Causal Inference in the Social Sciences:

Variance Theory, Process Theory, and System Dynamics

Don R. Morris
Miami-Dade County Public Schools
1500 Biscayne Blvd., Ste. 225, Miami, Florida 33132
Tel: 1 305 995 7531 / Fax: 1 305 995 7521
donr.morris@worldnet.att.net

The social sciences are in need of an alternative to the variance approach to causal inference, which—because it requires a counterfactual—restricts the claim of valid inference to conclusions drawn from experimental and quasi-experimental methods. Process theory proposes an alternative by way of demonstrating the presence of observable characteristics of the causal mechanism, a method that, while accepted in principle, has proven elusive in practice. It is suggested that system dynamics can improve the process approach, and in so doing open a path for itself to wider application in the social sciences. Educational research is suggested as a place to start, and results from two models drawn on topics from that field are offered as examples.

K**eywords:** **causal mechanism; causal process; educational research; modus operandi; nonexperimental methods; program theory**

I have read with much interest Nelson Repenning's (2003) account of his experiences in trying to introduce system dynamics to the other social sciences. I am encouraged by his attempts, and impressed by his candor. The rejection of contributions to the journals of the various social sciences is something many have experienced, even while articles by their own members embrace some of the very points that system dynamicists are trying to make. Repenning would cast a wider net than Radzicki (1990), who has sought closer ties with economics via a merger with its minority institutionalist faction. Repenning has recommended seeking opportunities among other of the social science disciplines that may hold out more favorable conditions for collaboration, applying a strategy of gentle familiarization to system dynamics via highly simplified models.

I applaud both efforts as much-needed steps toward greater collaboration and integration with the wider social science community. There is however, a key dissimilarity between system dynamics and other social sciences that is consistently ignored. System dynamicists differ greatly in the way that they view causality. System dynamicists take causality for granted, accepting it at face value. Richardson, for example, well aware of the historical controversy surrounding the concept, expressed the attitude this way: "I choose simply to presume that the concept of cause in the social and policy sciences has meaning, from which we can derive a meaningful idea of closed loops of circular causality" (1991, p. 8).

Because they take cause for granted, many system dynamicists fail to appreciate the very different role that causality plays in other social sciences. Causes are not considered directly accessible by most social scientists, and consequently cause is the central concern around which analyses are constructed. The Humean position (that causes cannot be observed) is generally accepted, and this acceptance has led to the view that the core purpose of scientific inquiry is the detection of causes. This in turn has resulted in a hierarchy of methods—at the apex of which is the randomized experiment—in which the only scientific analyses are the ones in which the variable(s) presumed to be causal can be manipulated. All other methods are considered less than scientific, and unless applied in support of experimental or quasi-experimental designs, the resulting analyses are considered "merely descriptive," or worse.

Opposition to this view of causation has been growing, and numerous factions have formed in all the social science disciplines, most of them opposed to the second class status allocated to nonexperimental methodologies. The strategy I propose for closer collaboration between system dynamics and other social sciences is to introduce a role for system dynamics methodology at the core of the social science purpose—causal inference—and by doing so exploit the methodological disputes that now infect all the social sciences. There are fields where these divisions are critical and the protagonists more evenly matched, and it is here that I believe system dynamics will find the greatest opportunities. Educational research is a discipline highly fragmented by this division, and in this paper I will draw on that field, together with aspects of the closely allied field of program evaluation, to indicate where opportunities for system dynamics may lie.

Inferring Cause: Variance Theory vs. Process Theory

There is a division that cuts across every social science—a dispute over what constitutes "doing science." At the heart of that dispute is a methodological disagreement over how we are able to determine—in a scientific sense—the reasons why things happen, what causes them. There are those who contend that science and the experimental method are one and the same, while others just as passionately maintain that other methods are equally scientific.

The defense of the experimental method rests on what has come to be called the "variance approach" to causal inference. The approach is deductive, and the result presumed conclusive. One deduces from theory the cause(s) of events in the ideal case, and constructs experiments in which the (presumed) causal variable can be manipulated such that results when the variable is present can be compared to results in its absence. The approach ignores the concept of a causal mechanism, since causes are considered unobservable. The analysis is of the "black box" type, in which the presence of cause is demonstrated by a systematic relationship between inputs and outputs.

Where experiment is not possible or feasible, data is collected with the intent of demonstrating that the effect is present when the cause is present, and not otherwise. This approach—called quasi-experimental—lends itself well to statistical methods, and in many social sciences structural equation models or their equivalents have become synonymous with causal inference. See Berk (1988) for an overview from a sociologist's perspective. There has been a longstanding uneasiness with correlational inferences of cause, particularly with complex statistical models, in educational research and elsewhere (see especially Freedman, 1987). Berk has noted that it is not that the statistical models are so good but that—on the assumption that there is no alternative to the variance approach—there is nothing better available.

An increasingly articulate defense of nonexperimental research is emerging (e.g., Johnson, 2002), and with it a contrast of approaches to causal investigation, a contrast well articulated for educational research by Maxwell (2004). The nonexperimental approach to causal inference is often referred to as the process approach, and the variance vs. process distinction has appeared in various disciplines; see for example Scott (1994) in sociology and Mohr (1982) in political science. In the process approach one observes an effect and identifies the suspected cause(s), but does not have the option of manipulating the causal variable(s), and/or does not have access to comparative data. Consequently, one draws instead on theory and/or experience to describe the mechanism by which the effect is thought to be caused, emphasizing the causal process.

The dispute is cast, then, in terms of a variance approach vs. a process approach, as Scott (1994) has termed it (the original terminology is attributed to Mohr, 1982). In Maxwell's words: "Process theory . . . deals with *events* and the processes that connect them; it is based on an analysis of the causal *processes* by which some events influence others" (2004, p. 5). A process approach requires a process-oriented conception of causal explanation, what Maxwell (2004) calls a realist causal approach, and Cook (2002) refers to as explanatory theories of cause. Maxwell asserts that (among other things): "A realist, process-oriented approach to explanation . . . . recognizes the explanatory importance of *context* of the phenomena studied, and . . . . relies fundamentally on an understanding of the *processes* by which an event or situation occurs, rather than simply a comparison of situations involving the presence or absence of the presumed cause" (pp. 8-9).

