than all three components. Problem Solving was selected over Concepts because it represents a more complex set of skills. For the Metropolitan Achievement Test in Reading, national percentile ranks were used as the dependent variables because the test was not criterion-referenced, as are the others. Regression models were based on a total of 43 schools for grade 10, 53 schools for grade 8, and 183 schools for grade 4.

Table 3 shows the results from the hierarchical regression analyses. Across all grade levels, the school SES variable produced significant effects in relation to student achievement. Schools with more economically disadvantaged students have lower student achievement across all subject areas and grades. Once the variation in school

SES is accounted for in the model, indicators of school context related to special needs (special education services and LEP/bilingual education programs) rarely have significant effects on student achievement. It appears that significant and negative effects in relation to student achievement become stronger among schools with higher percentages of special needs students as students get older. The last column of Table 3 represents the proportion of variances (fluctuations) in student achievement scores that were "statistically" explained (rather than "causally" explained) by the independent variables in the model. The model, for example, explained over 80% of the variances in student achievement in reading, across all grade levels. It appears that variances explained by the model get smaller for younger students (with the exception of MAT Reading), as the relationship between school SES and elementary student achievement is less prominent.

PREDICTING INDIVIDUAL SCHOOL PERFORMANCE

Regression equations produced by the models were used to create expected achievement scores (dependent variables) on the basis of selected known values of the school context (independent variables). These expected scores based on school context derived from the entire set of state achievement tests in school year 96-97, provide us with guidelines about what we can statistically expect for individual school performance in the 96-97 school year state assessments. Note that we cannot reliably predict the 97-98 state assessment performance from the 96-97 data. The techniques applied here only work in retrospective analysis. Standard (random) errors were computed in addition to the specific point estimates for the expected scores. The point estimates combined with the errors generated a band within which we would expect individual school performance to lie. The band reflects a confidence level of 95%, i.e., there is only a 5% or less chance that a school's performance would sit outside the band due solely to chance. Because of employing the 95% confidence intervals, schools with both a preponderance of the five variables we considered in our two block model and those who have small numbers for those variables within their entire student body have much larger bands. These schools at the two extremes are prone to more errors (recall the outlier discussion above), because they are outliers in the total population of RI schools. In other words, they are least like the other schools in the entire RI sample.

A VISUALIZATION OF THE RI MODEL

While we can not draw a graph for multivariate analysis like one can do with simple regression models, perhaps a virtual (spatial) image of the RI model would be helpful. Imagine a large empty room with transparent walls. Inside the room is a set of thousands of small balls suspended in space and stretching unevenly across the room. These balls represent the actual scores of all students in one particular grade who took one of the state tests. When you look into the room from the North (SES) side you notice that the balls spread across the room are variously tinged with red ranging from brilliant red (very poor) to very light, almost non-existent red (very wealthy). You note that most of the brilliant red balls are concentrated in one end of the room and most of the very light, almost non-existent red balls are concentrated in the far other end of the room. However, you do note a few brilliant red balls scattered here and there throughout the room, including even the far end where the very light balls hold sway.

Similarly, as you move to the West (Special Education) wall and peer into the room, you now notice that the balls viewed from this angle are tinged with two types of blue The deep blue balls represent students receiving services under special education law and the light blue represents students not receiving any special education services. Once again you note that most of the deep blue balls are near the same location in the room although there are exceptions. Some light blue balls can be found throughout the room, but most are concentrated toward one end of the room.

Now you move to the East (bilingual and LEP programs) wall and notice that there are two kinds of purple balls – dark purple (receiving services) and light purple (non-recipients). You note dispersal patterns among the purple balls similar to what you have noted with the previous colors.

As you now move to a platform to peer down into the room from the top, you now see that all the balls are colored from a bright yellow (very high achievement score) to clear (a zero). However, you note a crucial difference from what you have seen in the room from the other two walls. All of the bright yellow balls are in one end of the room and all of the clear balls in the other. In fact, balls with identical or very similar shades of yellow are uniformly found in the same approximate location within the room.

A person now opens the door into the room with a large transparent cube with colored balls in it speckled with the blue and red hues you have already observed. Each ball is also uniformly colored green (to represent the fact that these are all students from the same school). He releases the balls into the room from the cube and they magically head to various locations within the room. You note from your vantage point at the top of the room that they all have arranged themselves at varying places within the room with some concentrated at particular places and some existing solely by themselves with no other green balls anywhere near them. As you descend to the West (special education) wall, you see that nearly all the deep blue balls just released into the room have arrayed themselves in the end of the room where most of the deep blue balls are concentrated. Similarly, as you move to the North (SES) wall you notice that most of the brilliant red balls from this school have arrayed themselves with their counterparts (other very poor students from across the state) and the East (bilingual/LEP) wall similarly. Finally, as you move to the South (school) wall, you notice clearly that you can easily pick out where all the green balls are within the room. In fact you discover that you can count how many are in each section of the room. The majority of balls representing student achievement scores on this particular test from this particular school are concentrated within a given area of the room. The person who released the balls now explains to you that what you have seen is a statistical model of last year's student performance for the state. Each time they release colored balls representing an individual school within the room, 95 times out of a hundred, the balls end up in the approximate location where you now see them. "Is it chance, or just coincidence?," he asks you. You reply, "I will have to go to that school and see what I can find out."

