Educational Considerations

Volume 39
Number 1 *Class Size and Student Achievement*

Article 7

9-1-2011

A Journey, Not a Destination

James L. Phelps

Closing Essay:

A Journey, Not a Destination

James L. Phelps

Much of the motivation and ideas for the articles in this special issue originated with my dear friend, Maris Abolins, Professor Emeritus of Physics at Michigan State University. We started as neighbors and, as our kids grew up together, we socialized frequently. He is responsible for my interest in physics. I would read a physics book, which would become the subject of our next dinner conversation (while our wives talked about other, more social topics). Instead of a compilation of facts, physics became a way of thinking about problem solving. The "unified field" theory was the start of my new thinking. There are four fundamental forces in nature: The strong force holding the atom together; the weak force dealing with the decay of the atom; electromagnetism; and gravity. Subatomic particles are responsible for these forces. Einstein tried to combine these four forces into one comprehensive theory, but there was insufficient experimental information to be successful. While some of the forces have been united into a theory (relativity and electrodynamics by American physicist Richard Feynman), gravity remains illusive. Was it possible to unify the various aspects of achievement production into a comprehensive theory? I wanted to give it a go! A unified theory might provide ideas helpful for improving research; professional training and practice; and, therefore, student achievement.

The individual pieces of a unified achievement production theory were scattered about, but I had not taken the time to assemble them. According to Glass and Smith (1978), relationships might not be linear, which started my thinking.[1] There were some efforts in the field of mathematical programming, e.g., data envelopment analysis (Silkman 1986), but after investigating these I found them wanting. "Fixed effect" analysis was in the economics literature, but the idea that it represents educational effectiveness had not been fully developed. Again, there were possibilities. Cost-effectiveness was addressed more substantially by Levin (1988), but not in a way to influence policy decisions. There were large controlled experiments, but the emphasis was on class size and not on a wide range of potentially influential variables. Little attention was paid to how several variables might work together. Economists were largely in the forefront of research, and there was little integration of the instructional and organizational aspects as suggested by Walberg (1984)[2] and Levin (1997). There is a great deal of ambiguity as to the purpose and conclusions of research. The research seems to be divided between what advocates more resources and what advocates organizational changes in order to improve education. There is little discussion regarding how they might work together. I wanted to rethink the fundamentals and see if these scattered pieces could be combined in some meaningful way.

After a professional meeting where the idea of simultaneous equations was raised, I started by writing down a number of basic equations to see if I could find some uniting principles. When each of the equations was graphed,[3] there were straight lines going every which way. There was no rational way to unite or choose among the alternatives. The only interpretation was to provide unlimited resources for all variables with positive slopes, hardly a practical or unifying strategy. With enough money, all schools could get perfect scores, a doubtful outcome. And what would be done with the variables with negative slopes—eliminate them all together? There was no practical method of evaluating alternatives. There were logical contradictions among the pieces. Instead of clarity, the exercise caused anxiety and confusion.

What made Albert Einstein so unique was his willingness to take on problems characterized by contradictions between explanations and experimental evidence. His contributions were monumental because he was able to make sense out of those contradictions. Richard Feynman was also a maverick in much the same way. In his books, Feynman writes about returning to the "first principle" when tackling intractable problems. He would start with the first principles—the basic principles underlying the phenomenon. He would test these principles to determine if they could stand strict scrutiny. If not, he would replace questionable principles with better alternatives. With the new principles in place, new solutions evolved. In essence:

- Flawed first principles lead to contradictory explanations and inaccurate predictions.
- Superior first principles lead to improved explanations and more accurate predictions.

Reviewing the productivity research is a strenuous exercise, as demonstrated by the earlier articles. Even the most diligent and ardent observer of achievement productivity research will have difficulty in reaching meaningful conclusions. There is "something for everyone." There is least one study supporting every possible policy conclusion. As a result, research has little value in solving everyday problems. It raises the question: Why conduct further research if the inevitable conclusion is the same—every option is effective!

There is no set of rules consistently and effectively applied to the many diverse educational situations. Instead, there are different and conflicting rules applied universally, discounting the unique situations. What are those "achievement rules"? The "Glass Rule" is to lower class size to one even though there is not enough money to do so. The "Hanushek Rule" is reduction of class size sometimes

James L. Phelps holds a Ph.D. from the University of Michigan in Educational Administration. He served as Special Assistant to Governor William Milliken of Michigan and Deputy Superintendent in the Michigan Department of Education. Active in the American Education Finance Association, he served on the Board of Directors and as President. Since retirement, he spends a great deal of time devoted to music, composing and arranging, playing string bass in orchestras and chamber groups, as well as singing in two choirs. He resides with his wife, Julie, in East Lansing, Michigan.

works and sometimes does not work; it all depends. The "Hedges Rule" (Hedges, Laine, and Greenwald 1994) is not to spend money on reducing class size, but spend money on whatever local decision-makers think is important. The "Tennessee Rule" (Achilles et al. 1993) is to lower class size. The "California Rule" (Bohrenstedt and Stecher 1999; 200s) is not to lower class size. The "Walberg Rule" is to change the curriculum and instructional programs, but with little direction as to how much and under what circumstances. The "Levin Rule" (Levin 1997) is to select the most cost-efficient programs, but by how much and under what circumstances? There was one common scheme. Every positive result reached the same conclusion: Increase funding without limit. Clearly, contradictory conclusions proliferate in achievement production research!