Process as a formal approach to causal inference

Most people do not do science. It is a rare automobile mechanic who takes an interest in theory. Given a vehicle with a problem, he is not likely to make a list of the possible causes; he analyzes the situation informally, based on past training and experience. But, and here is the key, whether he is able to verbalize it or not, he can and does work from a reasonably complete list of those probable causes. And in fact, the ability to identify possible causes in such cases is neither unusual nor mysterious. Michael Scriven (1976) called attention to this fact, and used it as the starting point for a logical formalization of the analysis of causal process. He wrote:

> Most automobile mechanics. . . . rely, although they may not realize it, on a claim that . . . Most X's (in C) are caused by A, A', A'', . . . , A^n. Let us call this a *quasi-exhaustive causal list.* The sense in which [the mechanics] need this does not imply that they can state it, but it is easy to prove that they know it without relying on verbalization . . . and only a little less easy to show that they need it. (p. 104)

Scriven would standardize or formalize this everyday procedure, making as complete a list as possible of causes (including even the unlikely) to be conscientiously checked, eliminating those not present as well as identifying the actual cause. Being able to state that possible causes are not present strengthens the conclusion. And while these efforts may not fully qualify as scientific method, Scriven points out that "something can still be done even if we know only some of the entries in such a list" (p. 104).

Discriminating among multiple causes

A problem arises in this simple approach when we must show evidence that one or more causes is present, while other suspected causes are not. To address this problem, we must be able to identify one or more unique characteristics of each cause. For example, some causes are identified by their physical traces (certain poisons, for example, as a cause of death), others by detectable visual or audible patterns. The purpose of the analysis is to show "beyond a reasonable doubt" that all suspected causes are accounted for, as either present or absent. The approach is open-ended and inductive. In introducing a formal application of this elaboration, Scriven referred to it as the modus operandi (MO) check. He described it as follows:

> The MO of a particular cause is an associated configuration of events, processes, or properties, usually in time sequence, which can often be described as the *characteristic causal chain* (or certain distinctive features of this chain) connecting the cause with the effect. . . . The basic truism for MO analysis is that only real, that is, operative, causes fulfill their MO "contracts." Even if A and A' are both present, which we may determine directly or by inference from certain MO cues, one may not have completed the causal connection to the effect. The general nature of our task is thus one of pattern recognition . . . . In general terms, this part of the investigation focuses on discovering how many *complete* MO's are present. (pp. 105-106)

And here is the syllogism for the strongest inference from the MO under realistic conditions:

> [Quasi-exhaustive lists, and even partial lists] support a highly probable conclusion if the MO check is both complex and successful. The antecedent likelihood that an unknown factor will have the same MO is, in general, very low. Thus:
>
> i. A and A' can sometimes cause X.
>
> ii. Nothing else is known to cause X.
>
> iii. A but not A' was present.
>
> iv. The MO of A, which is highly distinctive, was present.
>
> A probably caused X.
>
> Again [as with the simple causal list], the third premise can be omitted, but, with this weaker inference [an incomplete list], its presence is valuable. The sense of "highly distinctive" that is needed for validity here refers to unknown other possible causes of X. Since they are unknown, the safest inferences will be those where the MO configuration is very complex. (p. 107)

Scriven's steps have been variously augmented over the years (see the discussion in Davidson, 2000), but the claim supporting a valid and independent process approach to causal analysis is still the MO and its claim to permit causal inference without resort to a counterfactual. Scriven insisted that the MO was capable of substituting for experimental and quasi-experimental methods, as well as supplementing them, and he went so far as to state that "I believe that the main thrust of efforts towards sophistication [in the design and support of evaluations] should now turn from the quasi-experimental toward the modus operandi approach" (p. 108).

Although the MO offers a way around the counterfactual, a way of determining cause in a single case, circumventing the variance approach was not Scriven's purpose, nor did he consider the approach one to be avoided. On the contrary, he began his article by stating that "control-group studies . . . are the method of choice [but] we must frequently face the need to do the best we can with nonexperimental data" (p. 102). He agreed with the superiority of the variance approach and the deterministic assumptions that underwrite it, acknowledging that his causal lists "are rather modest claims by contrast with . . . strong general determinism, which is usually said to be a required assumption for scientific investigation" (p. 104).

The examination of the components of the causal chain is for Scriven a presence-absence concern, the perspective of the experimentalist. However, in order to make the lists, and especially to utilize the MO signatures of various causes, it is necessary to know a great deal about the suspected causes; about the causal mechanisms, if you will. Scriven insisted that the experience that skilled and knowledgeable people possess needs to be tapped and incorporated into analyses, an argument that system dynamicists accepted from the beginning and routinely carry out. Scriven put it this way: "more attention needs to be given to externalizing the implicit knowledge of causal lists and MO's possessed by many specialists, be they master teachers or union leaders . . . . It is not that evaluators, social psychologists, sociologists, and others have never done this . . . . Rather, they have done it . . . with a degree of informality that leaves it in the category of anecdotal evidence" (p. 108).

Acceptance in principle and rejection in practice

Professional acceptance of the method Scriven, in his paper on the MO, referred to direct contacts with Donald Campbell. Campbell, often in collaboration with others, was a definitive source for experimental and quasi-experimental research methods in applied research and evaluation studies for half a century (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Shadish, Cook & Campbell, 2002). While his early writings strongly favored the experimental camp, Campbell "modified his view of qualitative research over time" (Johnson & Onwuegbuzie, 2004, p. 24). According to Maxwell, Campbell was a realist, and one of the foremost publications in the experimental tradition (Cook & Campbell, 1979) "is explicitly grounded in a realist epistemology" (2004, p. 9). Thomas Cook, who was Campbell's best known collaborator and is a widely quoted methodologist in his own right, states plainly that "the most esteemed theories of cause emphasize explaining relationships rather than merely describing them as 'if-then' causal connections" (2002, p. 179).

From the beginning, the MO concept and related ideas concerning the process approach have been cast in the applied context; that is, in program evaluations. Scriven set the stage by shifting the orientation from one of accounting for naturally occurring events to one of creating MOs to be used in linking interventions (deliberate causes) to outcomes desired to occur as their result. It is in this perspective that the MO concept was accepted into the matrix of approved methods. In their well-known 1979 work on quasi-experiments, Cook and Campbell acknowledged Scriven's MO to be not only a legitimate method of causal inference but a substantial improvement over their one-group posttest-only design, pointing out that in applying the MO, "the [original] experimental design . . . has become more complex and has many dependent variables that are expected to have different levels" (p. 97). The MO adds a series of intermediate cause-effect situations, and the presence of all (since a true cause always fulfills its contract) stands in for the if-and-only-if check provided by the control group.

Practical application and rejection Scriven had suggested that markers be embedded in the program process, to be picked up by evaluation. These would serve as "signatures" to assist in showing that intended causes of changes observed as a result of deliberate interventions really were the causes. In the same period (late 1960s and early 1970s), the idea that the evaluation of programs might be based on causal models of those programs began to be implemented. This became the field of program theory. The Theories of Change approach (ToC) is a program theory approach that became popular—with charitable foundations in particular—in the 1990s. The program/policy is assumed to be a cause of change, and the program as carried out is the causal mechanism. The theory predicts and the evaluators look for the expected "signatures" of the anticipated causes to determine if they fulfill their contracts. The design is expected to stand alone, without resort to control groups. As Cook has stated it: "The claim is that such theory-based evaluation can function as an alternative [to the randomized experiment]" (2002, p. 194). Among the arguments Cook lists as having been put forward in support of this claim are that: "First, [the ToC] does not require a causal counterfactual constructed through random assignment or matched comparison groups. . . . Only the group experiencing a treatment is needed. Second, obtaining data patterns congruent with program theory is assumed to validate that theory. This epistemology does not require explicitly rejecting alternative explanations, merely demonstrating a close match between the predicted and obtained data" (2002, p. 194).