ENDNOTE

The current RI model is by no means a finished product. In fact, the research team welcomes the advice and suggestions of other researchers or those community members with a bent toward statistically thinking as we seek to improve this model for next year. Already we are engaged in conversations with other localities that administer the New Standards Reference Examinations about sharing with us their 1998 academic achievement and demographic data. If we are successful in this venture, next year we hope to have enough schools in the entire “national” sample to apply the Hierarchical Linear Modeling (HLM) technique which was unavailable to us this year because it requires a bigger and more diverse set of schools than ours to accomplish the analysis.

Table 1. Student Characteristics Related Independently (Singly) to Student Achievement (Correlation Coefficients)

		% students in free/reduced lunch	% minority students	% students w below college Mother's education	% students in LEP	% students in special education
Mathematics						
Gr 10	Skills	-.851**	-.878**	-.758**	-.654**	-.173
	Concepts	-.705**	-.690**	-.764**	-.508**	-.181
	Problem Solving	-.722**	-.704**	-.771**	-.508**	-.131
Gr 8	Skills	-.755**	-.738**	-.604**	-.622**	-.124
	Concepts	-.706**	-.586**	-.705**	-.536**	-.238
	Problem Solving	-.575**	-.546**	-.516**	-.492**	-.251
Gr 4	RI Math	-.558**	-.462**	-.538**	-.404**	-.030
Writing (RI Performance Test)						
Gr 10		-.659**	-.656**	-.739**	-.534**	-.332*
Gr 8		-.818**	-.744**	-.694**	-.650**	-.232
Gr 4		-.467**	-.438**	-.379**	-.360**	-.077
Reading (Metropolitan Achievement Test)						
Gr 10		-.843**	-.886**	-.800**	-.729**	-.204
Gr 8		-.888**	-.870**	-.826**	-.793**	-.192
Gr 4		-.883**	-.818**	-.731**	-.703**	.077
RI Health						
Gr 4		-.414**	-.416**	-.341**	-.360**	-.040

Table 2. Basic Descriptive Statistics for All RI Public Schools

		N	Minimum	Maximum	Mean	Std. Dev
Dependent variables						
Mathematics (New Standards Reference Exam)						
Gr 10	Skills	43	21.7	89.5	60.99	18.64
	Concepts	43	0	47.6	17.30	12.19
	Problem Solving	43	0	62.5	24.08	14.49
Gr 8	Skills	53	14.6	86.5	55.30	17.47
	Concepts	53	0	50.3	17.77	11.30
	Problem Solving	53	0	46.2	18.51	11.95
Gr 4	RI Math	182	0	50	14.04	11.73
Writing (RI Performance Test)						
Gr 10		43	7.8	70.5	32.58	14.85
Gr 8		53	1.9	73.4	35.05	15.28
Gr 4		182	0	67.9	13.08	10.83
Reading (Metropolitan Achievement Test)						
Gr 10		43	19.07	68.26	48.53	13.82
Gr 8		53	19.09	69.74	50.41	12.58
Gr 4		182	14.87	75.83	53.27	13.59
RI Health						
Gr 4		182	0	46.9	11.96	10.06
Independent Variables						
% students in free/reduced lunch program		269	0.56	100	34.95	29.23
% non-white students		269	0	97.25	16.60	23.49
% students with Mother's education below college level		264	0	100	68.46	17.11
% student in bilingual or LEP programs		269	0	47.47	5.75	10.84
% students receiving special education services		269	0.46	100	16.38	6.74
Combined school SES		264	2.7	83.04	27.83	21.22

Table 3. Results from Regression Analyses for All RI Public Schools

		Constant		School SES		% LEP		% Special Ed		R^2
		Coeff.	S.D	Coeff.	S.D	Coeff.	S.D	Coeff.	S.D	Explained variance
Mathematics										
Gr 10	Skills	91.955**	4.251	-.988**	.092	.256	.193	-.624*	.300	.859
	Concepts	37.008**	4.346	-.616**	.094	.324	.198	-.473	.306	.656
	Problem Solving	45.580**	5.047	-.758**	.109	.420	.230	-.368	.356	.672
Gr 8	Skills	77.446**	7.781	-.694**	.119	.016	.237	-.214	.484	.651
	Concepts	37.970**	5.687	-.410**	.087	.024	.173	-.575	.354	.553
	Problem Solving	39.197**	6.281	-.398**	.096	.037	.191	-.660	.391	.490
Gr 4	RI Math	26.389**	3.124	-.338**	.051	.072	.097	-.206	.166	.340
Writing (RI Performance Test)										
Gr 10		62.094**	5.387	-.641**	.116	.197	.245	-1.167**	.380	.644
Gr 8		62.812**	5.824	-.582**	.089	.061	.177	-.694	.362	.725
Gr 4		25.133**	3.153	-.242**	.052	.061	.098	-.310	.167	.243
Reading (Metropolitan Achievement Test)										
Gr 10		71.638**	2.852	-.654**	.062	.041	.130	-.553**	.201	.885
Gr 8		71.241**	3.145	-.437**	.048	-.276**	.096	-.447*	.196	.878
Gr 4		71.134**	2.014	-.548**	.033	-.045	.062	-.122	.107	.807
RI Health										
Gr 4		20.632**	2.924	-.179**	.048	.058	.090	-.208	.155	.203