These "rules" are by-products of partial models; there is no single paradigm or comprehensive model encompassing the various aspects of the partial models.

The "reduce class size" or "spend more" rules are neither paradigms nor well-specified theories to test. Nevertheless, each piece of research has value in that it is a piece of a complicated puzzle. But the pieces have not yet been assembled into a mosaic for a clear image to appear. This is not to criticize the research as being bad. It points out the problem of reaching meaningful conclusions from research which has fundamentally different assumptions. What is missing is a set of first principles based on logic and evidence; and how the principles complement each other, and how accurately they explain and predict the phenomenon.

It is not possible to have multiple explanations for the same phenomenon—although it is possible to have several theories. After thorough testing, there must be just one theory which best explains and predicts the phenomenon. One of the basic assumptions of physics is that the physical laws apply everywhere in the universe. (It is science fiction when scientists apply different, untested laws.) Science is the pursuit of the best explanation with the best predictions. Regarding the explanation, the same laws apply in every situation, but when circumstances vary the solutions must vary. There cannot be identical solutions for varying circumstances. The influence of class size or any other variable must be the same in classrooms with similar conditions or it would be impossible to conduct research and to formulate explanations. Without this assumption, achievement production is reduced to opinion, with every opinion having equal, but not explanatory or predictive, value. But when school circumstances are different, there must be different solutions. The review of the achievement production research is abundant with contradictions regarding the statistical significance, shape of the relationships, effect sizes, and even the major determinants of achievement. Therefore, each piece of research produces a different explanation but the same solution, "unlimited more." I started to think in terms of some basic concepts, as follows: (1) *Similar circumstances must produce similar results; and there can be only one set of laws best explaining and predicting those results;* and (2) *Within the laws, different circumstances (parameters) must produce different solutions*. The challenge is to define the applicable laws and the influential circumstances.

Why the Contradictions?

Achievement research mostly relies on statistical models, which do not necessarily represent achievement production. Statistical models, in and of themselves, do not represent unified and coherent assumptions in all situations; they are tools to estimate the probabilities of relationships. Moreover, statistical models are not representations of the "real world." Rather, they are more like calculation machines providing a set of numbers in response to input numbers and instructions provided by the researcher. If the input numbers are good and the instructions are good, the conclusion might be good. Most importantly, the conclusions are not automatically good just because, "The model said so!"

Over time, statistical models have tended to become the mathematical representation of achievement production. In other words, the statistical models now de facto determine the first principles without further consideration of more appropriate principles. What is the first principle inherent in statistical models? The relationship between achievement and all explanatory variables is linear, so more of any explanatory variable will produce more achievement without limit. This principle is a primary source of the contradictions.

Should the researcher trust the conclusions and accept the model or trust the model and accept the conclusions? Can the conclusions be critiqued without fully critiquing the assumptions? Perhaps there is too much trust in the principles inherent in the statistical models and too much acceptance of the conclusions.

In many cases in the natural and behavioral sciences (gravity and the "learning curve," for example), mathematical representations were outgrowths of observations and possible explanations (theories). Only after the mathematical representation is developed are the predictions tested. In statistical analysis, the process is combined; the statistical model is the explanation (theory), the mathematical representation, and the testing mechanism. There is little questioning if the statistical model accurately represents the situation. As soon as the decision is made to use regression analysis, there is no further questioning if the relationships are nonlinear. Virtually all production function studies use regression analysis with the linear relationship principle. There is no follow-up to test the predictions, and the regression results are deemed to be reality. There is ample rationale and evidence to suggest that the achievement relationships are not linear and that nonlinear models should be considered. This is not to disparage these previous works. Without their efforts, it would be impossible to build something new.

There are reasons why a comprehensive, coherent, and unified modeling and testing process can be applied to achievement production. The purpose of this article is to identify those reasons. Are the proposed reasons perfect? No. Are they clear, comprehensive, unified, and coherent? Others will decide. It is not sufficient, however, to merely challenge the principles made herein; it is necessary to replace the principles with those better explaining achievement production and more accurately predicting achievement.