The method must be capable of providing an alternative to the counterfactual, else there is no known independent nonexperimental avenue to causal inference and the process approach is in fact subordinate to the variance approach. Cook concluded that the claim to stand alone had not been successfully demonstrated by the ToC approach. Listing a series of shortcomings in the substantive theory, Cook concluded that the approach failed to live up to its claims: "The . . . biggest problem with a theory of evaluation that depends on a program's substantive theory alone is that there is no valid counterfactual, no way of knowing what would have happened at any stage in the model had there not been the program" (2000, p. 31).

The System Dynamics Model and the MO

Cook did not conclude that the MO method had failed; he concluded that the substantive theory of the ToC plan had failed. He reaffirmed that, in principle, the claim of the MO approach (which he refers to as the "theory of signed causes") is sound: "One way to guard against [the ToC's inability to succeed in the absence of a counterfactual] is with 'signed causes,' predicting a multivariate pattern among the outcomes that is so unique it could only have occurred because of the reform under analysis" (2002, p. 194).

The legitimacy of the MO method (defined as its formal endorsement in the more prestigious publications of the field) is intact. However, "signed causes depend on access to considerable well-validated substantive theory" (2000, p. 31). Cook points out that unlike the detective solving a crime or the pathologist determining the cause of death:

> Rarely do social scientists have such specific background information available to them from substantive theory and experience, so discriminating among alternative causes is much more difficult. And rarely is the pattern of effects to be explained as clear-cut as the crime scene that a detective finds or the body that a pathologist dissects. So the theory of signed causes is not likely to be a widely applicable alternative to a valid counterfactual control group. (2000, p. 31)

Thus the problem lies in the theory that describes the causal mechanism. It is to be anticipated that the MO method will be difficult to apply, but it is accepted as a possible alternative to the experimental method. It remains to be shown that it can be a feasible alternative.

The system dynamics model as "substantive theory"

Can system dynamics models improve the application of the MO approach sufficiently to make the process approach to causal inference a meaningful alternative to the experimental method? Two of Cook's criticisms of the ToC theory play directly to the strengths of the system dynamics methodology. One recognizes the need to incorporate feedback concepts: "many of these theories seem to be too linear in their flow of influence, rarely incorporating reciprocal feedback loops or external contingencies that might moderate the flow of influence" (2002, p. 194). The other, on determining timelines, implies a need for a dynamic framework. This is strong evidence that system dynamics is a better vehicle for theory than any so far applied, where the intent is to use it in lieu of a counterfactual.

In the role of the substantive theory, the model would provide—having identified the events(s) to be explained—a rigorous formal explanation (a theory) describing the mechanism by which the cause produces the observed effect. In doing so, the signatures by which the presence of the cause may be verified (or not) are generated. Forrester has pointed out that "A model is a theory that explains the behavior created by the model. The structure of the model and the policies in it are clearly the reasons for the behavior that results from the interactions of the parts of the modeled system" (2003, p. 13). In assessing the result, Forrester observed that: "Confidence in such a model depends on whether or not the structure of the model can be identified in the real world [and] how similar the model behavior is to the kind of behavior that has been observed" (ibid). The model embodies the substantive theory, and the "identification in the real world" is the pattern match.

The "well-validated substantive theory" to which Cook refers is theory derived from available knowledge—research literature and experience. There is no guarantee that the resulting patterns will possess characteristics sufficiently unique to serve in the role of substitute for a counterfactual. It is important to bear in mind that the concern here is not for model complexity, but for complex or "highly distinctive" outcomes that are observed and/or theoretically predicted. The causal mechanism may or may not produce "signatures" that are highly distinctive or unique. However, different models do reflect different representations of theory. Linear models are not so likely to produce convincingly unique matches. The advantage of a system dynamics model (even a simple one) is that its non-linear outcomes are more likely to produce patterns that reflect empirical complexities. On occasion, the model may assist in their discovery. I offer as examples brief descriptions of the results of two models drawn from my own work in educational research. Both have been described in more detail in papers presented at past system dynamics conferences (2001, 2003).

Example one

In the first example I seek to explain the drop in entrance exam scores at a community college following the abrupt termination of a state policy that maintained high rates of retention in grade at local public schools.

The cause to be inferred The cause to be inferred is indirect. Retention in grade of secondary school students who do not meet graduation standards is taken to cause increased dropout, which in turn affects the percentage of high school graduates who perform at or above grade level.

The pattern to be matched The pattern to be matched is a year-to-year decline in the percent of high school graduates scoring above the cutoff on the entrance exam to a local community college, over a twelve year period. This pattern is displayed by the black circular symbols in the chart at the bottom of Figure 1.

The pattern match The inference of cause is based on demonstrating that the empirical pattern is matched by a model-generated pattern that is so unique that it can be clearly traced to a prior and more general empirical pattern attributable to the presumed cause. A lot of curves would fit the cascading points shown in the graph, but the uniqueness lies in closely matching the point-to-point variations of direction in the single pattern over time. The red line in the bottom panel of Figure 1 shows that the fit of the model-generated pattern to the empirical time series closely follows its erratic path. The closeness of the fit to a unique pattern is the basis of the claim of uniqueness.

Linking cause to pattern Theory and reason both assert that increased retention causes increases in dropout rates, and that the more under-performing students who drop out, the greater the percentage of the remaining will be students performing at or above grade level. When policies that mandate high retention rates are abruptly terminated, the reverse will occur.

Retention-in-grade exhibits a very distinct and commonly occurring pattern. The top panel in Figure 1 shows data assembled from reports of eleven states for the year 1985 (from Morris, 1993). The symbols show the aggregated percent retained, and the line represents an exponential decay function started at grade one and reset at grades seven and nine. The process that produces this pattern is roughly as follows. A standard is set against which to judge student performance, and students are to be held back when they are found not meet that standard. Teachers then assess and retain those whom they find have not met the standard. Teachers in the next grade will find more, and retain them. This will continue in each consecutive grade until all those who are deficient in the standard are detected and retained, or until the situation changes, as when students go on to the next level. System dynamicists will recognize the process as a simple goal-gap structure tracing out an exponential pattern of decay.

