DISCUSSION

James McPartland, The Johns Hopkins University

The Parke and Sheldon paper raises many important points concerning the relationship of social statistics and public policy. I would like to give emphasis to some of their ideas by applying them to a different example: the state and local educational accountability and assessment reports which have developed in the past couple of years and seem to promise to become quite widespread in the next several years.

State Educational Assessment Reports

A recent survey by ETS finds that 30 states are now operating educational assessment programs.¹ Sixteen are required by state educational accountability laws to do so. In the other states, the idea was introduced by the statisticians and professionals in the education agencies. These assessment programs and reports use standardized achievement tests and exercises for students, sometimes tests constructed by the states themselves.

Generally, the reports of the test results are presented in one of two ways: either (a) unadjusted school or district mean achievement scores, or (b) adjusted residual mean achievement scores (adjusted for student inputs via measures of student socio-economic status, previous year's test scores, or some so-called ability test).

Applying Four Ideas from the P-S Paper to this Example

There are four points from the Parke-Sheldon paper which I will apply to this example.

First, and most obvious, this example shows the need to "value the importance of policy problems of states and localities" as stated by Parke and Sheldon. These local laws and activities have come about because of the clear need for better information to justify and allocate the local monies which go to public education. (This expenditure is now more than 40 percent of total state and local outlays, on the average.)

However, this need is not likely to be solved in this case simply "through projections of (federally collected) social data for areas below the national level," but seems to require the actual development at the local level of separate statistical collection and reporting activities. In other words, the example of educational accountability emphasizes the need expressed in Parke and Sheldon, but also suggests that not all of the "action" is at the federal level, with federal statistical results projected and reported at the local level. In this example, while national programs on school tests are helpful (such as National Assessment of Educational Progress or the development of the "anchor test", which are partially supported by federal agencies), it seems for political and practical reasons that the local agencies

are needed to do the real work of developing statistical indicators and report formats. At any rate, much of the leadership in this development has come from statisticians working on the state and local level.

A second need expressed by Parke and Sheldon --to value "the role of public understanding in policy development"--and their prescription for statisticians to develop indicators and reports accordingly, is of great importance in the state assessment example.

Each of the two approaches to state educational assessment I mentioned (unadjusted school means and means adjusted for student inputs to schools) are usually faced with erroneous public interpretations unless special precautions are taken in reports.

The first kind of statistic (the unadjusted school means) tells something about the general level of learning in the school-aged population; and a comparison of schools on this measure locates where the poor students go to school. But this unadjusted statistic does <u>not</u> say anything about program evaluation, *i.e.* which schools are doing the best job and why. This is because the average student achievement scores in a school have been found to be more a function of family background and previous experiences of the students, rather than a function of the school's program itself.

As obvious as this may be to this audience, a large fraction of the public (and legislators) misinterpret the unadjusted scores as saying more about the quality of various school programs than about where the poor students happen to go to school. So, statisticians working or advising on the state reports need to be careful to prevent this misinterpretation, perhaps by presenting unadjusted results only in aggregation above the school level. In doing so, the reports educate the public that schools are only one of several influences on children's learning.

The second kind of statistic (adjusted residual school means standardized for student inputs) must face another set of public misunderstandings. One difficulty arises from popular assumptions about the particular input variables which should be controlled. A second problem comes from the degree of precision which may be assumed for the results.

Testing specialists generally agree that most of the so-called ability, aptitude or IQ tests which are administered to large groups are simply <u>other</u> achievement tests (that is, there is no basis on which to choose one test as more indicative of native aptitude than another, even though one test may carry the label "aptitude" or "ability" test). This means that one should <u>not</u> use a concurrent "ability" test as an input control in analyzing residual achievement influences of school programs.² Nevertheless, the essential similarities of many "ability" and "achievement" tests are not well known by the public or by **many** educational professionals. Unless statisticians and measurement specialists assert their proper role in developing report and analysis methods, essentially technical questions will be decided politically--by votes of lay advisory committees--and important public misunderstandings will go uncorrected.

In addition, the public tends to make a great deal over small differences between schools in adjusted scores, even though many of the differences may be well within the range of likely random error. As suggested by a recent RAND report on educational outcome measures, this kind of public overinterpretation can be minimized by using a coding scheme of broad categories to present results, rather than suggesting a psuedo-exactness by employing continuous cardinal residual measures.³ In RAND's words, "Crude measures should be employed crudely."

Finally, the example of state activity in school assessment serves to emphasize two other ideas offered by Parke and Sheldon. Now, I refer to their comments on "measures which derive their significance from models of social processes" and also their remarks concerning reports and indicators which narrow the range of options rather than point definitively to a single policy choice.

Underlying the approach using adjusted means is a general model of student input - school learning processes - and learning outputs, but in a simplified version. A major goal in using the adjusted means for a comparison of school programs is to learn which aspects of the total instructional program work best for specific learning outcomes: for example, is reading most affected by the individualization of assignments, by the kinds of staffing, by the way grades and rewards are managed, by the grouping of students? In general, what elements of the school's program are most important for specific outcomes, and thus require more emphasis and future investment in the least successful schools?

To truly answer this question analytically requires a complicated causal model that includes all student input factors, plus all the possible school factors that may distinguish one school program from another, expressed for each separate kind of learning. Clearly, such a specific model is not likely at present: it is beyond our present knowledge to specify all the variables, and the various factors are probably not statistically distinguishable across schools but are highly correlated with one another. Even if we knew the variables, the data requirements to implement such a model are beyond the capacities of most state and local agencies to collect and analyze.

So, in those state plans that include residual adjustments for student input, a more simplified model is used to initially narrow the range of possibilities to explain present differences in school effectiveness. This simplified model seeks only to *identify* exemplary schools by name, but not (at first) to specify the elements of their programs which make them exemplary. (The simplified model is to regress achievement on student background factors, and then identify schools which are consistently above or below the expected level on several indicators.) Thus, a general social model is in mind in developing the approach, but to begin with, a simplified model is used and the information gathering process is divided into separate stages.

After the exemplary schools have been identified, data can be collected formally or informally about how their school programs and staff are actually different from the rest. This second step provides some direction to program changes which might be attempted and evaluated in the least successful schools, and incorporated in more sophisticated models of school effects for later reports.

Thus, both of the points raised in the Parke-Sheldon paper (using social models, and the virtue of narrowing the range of possibilities through practical statistical indicators) seem to be a major part of some state and local educational assessment programs.

My comments on the Parke and Sheldon paper reflect that many of their main points on the relationship between social statistics and public policy seem to apply very well to some important recent state and local policy concerns and statistical projects in the field of education. The example I chose also suggests that federal statistical agencies may not be the only locations where important developments are underway: state agencies and localities also are at work on the problems of social statistics and public policy.

Footnotes:

- 1. <u>State Educational Assessment Programs</u>, <u>1973 Revision</u>. Princeton: Educational Testing Service, 1973.
- Dyer, Henry S., "Toward objective criteria of professional accountability in the schools of New York City," <u>Phi Delta Kappan</u>, 52, 1970, 206-211.
- 3. Klitgaard, Robert E., <u>Achievement Scores</u> and <u>Educational Objectives</u>. Santa Monica: Rand Corporation, 1973.