While overstated, there is an underlying truth to the saying: "If you keep on doing what you're doing, you will keep on getting what you're getting." If the same achievement production research is continued, the same conclusions will inevitably result. There seems to be sequence in bringing about change in what Kuhn (1970) calls "normal science."[4] First, there must be a new set of unifying and coherent principles, which become the basis of

https://newprairiepress.org/edconsiderations/vol39/iss1/7
DOI: 10.4148/0146-9282.1104

research. The purpose of the research is to verify the principles. Once the principles are verified, they are used to train people who choose to apply these principles as a part of their profession. If the principles are correct, the research carefully conducted, the training effective, and the professional practice successful, the results will be rewarding.

Proposed First Principles

A set of first principles is proposed to address the contradictions associated with achievement production. The details and rationale for these principles are in the earlier articles. Here they are summarized in a different context, to be a foundation for future research, professional training, and practice.

These first principles were not conceived all at once. When I discovered what I thought was an inconsistency, I looked to a different knowledge base for possible answers. In essence, I was on a journey, which I briefly describe as a part of the first principles. You, the reader, are invited to retrace the journey, in the event you might discover another path.

Principle 1: Nonlinear Relationships

What started my analytical journey was the realization that achievement testing, like light, has its own "speed limit"—a perfect score—and as a consequence, the mathematical relationship between achievement and class size cannot be linear. Most certainly, it cannot be the curve suggested by Glass and Smith. The mathematical functions representing the theory of relativity are based on the idea that one can get closer and closer to the speed of light but can never exceed it. By demonstrating the mathematical difficulties in the Glass and Smith proposition, new thoughts came to mind regarding the nature of the determinants of achievement—the relationships must be nonlinear because there is a test ceiling and floor, and most likely the curve has a maximum and minimum (asymptotic at the top and bottom).

Years ago I heard a talk (I unfortunately do not recall where, or when, or by whom) about providing textbooks to classrooms in poorer parts of Africa. The speaker was raising the question, was it necessary for every student to have his or her own book? He concluded that it was not necessary. Students could share books and by doing so it was possible to save the expense and purchase books in other subjects. He drew a curve estimating the benefits of the number of textbooks—a diminishing returns curve. Ever since that talk, I have tried to identify circumstances where "more resources" do not eventually lead to diminishing returns. I have not identified any. It was important for me to know something about the research on learning, especially the "learning curves." Indeed, there is empirical evidence for a "learning curve," flat at the top and bottom.

By accepting the principle of nonlinear relationships, there are corollary principles.

- Every school has unique circumstances, identified by different points on the curves, meaning there is a different solution for every school rather than a single solution for all schools (principle of regression).
- There is a point where there become diminishing returns for all explanatory variables, rather than constant returns (principle of regression).
- There is an optimal point on each curve, allowing curves to be compared.

By changing one principle from linearity to nonlinearity, many of the contradictions were addressed.

Principle 2: Consistency of Components and Uncertainty of Measurement

In an publication using fixed effects analysis, I obtained a different set of explanatory variables for each year of data (Addonizio and Phelps 2006). There was no reason why the regression results should vary so much year to year. Then I realized slight changes in the correlation matrix would produce substantial changes in the order of significant variables in the step-wise regression results. As a result, the coefficient varied widely year to year. Simply put, basic laws cannot change year-to-year (if they could change by year, they could change by month, day, hour, or minute).

There were too many variables, and they were correlated. Merely entering all possible explanatory variables into a regression equation was not satisfactory; there was no theory driving the decision. The data were collected in categories: Staffing quantity; staffing qualifications; instructional materials; and proxies for socioeconomic status (SES). Rather than all variables working independently, it made more sense to have them working together; e.g., all staff work toward a common goal of achievement. The variables in each of the categories were used as explanatory variables against the various achievement measures. Averaging the coefficients over time addresses the time consistency of variables and consistency of coefficients issues. More importantly, the method represented a better explanation--conceptually similar and statistically correlated variables work together, not individually.

There was a second issue: The coefficient between achievement and an explanatory variable provides one estimate of the relations, but when a second explanatory variable is added, the results change. According to factor theory, two explanatory variables each make a unique contribution as well as a common or shared contribution. In essence, the contribution of any combination of correlated variables cannot be precisely measured. As is the case in quantum mechanics, there is inherent uncertainty of measurement. To deal with this uncertainty, the conceptually similar variables were grouped into factors and transformed indices by combining all the unique and common variance into the index. This provided an estimate of the contribution of the factor and upper and lower limits for each of the component variables.

Then there was the realization that educational research did not have an all-encompassing theory describing how all the various components fit together in a measurable and predictable way. Research mostly focuses on the pieces and not on the whole. Studies using different variables will undoubtedly get different results. Studies using the same variables get different results in different years. In order to estimate the basic laws:

- The basic laws must be comprised of the same explanatory variables although the coefficients can be different depending on grade and subject.
- Conceptually and statistically related variables must be combined in such a way to estimate the contribution of the variables within the group, and thus boundaries for the individual components.
- The coefficients of the basic laws are best estimated by averaging over time.