High-stakes policies that abruptly introduce high retention rates leave a similar distinctive trail. In the baseline run of the school district model with no dropout (the lower blue line in the middle panel), retention affects the graduate performance in spikes of increase in the percent qualified, clearly mimicking the pattern of the panel above at each of the three levels where retention was introduced. When high dropout is introduced as a direct cause of the retention, the pattern—though blurred—remains as the percent of qualified graduates (those not dropping out) increases, as can be observed in the red line. And when the policy is terminated, the pattern appears in reverse.

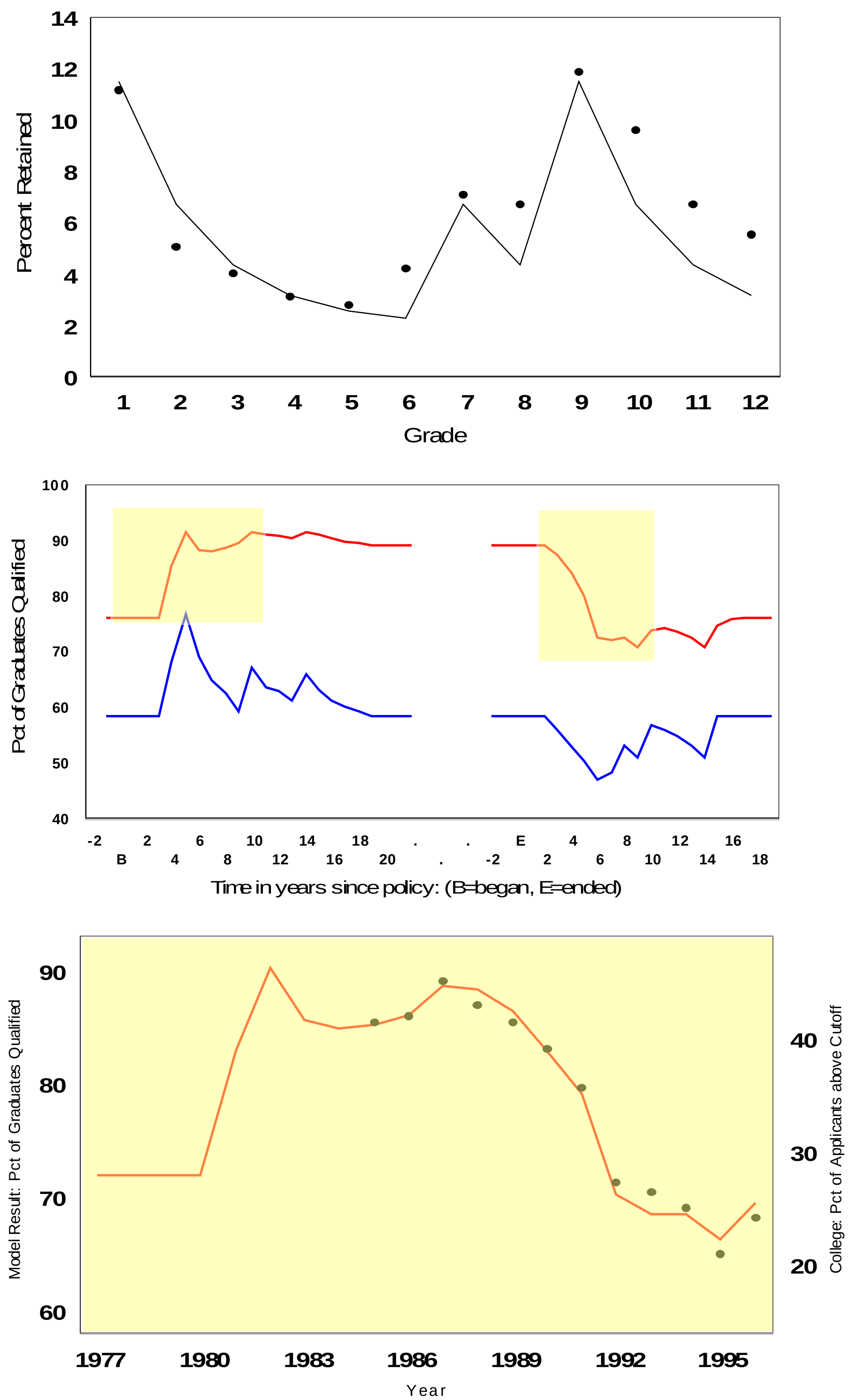

Figure 1 Retention in grade as the cause of the drop in qualified applicants to the community college. The top panel shows the aggregated retention rate of ll American states in 1985 (Source: Morris, 1993). The middle panel shows baseline runs of a school district chain model with a negative feedback loop representing retention. The blue line shows the pure retention effect, with no dropout. The red line shows the retention effect under conditions where a high proportion of retained students drop out. The lower panel shows the application of the model to the community college admissions scores data. The yellow areas in the center panel indicate the correspondence to the lower panel chart. All these graphs can be viewed in the context of a more complete description of the model in Morris, 2001.

When calibrated to the local situation, the model produced the pattern observed in the lower panel of Figure 1 (red line), where it can be seen to follow the annual rises and falls of the data points with considerable faithfulness. The shading (yellow background) of the bottom panel is a reflection of the areas marked in yellow in the panel above, where the pattern is clearly seen to be a result of variations in the dropout rate as conditioned on the retention rate. The signature of retention as a cause of the decline in entrance exam scores is clearly evident.[1]

Example two

Example two is an explanation of the reasons why the aggregate relationship between achievement and poverty changes from linear to nonlinear when the achievement measure is broken down into performance categories.

The cause to be inferred The cause of nonlinearities in the relationship between poverty and achievement is taken to be the social process of students in schools, when achievement is broken down into the percent of students scoring in categories from low to high.

The patterns to be matched To be matched is a set of patterns that occur in every observed instance of the event to date. There are several interrelated patterns. Each pattern consists of shapes of measures of student achievement plotted across the range of a poverty measure. These patterns are shown in the left column of Figure 2. They are the results from the Florida Comprehensive Assessment Test (FCAT) for grade 3 of 2003, selected randomly from fourteen processed data sets.[2] The test results are published by school as percentages of students scoring in five performance levels from lowest (level 1) to highest (level 5).

The first pattern is found in the top panel, which shows the plot of the weighted averages by school of all the levels (the achievement measure) across the measure of poverty (the percent of the school's students qualifying for Free or Reduced-price Lunch). At this highest level of aggregation the pattern of the relationship between achievement and poverty is shown to be a linear decline in achievement as poverty increases. This is the common relationship found in results from normed-referenced tests, and the test in question—though criterion referenced—is correlated highly with a sister test that is norm referenced.

In the center left panel, the results are divided into those percentages of students performing below grade level (levels 1 and 2) called At Risk (AR) and those who are not at risk (NAR) and performing at grade level or higher (levels 3 through 5). The chart shows that these groups are also linear on poverty, with the former (the red line) increasing while the latter (the blue line) decreases.

Finally, the third pattern, in the lower left panel, shows a curvilinear relationship. The five levels are broken out into four subgroups, two each for the AR and NAR groups. The two subgroups in the NAR group represent students who are average performers (level 3, and referred to as the Average subgroup)—the dark blue line—and the high achievers (levels 4 and 5, referred to as the Above-Average subgroup)—the light blue line. The two blue lines are observed to curve away in opposite directions, creating nonlinear relationships with the poverty measure.

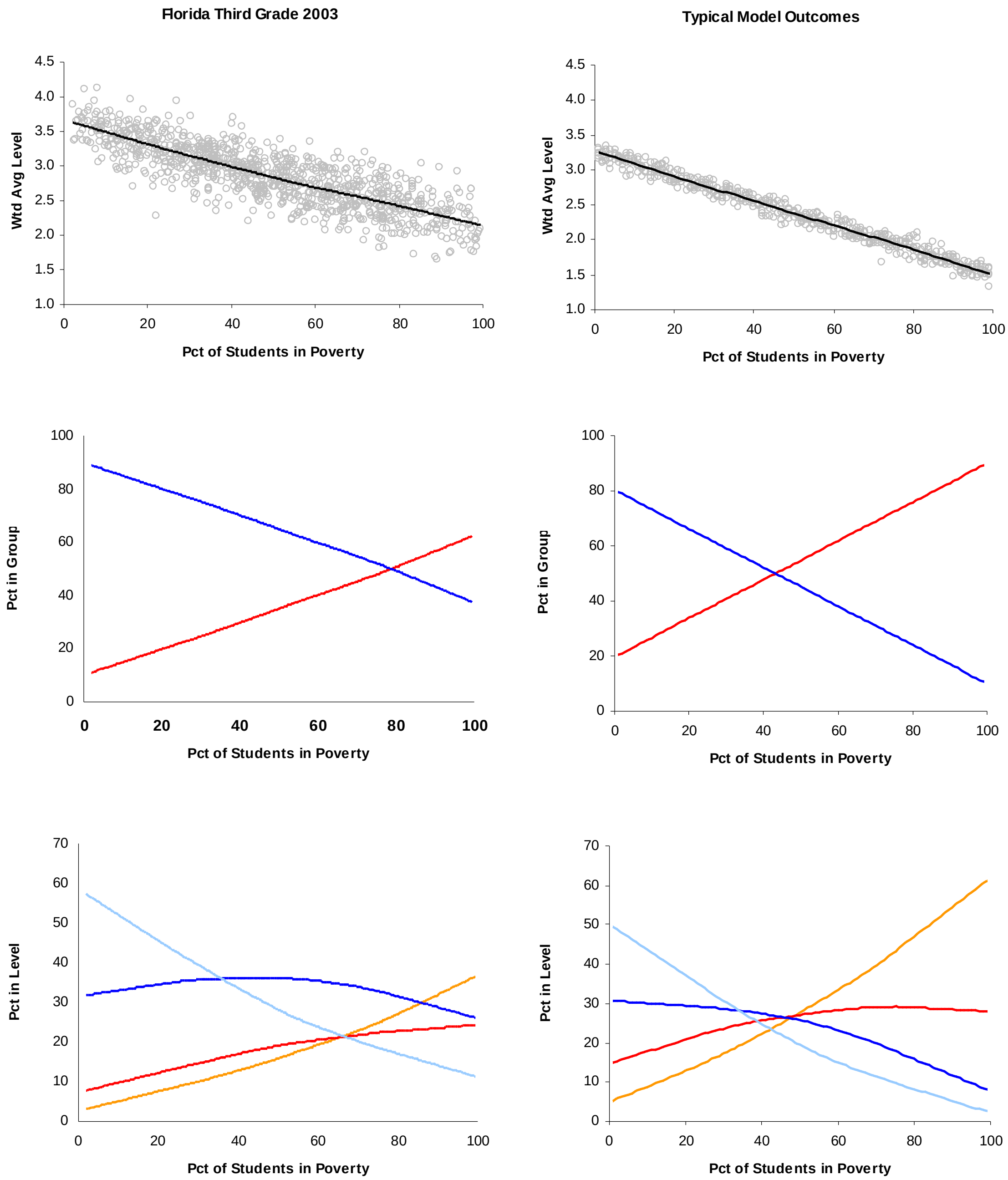

Figure 2 Florida Comprehensive Assessment Test results (left) and model simulation outcomes (right). In the top panels the FCAT levels are combined by weighted averages of the percentages in the levels, yielding the familiar linear decline of performance with poverty that is characteristic of standardized achievement test results. In the middle panels, the blue line represents the group of students not at risk (levels 3, 4, and 5), and the red line represents the group of students who are at risk of failure (levels 1 and 2). In the lower panels the light blue line represents the highest performing students (levels 4 plus 5), the dark blue the average performers, the red the basic (level 2) performers, and the orange the lowest (level 1) performers. The Florida data is for 2003 third grade (chosen randomly from grades three through eight and years 2001-03). The data represent approximately 1,000 schools. Source is Florida Department of Education. The model results plotted represent 500 runs evenly spaced across the poverty range. All lines are produced by the local regression algorithm lowess using a smoothing parameter of 0.5 (see Cleveland, 1979).

The two subgroups in the AR group represent students who are marginally below grade level (the Basic performers, represented by a red line) and those who are very much below (represented by the orange line). The Below-Basic performers are considered leaders in their own way, as representing a rebellious anti-school attitude. These two lines also, the red and the orange, curve in opposite directions and display a nonlinear relationship with poverty.

The pattern match All of the patterns just described are to be matched by output from the model. The number of pattern-shapes and the characteristics of each are listed, along with the method by which they are produced. Each and every characteristic in the list is to be qualitatively matched (the shapes reproduced) from the same lower level units (schools). The task is to reproduce all characteristics in the compiled list, in the same way that the empirical data are processed. The appeal to uniqueness lies in the qualitative matching of all the shapes without exception, arrived at by the same procedures as were applied to the empirical data. The matching model output is shown in the right column of Figure 2.

Linking cause to pattern The social process that is proposed as the cause of the nonlinearities in the achievement-poverty relationship is modeled at the school level.[3] Grounded in the research literature on peer influence, it is a process of attraction of more influential peers on like-minded students who come to adopt their attitudes and behavior in a series of events occurring over the course of a school-year. Above-Average students attract Average performing students, and Below-Basic students (with anti-scholastic attitudes) attract Basic-performing students. The relationship is one of positive feedback, and in schools where most students are of the NAR group, the number of Above-Average performers will show a noticeable and nonlinear increase at the end of the run, at the expense of the Average subgroup. Conversely, in schools where most students are of the AR group, the Below-Basic subgroup will show an observable nonlinear increase while the Basic subgroup declines.