These principles are not a matter of personal preference; rather, they are a matter of statistical necessity. They explain some of the contradictions in the research--different variables and measures were used.

Principle 3: Accurately Representing Achievement Production

Education pursues multiple goals simultaneously. As a consequence, a single equation is not an accurate representation of the achievement production process, and a different formulation is required.

First, the achievement production system must be represented by simultaneous equations. There must be a separate equation for each achievement outcome and a way to control the cost of each of the variables, again in separate equations. This conclusion directed me to the field of mathematical programming, especially the books by Williams (1985) and Schrage (1991). None of the linear programming models worked because achievement was nonlinear (Principle 1). Was there a function representing the achievement/variable relationship that could be measured through some statistical process and could be solved using simultaneous equations? This became another dinner conversation, and Maris Abolins gave me *An Introduction to Error Analysis* (Taylor 1982). For the first time, I started to understand the reasoning behind the normal curve. I realized that the integral of the normal curve was the appropriate nonlinear function that could be measured by statistical regression. (It has a similar shape to the "learning curve" I was reviewing in another book. Both have the upper and lower limit properties.) All I had to do was find a way to formulate the necessary equations and solve nonlinear simultaneous equations. Back to mathematical programming I went and soon found software capable of accomplishing the task. Earlier software was cumbersome, but Microsoft Excel was easily available and easy to use.

Achievement production must be represented by a set of simultaneous equations representing each goal to be achieved, and must include an equation representing the costs. This addresses some of the contradictions.

Principle 4: Effectiveness Is An Integral Part of Achievement Production

I returned to Taylor (1982) and took note of the section dealing with systematic and random error. As a golfer, I immediately realized my hitting the ball consistently to the right was not random error, it was systematic. Systematic error can be separated from random. I had to correct my systematic error to improve my game. Now my topic became "fixed effect estimation' in econometrics. Because of my role in the Michigan Department of Education dealing with reporting school progress, I wrote the paper, "Measuring and Reporting School and District Effectiveness," (1988) building on my thoughts regarding factor theory and fixed effects. To borrow from my golf swing analogy, schools must correct their "slice" in order to improve student achievement. Including the notion of effectiveness in the simultaneous equations addresses some of the contradictions.

Principle 5: Achievement is derived from behavior

Again, the "eureka" moment came from reading physics, this time about gravity. The discussion was, how long would it take for the effects of the sun's collapse to reach earth? The answer is: At the speed of light. How long will it take for a change of class size to improve achievement? Surely, not at the speed of light. Actually, the change would not even be guaranteed. A change in achievement cannot be related to the number of students in the rooms, it must be related to the behaviors of the teacher, students, and parents. Somehow, the notion of behavior must be incorporated into the explanation and model. This notion explains some of the contradictions in research where the assumption of the regression model is that change is automatic.

Principle 6: Policies and Incentives Influence Behavior

The realization of the effectiveness and behavior notions brought new insights into my appreciation of the work of Walberg and Levin. Simply put, their ideas combined to make a plausible explanation. Policies influence behavior, and behavior influences achievement. In other words, their ideas were the reasonable explanations for the mysterious unobserved fixed effects or effectiveness. Even though there is much more research to be conducted in these areas, they do deal with some of the contradictions.

Principle 7: Policies Are Subject To Cost Constraints

Levin's influence on my thinking was substantial; cost-effectiveness must be included in any explanation of achievement production. With the simultaneous equation formulation, this was easily accommodated. This was the final piece of the puzzle and addresses what is perhaps the biggest incongruity in the regression formulation; that is, it is a basic inconsistency to advocate more of everything where there are fiscal constraints.

I have tried to carefully articulate the first principles in order for the reader to have the full context on which to critique the model.

Implications for Research

Are these principles valid? More accurately, are theses principles generally accepted as explaining achievement production? These principles are intended to be a beginning, not an end. It is important for there be a comprehensive discussion among those who are interested in the topic of achievement production in which they express their views and suggest improvements. As consensus is gained on the principles, attention can then be direct to research, training, and practice.

Are the opposite principles false? In most cases, each of the principles can be expressed in the negative, e.g., the relationship cannot be nonlinear and must be linear. By doing so, the distinctions are sharpened making the analytic process clearer.

Are these principles the foundation of current research, training, and practice? This is highly unlikely. There is little in the research literature regarding comprehensive theory; attention is mostly on specific issues. If I would identify the major weakness of research, it is the lack of consensus regarding the components of the underlying theory. After all, science is the testing of comprehensive theory, not the testing of unrelated assumptions.

Could these principles form the foundation of a new paradigm? Obviously, I think this is the case; it is why I have devoted my time and energies to this project. I wonder if others share this observation?

Could the new paradigm constitute the foundation of a normal science? My experience in academia and in the Michigan Department of Education leads me to believe that the pursuit of achievement excellence is not a scientific matter—it is mostly political. More emphasis is placed on more money and who gets the money

https://newprairiepress.org/edconsiderations/vol39/iss1/7
DOI: 10.4148/0146-9282.1104

than how the money is used to improve the performance of students. Old research methods are repeated in hopes that they will miraculously produce different results.

If these principles are the foundation of the normal science of achievement production, will the practitioners of this normal science adhere to these principles? Schools of education are at a crossroad: Are they a branch of political science where opinion and perceptions are key, or will they move more toward normal science where theory, experimentation, and evidence are key?

As previously noted, the achievement paradigm must be thoroughly tested. First, individual profiles would be established for each school describing their unique situation regarding their standing on resources, SES, and effectiveness. Second, based on this information, the school would be asked to develop a set of policies and evaluate them based on the paradigm model and the predicted gain in achievement, and then select one for implementation. Third, they would implement the policies and collect information regarding the implementation. Finally, the information would be analyzed along with the actual achievement results to identify any relationships. Surely, such a planning process could do no harm. In contrast to the controlled class size experiments, such a regimen would provide a great deal of information upon which to address some of the unanswered questions:

- Do school organizations respond to policy changes, i.e., can good policies change the behavior of the instructional staff?
- What are the successful policies and effective implementation strategies?
- How does a change in instructional staff behavior influence a change in student, family, and community behavior?
- Can school policies influence family and community behavior?
- How are the changes in behavior translated into higher achievement?

Implications for Professional Preparation

Forrester (1980, 11) had some perceptive and instructional observations regarding organizations directly applicable to education:[5]

> For the most part, and in spite of lip service to the contrary, managers are usually decision-makers, not policy makers. The distinction is crucial. People can make decisions without knowing why. Decisions tend to be capricious and are dominated by short-terms pressures. A decision-maker runs an organization, but a policy-maker designs an organization. The distinction is like that between an airplane pilot and the airplane designer. It is the challenge of the designer to create a system that can function as intended in the hands of the kinds of operators who will be available. Seldom are school systems designed. We know that aircraft must be skillfully designed to operate properly, but the same attitude has not yet been generally extended to the much greater complexity of a school system. Here is the challenge and the opportunity for the teaching of management policy—teaching the design of the school systems rather than piloting. Modeling can provide the process for shifting the more responsible levels of management from being school system pilots to school system designers—to shift from coping with day-to-day crises to creating a social system that can be run by ordinary people without continuously recurring crises.

Actually there are many specialized people involved in airplane design: aeronautical, mechanical, and electrical engineers, to name a few. They work together in building a sophisticated product because they were trained within a common scientific paradigm and with particular knowledge and skills within the paradigm. Based on a set of scientific principles and mathematical laws, each discipline is trained to extend the laws to represent new situations.

It is not clear as to what is being taught in universities and what is being practiced in terms of theories and models of improving academic achievement. It is highly doubtful that graduate education students have been asked to solve the Glass and Smith (1978) equations or asked to replicate the results using actual statewide data. If these exercises were attempted, the flaws in the theory and mathematical model would have become apparent. The same can be said of the Hedges et al. equation. Most likely, students are never asked to test the underlying theory and model of achievement production either as a simulation or on actual data. In contrast, a fundamental part of aeronautical, mechanical, and electrical engineer training is the solving both simulated and "real" problems.

Here is a classroom exercise: The current achievement production function is:

$$A = \sum \beta D_{(Z)}$$

where A is Achievement measured in Z-scores; β is the standard regression weight; D is the explanatory variable measured in Z-scores; and Z is the Z-score. The problem: Using the information contained in these articles, sum the possible variables and estimate the value of A for Z = 0 and Z = 1. How much will achievement improve by increasing every variable by one standard deviation? What is wrong with this picture?

A three tier policymaking taxonomy was suggested in earlier articles starting with opinion, progressing to reliance on research, and ending with a comprehensive process of stating the underlying assumptions and evaluating the alternatives. The observations by Forrester tend to explain why most instructional policy-making is based on opinions (tier one) rather than on a common set of skills and knowledge developed from research (tier three). Following the thoughts of Kuhn, this is because there is not a common theory, a common set of laws, and a common methodology guiding research, which is used to prepare individuals to actually apply the theory, laws, and methodology. When there is a shortage of people with requisite knowledge and skills, opinion fills the vacuum. To use Forrester's metaphor, the crew and passengers without the requisite training are designing airplanes rather than the aeronautical, mechanical, and electrical engineers! Before this situation will change, a new set of specialized individuals must be trained. Before the new individuals can be trained, the existing examples of achievement productivity must be replaced with a more functional paradigm with a more clearly defined set of principles, knowledge, and skills.