Measures of achievement (the test results) are taken only at the end of the year, so there are no temporal patterns in the empirical data to be matched. However, by collectively plotting model outcomes across the range of poverty, we can observe the social process in cross-section. Each run of the model produces four “end-of-the-year” subgroup outcomes in simulation of the performance levels in the empirical data produced by a single school. When plotted against poverty, the simulated schools perform in all the ways that real schools do, producing results that decrease with increasing poverty, and combining in the same ways to form linear relationships. They also reproduce the curvature that the individual FCAT levels show, and they do it in accord with theory, research, and experience accruing over a long period—the common sense of the field.

In the right column of Figure 2 the results from 500 model runs have been processed in the same manner as the Florida school data. In the upper right panel, the results of each model run have been combined as weighted percents—just as in the empirical data—and plotted to create a graph matching the empirical results on the left. Similarly, the levels are combined in pairs to produce a result in the center right panel that matches the linear empirical results on the left. Finally, the four model-generated outcomes are plotted separately. Here one observes that the levels curve in the same manner and direction as their counterparts on the bottom left chart. The model-predicted nonlinearities constitute the signature of the social process.

The matching of so many patterns simultaneously with no toleration of exceptions is the basis of the claim to uniqueness through complexity. The claim is further supported by the additional step of applying the model at a lower analysis level and aggregating up. The aggregation procedures are matched as well.[4]

Variance vs. process, a reassessment

The MO approach was necessary in both of the examples presented. In the retention study, the approach was necessary because it is a case study, an N of one. In the achievement and poverty study, there were no schools without a social process, and so no possibility of comparing the patterns of those with and without. These examples represent two quite different appeals to the "MO criterion," two different interpretations of complexity or uniqueness. The first example conforms most closely to the concepts that Scriven and later Cook described. The second claims uniqueness in the quantity and consistency of the match, and the decision to allow no exceptions.

In both examples the causes are remote and basically simple—one or two feedback loops. I have tried to show that the signatures can be clearly traced back to the causes, so that the MO is "present and highly distinctive." It is unlikely in either case that the connection between cause and effect could have been made without applying a technology such as that offered by system dynamics. The examples thus demonstrate that with system dynamics assistance there are practical applications in which the MO is successfully applied. This in turn indicates that causal inference is possible using the process approach, so that it is in that sense the equal of the variance approach.

I do not suggest that constructing system dynamics models for this purpose is easy, and there is a question of how frequently the MO method can be successfully applied. The uniqueness criterion appeals to reason for a qualitative estimate, rather than to a statistical probability. The concept of uniqueness is not a dichotomy, but a spectrum where the pattern match varies from less to more unique, and that spectrum constitutes a kind of "confidence index." It remains to be seen how far along it a reliable process methodology will be worth the effort, and in what circumstances.

But there are difficulties with the variance approach as well. As Maxwell (2004) has pointed out, conditions in educational research (and, I submit, in the other social sciences as well) are rarely optimal for application of the experimental method. The control group is an expensive and troublesome accretion, and its reliability in practice is often overrated. Any educational researcher who has ever tried to select an equivalent control group or maintain the integrity of randomized groups over the course of an evaluation knows that the true reliability of the result is likely to differ substantially from the reported statistical significance. In theory it is a valid answer—where it is truly feasible arguably the preferred one—but in practice it, like the MO, faces "a daunting set of conditions."

The MO is to the process approach what controls are to the variance approach, the means by which the causal inference is confirmed. Both approaches have various alternative "carriers." I suggest that the system dynamics model is to the MO what the randomized control group is to designs that manipulate the causal variable, the best of several choices.

Commentary

The social sciences have long suffered from the fact that they are measured against a popular image of science held over from the late nineteenth century, when there was a strong public belief that science would soon provide the answers to all of society's ills. This was the era of strong determinism, the heyday of the classical physical sciences. Strong determinism (causal relationships in nature brooked no exceptions) has been a required assumption for scientific investigation—at least above the sub-atomic level. The basic aim of science was understood to be the discovery of fundamental laws of nature, universal causes that by definition do not vary with time or location.[5] This is the environment in which attitudes toward causation have been formed. The unscientific everyday use of the term refers to particular causes, which are as many and as diverse as there are events to be explained.[6] It has been the scientist's job to seek invariant causes—the essence of natural laws. That view has been, and remains today, the most widespread view of what all science should be about.

However, Goldman (2004) has described many changes that took place in the way science came to be viewed over the course of the twentieth century. A number of Goldman's topics—evolutionary economics, path determinism, self-organization and complexity—have been acknowledged by system dynamicists (e.g. Radzicki & Sterman 1993; Radzicki 2003). Of particular interest is the growing shift to an evolutionary perspective. Since the middle of the last century, the evolution paradigm has come to challenge determinism in many sciences, and this challenge has provided a foundation for skepticism about the universal appropriateness of the experimental method. Stephen Jay Gould (1989) distinguished between two approaches to science: the deterministic and what he termed the "sciences of history." Among the latter are evolutionary biology, geology, and Gould's own discipline, paleontology. Reflecting his mentor Charles Darwin, Gould makes a distinction—applied to all the natural sciences—between laws in the background and contingency in the details. Thus the universe runs by laws, but "with details, whether good or bad, left to the working of what we may call chance" (1989, p. 290). In the social sciences, the role of contingency is very large, with the consequence that one cannot understand the present without understanding the past, and non-deterministic in the sense that one cannot predict the future with any certainty.

Maxwell (2004), who quotes Gould, points out that "what Gould calls 'historical explanation' is better seen as process explanation. . . . [and] this approach is also widely used in ecology, psychology, anthropology, and sociology" (2004, p. 5). Sklar in history (1991) and Burnham in political science (1994) embrace Gould's "pattern model" in a way similar to that in which Radzicki (1990) described it for institutional economics. And in the wake of the economic institutionalists, there are neo-institutionalist movements in sociology, political science, and organization theory (see DiMaggio and Powell's introductory comments in Powell & DiMaggio, 1991).

With the adoption of the evolutionary perspective, concentration is shifted to the particular, in order to deal with that large role of contingency. In dealing with contingency, we must deal with particular causes, and in doing so it is helpful to employ the explanatory device of a causal mechanism. Process theory can often handle the problem of causal inference better than variance theory in the "high-contingency" conditions of the social sciences, and there is a growing awareness and acceptance of this fact.

Toward greater collaboration

Caught in this changing intellectual environment, the situation in many social sciences today is fluid, and this confronts system dynamicists with both challenges and opportunities. Change brings opportunity, but it provokes resistance as well, and that resistance channels the paths to opportunity, blocking some avenues and opening others. Resistance is strongest in those areas which were most successfully integrated with the old regime, and unfortunately, these are the areas with the most contemporary appeal, since they still carry prestige and the aura of acceptance. Economics is one of these areas, and although Radzicki has made a compelling case for the integration of system dynamics with economic institutionalism, the economic institutionalists would appear to have a hard road ahead of them. As Repenning has suggested, this may not be the optimum path to interdisciplinary collaboration.

So while change may be “in the air,” questions of where and how to take advantage of the opportunities it generates are never clear or easy. I have identified what seems to me to be the best path to successful collaboration and greater integration into the full breadth of the social science community. The strategy—as I have tried to show it—is an approach through the fundamental methodology by which all the social sciences conduct research and attempt to find answers to the basic question: what caused this event to come about? While this strategy cuts across all the social science disciplines, actual tactics of successful collaboration do center on specific academic disciplines (where vulnerability will be a chief consideration), and through evaluation affiliates (where system dynamics has a technical advantage). This gives the flexibility of breadth while playing to system dynamics’s strong suite—its systems methodology.

I have drawn my examples from educational research and program theory. Educational research, like system dynamics, is a field oriented toward application (and so exhibits many of the qualities Repenning would preserve), and it is already a field that is accustomed to drawing on the work of various disciplines; especially psychology, sociology, and public administration, and more recently organization theory. Some in the field have developed an interest in what are termed systemic approaches (see Reigeluth & Garfinkle, 1994 and Banathy, 1991). Finally, educational research is a field under strong pressures to make major changes in the way its members conduct research.[7]

Program theory is another area that shows promise as a particularly fertile area for system dynamics, for reasons that I have discussed elsewhere (Morris, 1996). Program theory has even ventured to apply a system dynamics model in one reported instance (McClintock, 1990). None of the program theory research shows anything like the sophistication in systems analysis that system dynamicists take for granted, and few of its practitioners show any awareness of system dynamics. Supporting evidence for this assertion comes from program theory itself, where both the lack of and the need for systems approaches are acknowledged.[8] Judging from the state of the program theory literature, system dynamics’s group modeling techniques appear to be far superior to the methods currently being discussed there. Compare for example the writings in Rogers, et al. (2000) with the projects described in Morecroft and Sterman (1994).

For these reasons educational research and program theory present opportunities for major system dynamics contributions, given a sufficiently sensitive strategy. Strengths should be shown (do not simplify, that conceals your contribution) but not flaunted. Begin by becoming thoroughly familiar with a field's literature, and when working within its boundaries, adopt and use its terminology. Welcome to the (other) social sciences, system dynamicists.

References

Banathy, B. H. (1991). *Systems design of education: A journey to create the future*. Englewood Cliffs, NJ: Educational Technology Publications.

Berk, R. A. (1988). Causal inference for sociological data. In N. J. Smelser (Ed.). *Handbook of sociology* (pp. 155-172). Newbury Park, CA: Sage Publications.

Burnham, W. D. (1994). Pattern recognition and "doing" political history: Art, science, or bootless enterprise? In L. C. Dodd and C. Jillson (Eds.), *The Dynamics of American Politics*, pp. 59-82 . Boulder, CO: Westview Press.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton-Mifflin.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74,* 829-836.

Cook, T. D. (2000). The false choice between theory-based evaluation and experimentation. In P. J. Rogers, T. A. Hacsi, A. Petrosino, and T. A. Heubner, (Eds.), *Program theory in evaluation: Challenges and opportunities*. New Directions in Program Evaluation (no. 87, pp. 27-34). San Francisco: Jossey-Bass.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis, 24,* 175-199.

Cook, T. D. &. Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton-Mifflin.

Davidson, E. J. (2000). Ascertaining causality in theory-based evaluation. In P. J. Rogers, T. A. Hacsi, A. Petrosino, and T. A. Heubner, (Eds.), *Program theory in evaluation: Challenges and opportunities*. New Directions in Program Evaluation (no. 87, pp. 17-26). San Francisco: Jossey-Bass..

Forrester, J. W. (2003). Economic theory for the new millennium. *Proceedings of the 21st International Conference of the System Dynamics Society*. Albany, NY: System Dynamics Society.

Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics*. *12* (2), 101-128.

Goldman, S. L. (2004). *Science in the twentieth century: A social-intellectual survey*. Chantilly, VA: The Teaching Company. DVD series.

Gould, S. J. (1989). *Wonderful life: The Burgess Shale and the nature of history*. New York: Norton.

Kaestle, C. F. (1993). The awful reputation of education research. *Educational Researcher*, *22* (1), 23-31.

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, *30* (2), 3-13.

Johnson, R. B., & Onwuegbuzie, A.. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33* (7), 14-26.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33* (2), 3-11.

McClintock, C. (1990). Evaluators as applied theorists. *Evaluation Practice, 11,* 1-12.

Mohr, L. B. (1982). *Explaining organizational behavior*. San Francisco: Jossey-Bass.

Mohr, L. B. (1996). *The causes of human behavior: Implications for theory and method in the social sciences*. Ann Arbor: The University of Michigan Press.

Morecroft, J. D. W. & Sterman, J. D. Eds. (1994). *Modeling for learning organizations*. Portland, OR: Productivity Press.

Morris, D. R. (1993). Patterns of aggregate grade-retention rates. *American Educational Research Journal, 30*, 497-514.

Morris, D. R. (1996). Institutionalization and the reform process: A system dynamics perspective. *Educational Policy*. *10,* 427-447.

Morris, D. R. (2001). Causal inference and system dynamics in social science research: A commentary with example. *Proceedings of the 19th International Conference of the System Dynamics Society*. Albany, NY: System Dynamics Society.

Morris, D. R. (2003). Peer influence in educational reform: A system dynamics approach. *Proceedings of the 21st International Conference of the System Dynamics Society*. Albany, NY: System Dynamics Society.

Olson, L. & Viadero, D. (2002). Law mandates scientific base for research. *Education Week*, vol. 21 (20), pp. 1, 14, 15.

Powell, W. W. & DiMaggio, P. J., Eds. (1991). *The new institutionalism in organizational analysis.* Chicago, IL: The University of Chicago Press.

Radzicki, M. J. (1990). Methodoligia oeconomiae et systematis dynamis. *System Dynamics Review*, *6*, 123-147.

Radzicki, M. J. (2003). Mr. Hamilton, Mr. Forrester and a foundation for evolutionary economics. *Proceedings of the 21st International Conference of the System Dynamics Society*. Albany, NY: System Dynamics Society.

Radzicki, M. J., & Sterman, J. D. (1993). Evolutionary economics and system dynamics. *Proceedings of the 1993 International System Dynamics Conference* (pp. 388 I-388 XXVI). Cancun, Mexico: System Dynamics Society.

Reigeluth, C. M. and. Garfinkle, R. J. (Eds). (1994). *Systemic change in education*. Englewood Cliffs, NJ: Educational Technology Publications.

Repenning, N. P. (2003). Selling system dynamics to (other) social scientists. *System Dynamics Review*, *19,* 303-327.

Richardson, G. P. (1991). *Feedback thought in social science*. University of Pennsylvania Press.