Please return to and read the "achievement production rules." Engineers could not build aircraft under these conditions; yet schools are expected to "produce" high levels of achievement with multiple sets of ambiguous and contradictory rules. Amazingly, many schools do quite well.

A new achievement production paradigm would have similar characteristics and steps as building an airplane.

(1) What is to be accomplished—the specifications?
(2) What are the applicable laws?
(3) How is the system to be modeled?
(4) What are the initial conditions, and how should these conditions be changed?
(5) How much will the changes cost?

After repeatedly testing and evaluating various simulation models, an actual test model is carefully constructed and extensively examined. After evaluating the results and making the necessary corrections, the model is put into production. After production, the operations are continuously monitored, so improvements can be made. Increasingly, modeling is being used in many types of organizations. Is it possible for modeling to be applied in education?

Implications for Normal Science

Many of the ideas for this series of articles came from Kuhn's thoughts regarding paradigms and normal science. Importantly, these articles are not designed to reach specific conclusions regarding specific variables associated with achievement. Rather, they are designed to propose a different way of thinking about relationships—a paradigm. To follow are some relevant quotes from Kuhn with an explanation of how the proposed paradigm compares with his writing.

> By choosing "paradigm," I mean to suggest that some accepted examples of actual practice—examples which include law, theory, application, and instrumentation together—provide models from which spring particular coherent traditions of research" (p. 11).

This series of papers proposes an achievement production paradigm with an articulated theory, a mathematical law, a practical application, and instrumentation (a process of optimization). Many of the ideas spring from strengths of previous productivity research and, in some cases, apparent contradictions.

> Paradigms share two essential characteristics: 'their achievement was sufficiently unprecedented,' and 'sufficiently open-ended to leave all sorts of problems.' A paradigm 'is an object for further articulation and specification under new or more stringent conditions' (p. 23).

Clearly the theory, law, application and instrumentation is unique compared with other productivity research, and it is open-ended. There is substantial opportunity for further articulation and refinement under wide ranging conditions.

> To be accepted as a paradigm, a theory must seem better than its competitors, but it need not, and in fact never does, explain all the facts with which it can be confronted (p. 18).

Theories and mathematical models are representations of a phenomenon, and, hence, not the "real thing." Therefore, theories and models must be judged based on: (1) How well they explain the phenomenon; (2) how well they predict the outcome; and (3) how well the prediction can be verified.

A "policy behavior achievement" (PBA) paradigm is a better explanation of achievement production than a "resource achievement" prescription for a fundamental reason: Achievement is a form of behavior, and school behavior is directly influenced by policy. If, over time, the behaviors of the teacher and students change, then an improvement in achievement is likely. However, it is more likely for behaviors to change through wise policies.

Regarding the ability to predict achievement, the PBA paradigm is more accurate then the "resource achievement" prescription for several reasons. First, the PBA paradigm recognizes the ceiling effect of achievement and includes a law more accurately representing that characteristic. Second, it includes data regarding the effectiveness of existing policies even though the data are derived indirectly rather than observed. Because the effectiveness variable explains a considerable amount of the variance, its inclusion makes the predictions of achievement more accurate.

The PBA paradigm allows for, indeed requires, the testing of various theories or scenarios through the simulation process not available with other theories or models. This is possible because of the nonlinear functions enabling the use of simultaneous equations and the inclusion of cost as a variable. With a refined model identified, a comprehensive experiment can be conducted. This is not the case with existing achievement production theories and models.

> 'Normal Science' means research firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice. Today such achievements are recounted [by textbooks], elementary and advanced. These textbooks expound the body of accepted theory, illustrate many or all of its successful applications, and compare these applications with exemplary observations and experiments (p. 10).

Achievement production has not yet become a "normal science," as suggested by Kuhn, because there is no accepted paradigm or successful applications. Students are not asked to solve simulated problems replicating successful applications as students of engineering are asked to do.

> The study of paradigms...is what mainly prepares the student for membership in the particular community with which he will later practice (p.11).

As some point, after further articulation and refinement, the PBA paradigm could be valuable as a subject for professional training and practice.

> Men whose research is based on shared paradigms are committed to the same rules and standards of practice (pp. 10-11).

It is unclear what the current rules and standards of practice are. It is unlikely that some form of unification will take place until there is a unification of purpose among many institutions including universities, departments of education, foundations, and other organizations interested in improving the academic performance of students. For example, it is doubtful whether the various areas of education preparation—curriculum and instruction, administration, social foundations, finance—agree on common research and teaching efforts based on the same model.