Rogers, P. J. (2000). Causal models in program theory. In P. J. Rogers, T. A. Hacsi, A. Petrosino, and T. A. Heubner, (Eds.), *Program theory in evaluation: Challenges and* *opportunities*. New Directions in Program Evaluation, (No. 87, pp. 47-55). San Francisco: Jossey-Bass.

Scriven, M. (1976). Maximizing the power of causal investigations: The modus operandi method. In G. V. Glass (Ed.), *Evaluation studies review annual* (Vol. 1, pp. 101-118). Beverly Hills, CA: Sage Publications.

Scott, W. R. (1994). Institutional analysis: Variance and process theory approaches. In W. R. Scott, J. W. Meyer, and Associates (Eds.), *Institutional environments and organizations: Structural complexity and individualism* (pp. 81-99). Thousand Oaks, CA: Sage Publications.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sklar, M. (1991). Periodization and historiography: Studying American political development in the progressive era, 1890s-1916. In K. Orren & S. Skowronek, Eds., *Studies in American political development*, vol. 5. New York: Cambridge University Press.

Notes

About the Author: Don R. Morris received his PhD in political science in 1971 at the University of Wisconsin at Madison. He served on the faculties of Vanderbilt University and the University of Miami's Center for Advanced International Studies before joining the Miami-Dade school district's evaluation and research team in 1982. He is the author of more than 40 papers and articles, and winner of the AERA's 1998 Division H Outstanding Publications Award in the category of Advances in Methodology. His professional interests include research methodology and educational policy.

[1]The only serious alternative explanation was that the school district had abruptly and substantially lowered its standards some years before the drop in graduate quality began, but test scores from the ninth grade level High School Equivalency Test—used to determine minimum graduation standards—did not show any pattern to match the drop observed in the community college scores.

[2]Statewide testing of grades three through nine with the full battery of tests began in 2001. Data are available by school for all Florida counties from 2000 (partial) and 2001 for all tests. The data plotted and examined included approximately 1,000 elementary schools and 700 middle schools, depending on grade configuration, for each of the years 2000 through 2003. At the time of the analysis, test data were available for 2004, but not data on the poverty measure. The poverty measure is not available for the high schools, and so ninth grade was not examined.

[3]When I constructed the original model several years ago, I was not aware of the FCAT patterns. I assumed that the model would apply primarily at the middle school level, and ran the model against different data that yielded a much simpler pattern. That analysis was presented at the 2003 System Dynamics Conference, and it may be examined in detail in the paper included in the Proceedings for that year. A paper on the FCAT patterns is planned for presentation at the Florida Educational Research Association's conference in Miami in November of 2005.

[4]Data from an earlier Florida reform (1977-1989) showed that test results of a performance range equivalent to that of level 1 in the FCAT, administered at the first of the school-year, result in a linear pattern across the poverty measure, supporting the contention that the curvature is a product of events occurring during the school year.

[5]Members of the social sciences have long been urged to emulate the "hard sciences," and have frequently sought advice on how to do so. In 1958 for example, the leaders of the American Political Science Association invited the mathematician Anatol Rapoport to address its members on the essential qualities of a science. In the exact sciences, Rapoport counseled, theory and prediction are closely related. A theory is a set of assumptions or theorems, both assumed and derived, about how things should happen under ideal conditions. The more theorems derived or deduced, the more powerful the theory. Discrepancies between predictions and observations should lead to further refinements in the theory. To the extent that conditions of strong determinism hold, these refinements—undertaken as a series of ever more sophisticated experiments—will ultimately lead to increasing accuracy of prediction.

By this reasoning, it is the collection of derived theorems that makes a discipline a science. It is the accuracy of the predictions that makes it successful. Economics for example (that is, classical economics), has such a collection of deduced theorems, although the agreement between theory and observation has not been very good. Rapoport did not think that this discrepancy should be taken as indication that economics is not a science, or that the approach to research is not the correct one. In early meteorology, he pointed out, large predictive errors were also frequent, yet no one questioned whether its methods of research were inappropriate.

Rapoport went on to make some suggestions about how other social sciences might make their disciplines more scientific, in the exact sciences image. His favorite was game theory, which was to provide the theorems of ideal behavior by rational actors for decisions, a suggestion that the political scientists adopted enthusiastically over the following generation.

[6]Everyone "reasons causally" as a matter of course. We all "know" that the window broke because the baseball hit it, but no one thinks of that as scientific causal inference. To the educated layman, process thinking is considered mundane, whereas there is a tendency to regard the experimental approach as ipso facto scientific. If they think about it at all, the educated are easily convinced that the conclusion about

cause in a specific case—the broken window for example—is at root based on counterfactual reasoning, the product of subconsciously synthesizing innumerable results and non-results accumulated over a lifetime of observations. For a well-argued objection to the idea that people reason this way see Mohr (1996).

[7]For a variety of reasons, research in the education field has acquired an exceptionally poor reputation (see Kaestle, 1993). There has been a lot of soul-searching over the past several decades, and the *Educational Researcher*—the field's major journal for professional commentary—has been filled with passionate polemics over philosophical issues of positivism, its demise, and its proposed replacements. The field has been as vulnerable as any to the rivalries of the science wars, and a variety of methodological camps have formed, each proclaiming itself the wave of the future. This has resulted in a methodological pluralism in which the remnants of the older, positivist quantitative approach retain a weak hegemony.

Since 2001, a strong emphasis on scientifically based research by the Bush Administration (see Olson & Viadero, 2002) has generated a controversy among researchers, further fueling the acrimonious division among the various methodological camps. The controversy has come to center on how scientifically based research is to be interpreted, and in particular on a hierarchy of methods with the randomized experiment at the top. With the experimental approach at the top of the hierarchy, followed by quasi-experimental approaches where true experiments are not feasible, all nonexperimental methods—quantitative and qualitative—have been cast in a role of second-class assistants. This has generated a great deal of resentment. The official position of the field's major professional organization, the American Educational Research Association—to the chagrin of many—has been to emphasize its defense of the nonexperimental methodologies while acquiescing in their secondary status. Here is the opportunity for the intervention of a new approach, one that can demonstrate a credible claim to holding its own against the experimental method's claim to superiority.

[8]In her 2000 overview of causal models in program theory, Rogers wrote that "causal models are at the heart of program theory evaluation, yet there has been surprisingly little discussion of the different types of causal relationships that might be useful for program evaluation" (p. 47). She acknowledged that causality is complex, and that the simple causal chains in PTE theories are usually gross oversimplifications. She was further aware of the possibility that the relationship between cause and effect is not linear, and noted that feedback loops are rarely if ever included in the program logic. However, although she acknowledged that a few "causal models from systems theory . . . appear to be potentially useful for program theory" (p. 52), her sole reference outside her own field was Senge's *Fifth Discipline*.