> In the absence of a paradigm...all of the facts that could possibly pertain to the development of a given [phenomenon] are likely to seem equally relevant. As a result, early fact-gathering is a far more nearly random activity (p. 15).

The many contradictions in the research conclusion suggest that current fact gathering is a "nearly random activity." As the critique of the paradigm evolves, the shortcomings of the data being collected would become apparent, and there would be more specific purposes for refining the collection process.

https://newprairiepress.org/edconsiderations/vol39/iss1/7
DOI: 10.4148/0146-9282.1104

It suggests which experiments would be worth performing (p. 18).

Based on the paradigm, there are several immediate questions worthy of further investigation:

- Is there an achievement ceiling effect?
- Is the relationship between achievement and the determinants nonlinear?
- Is there an appropriate nonlinear measurement of effect size?
- Do individual school circumstances matter in improving achievement?
- Are some schools more effective in producing achievement?
- What make these schools more effective?
- Do policies influence behavior?
- Do behaviors influence achievement?

The "Policy-Behavior-Achievement" Paradigm as Normal Science

According to Kuhn, normal science is the articulation of the theories already supplied by the paradigm. It is "the empirical work resolving some of its residual ambiguities and permitting the solution of problems to which it had previously only drawn attention" (p. 27). There are substantial questions regarding the responsibility for expanding the knowledge of the normal science of achievement production. In other disciplines, the responsibilities of the various institutions are far clearer, heavily relying on the efforts of higher education. What are the responsibilities of universities, departments of education, and other institutions interested in educational policy?

Universities are guided by three major purposes--teaching, research, service. The PBA paradigm is a possible vehicle for addressing all these purposes in preparing school policymakers. First, the necessary data for seeding the model are available from departments of education. The examples in these papers are from Minnesota Department of Education. The method to prepare the data for seeding into the model is described by the author in a 2009 article titled, "Reporting and Measuring School and District Effectiveness." The information for the profile, estimates of effectiveness, and the boundaries for the factors come from these data. Replicating this information could be a practical exercise for graduate students as a part of their statistics training, but state departments of education have the responsibility for the data and presumably for reporting this information to policymakers and the public. From my experience, there is little collaboration in this effort. Working together would be a good start.

With the necessary data available, all university departments contributing to graduate education could use the PBA paradigm to investigate the achievement policymaking process by means of the simulation model. The materials presented in the classroom, readings, and individual research would provide background for exploring various policy options. Rather than writing papers, the students would be asked to "test" the policy options using the simulation model. The very process of exploring policy options has value. The product of the exercise would be a critique of various policies, leading to the development of an achievement improvement strategy.

There are opportunities for the faculty and student to improve the paradigm by focusing on the theory, laws, applications, and instrumentation. Also, testing selected policy options in an experimental setting would also be valuable. From these experiences, a collection of case studies, valuable in the teaching process, would evolve. Even if a final testing of the strategies did not transpire, identifying and testing the underlying assumptions has value in developing skills and knowledge.

After over 25 years, my journey is at an end. It is possible to combine several seemingly unrelated aspects of achievement production into a single explanation and make predications based on that explanation. Indeed, achievement, various resources, SES, different notions of effectiveness, and cost can be coherently unified and incorporated into a method to predict changes in achievement. My original dream has been fulfilled. This is not to say that I have found THE answer, merely AN answer. It would be most gratifying if others would find better explanations and models, and better yet, use the explanations and models for training and in practice.

For those who have managed to wind their way through the morass of data and arguments, some might be disappointed because there is no definitive conclusion regarding the influence of class size or resources. Others will be disappointed because it is too complicated. Hopefully there will be a few who will see a future for these ideas. To me, the purpose was the journey and not the destination; it changed my way of thinking! Improving achievement is complex, requiring an explanation and model commensurate to the task. The ideas of the paradigm were emphasized in order to encourage researchers, trainers, and practitioners to broaden their thinking away from the traditional issues—lower class size or more money—to the holistic issue: How can a complex organization be designed and operated to reach its achievement goals? As it has been emphasized repeatedly, the focus must be on critiquing the underlying principles and not accepting "common-wisdom" conclusions.

Like Newton, "I stood on the shoulders of giants," such as Henry Levin, Herbert Walberg, Eric Hanushek, John Taylor, Thomas Kuhn, Linus Schrage, and Hilary P. Williams. Ironically, Glass and Smith were instrumental in molding my thinking (even though we disagree on the conclusions). I benefited substantially from their ideas and incorporated them freely. They deserve credit for building the foundation.

References

Achilles, C.M., B.A. Nye, J.B. Zaharias, and B.D. Fulton."The Lasting Benefits Study (LBS) in Grades 4 and 5 (1990–1991): A Legacy from Tennessee's Four-year (K–3) Class-size Study (1985–1989)." Project STAR. Paper presented at the North Carolina Association for Research in Education, Greensboro, North Carolina, January 14, 1993.

Addonizio, Michael, and James L. Phelps. "How Much do Schools and Districts Matter? A Production Function Approach to School Accountability," *Educational Considerations* 33 (Spring 2006): 51-62.

Bohrenstedt, George W., and, Brian M. Stecher, ed. *Class Size Reduction in California: Early Evaluation Findings, 1996-98*. CSR Research Consortium. Sacramento, CA: California Department of Education, June 1999. http://www.classize.org/techreport/index.htm.

_____________. *Capstone Report: What We Have Learned about Class Size Reduction in California*. CSR Research Consortium. Sacramento, CA: California Department of Education, August 2002. http://www.classize.org/techreport/index-02.htm.

Forrester, Jay W. "Systems Dynamics: Future Opportunities." In *Studies in Management Science*, ed. Augusto A. Legasto, Jay W. Forrester, and James M. Lyneis, 7-21. Vol. 14, Systems Dynamics. Amsterdam, North Holland: 1980.

Glass, Gene V., and Mary Lee Smith. *Meta-analysis of Research on the Relationship of Class-size and Achievement*. San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23 (April 1994): 5-14.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press, 1970.

Levin, Henry M. "Cost-Effectiveness and Educational Policy." *Educational Evaluation and Policy Analysis* 10 (Spring 1988): 51-61.

____________. "Raising School Productivity: An X-Efficiency Approach." *Economics of Education Review* 16 (June 1997): 303-311.

Phelps, James L. "Measuring and Reporting School and District Effectiveness." *Educational Considerations* 36 (Spring 2009): 40-52.

Schrage, Linus E. *Lindo: An Optimization Modeling System*. San Francisco, CA: Scientific Press, 1991.

Silkman, Richard H., ed. *Measuring Efficiency: An Assessment of Data Envelopments Analysis*. San Francisco, CA: Jossey-Bass Inc., 1986.

Taylor, John R. *An Introduction to Error Analysis*. Mills Valley, CA: University Science Books, 1982.

Walberg, Herbert J. "Improving the Productivity of America's Schools." *Educational Leadership* 41 (May 1984): 19-27.

Williams, Hilary P. *Model Building in Mathematical Programming*. 2nd ed. New York: John Wiley, 1985.

Endnotes

[1] All subsequent references to Glass and Smith in this article refer to Gene V. Glass and Mary Lee Smith, *Meta-Analysis of Research on the Relationship of Class-Size and Achievement* (San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978).

[2] All subsequent references to Walberg refer to Herbert J. Walberg, "Improving the Productivity of America's Schools," *Educational Leadership* 41 (May 1984): 19-27.

[3] All in standard scores with lines passing through the Z-score coordinates of 0, 0.

[4] All subsequent references to Kuhn in this article refer to Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press, 1970).

[5] Note that I have substituted "school system(s)" for "corporation(s)," and "modeling" for "system dynamics" in the quotation.

Acknowledgments

Over the years several friends and colleagues have listened to my thoughts and have given me valuable perspectives. They are listed below.

An indispensable companion during the writing phase has been a senior at Macalaster College in St. Paul, Minnesota, Michelle Neary. A superior student and an accomplished musician, she edited the papers for submission, and more importantly, made many perceptive suggestions for improving the clarity of logic and writing. I am proud to say she is our granddaughter. After graduation, Michelle plans to pursue a Ph.D. in chemistry.

Over the many years of pursuing these issues, my lovely, supportive, and helpful wife of 50 years frequently asked, "When will you be finished?" I can reply, "Now." However, I hope the issues and questions raised in this series of papers will motivate others to continue the journey.

My deepest appreciation to:

- Charles Greenleaf, former Michigan and White House official for education, and dear friend who got me started;
- William Milliken, former Governor of the State of Michigan who gave me my most professionally rewarding opportunity;
- Maris Abolins, Professor Emeritus of High Energy Physics at Michigan State University, and dear friend who opened my eyes to new ways of thinking;
- Doug Roberts, former state education official, former State Treasurer of Michigan, and dear friend who always encouraged my unorthodox efforts;
- John Porter, former Michigan State Chief School Officer, who started my thinking about educational accountability models;
- Wilbur Brookover, former Professor of Sociology at Michigan State University and dear friend who tutored me regarding the importance of community and family in improving achievement;
- Mike Addonizio, Professor of Education at Wayne State University, a colleague and frequent coauthor;
- Mal Katz, former state education official, who frequently challenged me on the relationship between laws of achievement and individual circumstances in education;
- Mike Moch, Professor of Business at Michigan State University, who tutored me regarding the role of policy and behavior in organizations.

https://newprairiepress.org/edconsiderations/vol39/iss1/7
DOI: 10.4148/0146-9282.1104