The Logic of Program Evaluation: What Should We Evaluate in Teacher Enhancement Projects?

George E. Hein,*
Director, Program Evaluation
and Research Group
Lesley College

Introduction

Tip O'Neill, the late, former speaker of the House of Representatives, said, "All politics is local." So is education. No matter how grand our educational schemes for systemic change, they need to be assessed for their local impact; for their effect on children. The purpose of public schools is to educate children. Any change—any scheme to reform schools—must be judged by the extent to which it helps children become fully functioning, competent, educated adults.

However, the truism above does not immediately translate into the conclusion that evaluation of educational programs must necessarily involve an assessment of the extent to which a program has resulted in increased learning by children.[1] In fact, I will argue in this paper that the proper subject for the evaluation of teacher enhancement efforts is the impact of

---

[1]A current, controversial issue concerns whether children's learning should be assessed through traditional, standardized tests or other means. Important as that debate may be, it is not the subject of this paper. In this paper I shall only explore *whether* student assessment is appropriate as part of the program evaluation task, not *how* student performance should be assessed.

these efforts on *teachers.* The argument will be based on the analysis of a simple, visual model of the structure of schooling.

A Structural Model for Schooling.

 The U. S. education system can be conceived as a hierarchical structure ranging from the national level to the level of children as indicated in Figure 1. This figure represents the *logical structure* of the national education system, not necessarily the relative importance of the various levels. It emphasizes the systemic, bureaucratic elements of the system, not all the aspects of society that contribute to children's learning.[2]

The model is particularly appropriate for this discussion of the logic of evaluation of teacher enhancement programs because much of the initiative and most of the funding for reform, no matter what level targeted, comes from national and state sources. As far as program evaluation is concerned, the national and state funding levels play a disproportionately significant role. Most of the local school funds—the property taxes that supply the basic, day to day operating expenses of schools—are usually not subject to external program evaluation.

As indicated in the introduction, all interventions, no matter what their intended level, must ultimately be justified by their effect on children's learning. Thus, the shaded section emphasizes the level of the child.

The schematic diagram also indicates the traditional view of the education system as a hierarchical structure in which policies developed at any level are likely to influence those levels below[3] it. There is an expectation that national

---

[2] The United States is one of the very few countries in the world that does not have a formal, national school system and a national curriculum. The actual organization of education in the United States consists of approximately 15,000 autonomous, local school districts, with state and national requirements superimposed for only a fraction of their educational policies and practices.

[3] The terms "below" or "above" (or "higher" and "lower") applied to this model refer to the level of aggregation of the system, they have no value implications. A state system of education is a "higher," i.e. larger, organizational level than a single school, although it is farther removed from the children.

educational policy will somehow trickle down to the level of the child through the various layers of the structure. An open question, and one that is currently much discussed by writers about organizational change[4] is the extent to which reform at any lower organizational level can and needs to influence higher organizational levels in order to bring about systemic change. The arrow along side the model indicates these expectations of influence, with the question mark emphasizing that the degree of influence from lower to higher levels is open to debate.

The schematic in Figure 1 can also be inverted to emphasize the importance of the child in education; the lowest organizational unit is, in fact, the most important one, as illustrated in Figure 2. In addition, the many components that influence education, such as family and community, business and industry, and, especially, various agencies dedicated to improving education, can be added to provide a more complete picture of the structure of the U. S. education system. This is illustrated in Figure 3.[5]

Examples of national interventions to bring about school change abound. When the Secretary of Education publishes a document such as *A Nation at Risk* the intention is to influence education across the country. Examples of other initiatives intended to influence education nationally would be the National Standards effort, movements to require (or ban) prayer in schools across the United States, or Supreme Court decisions that influence racial distribution of children nationally.

Another set of efforts is directed primarily at the state-wide organizational level. Such initiatives may arise within a state, such as the various state frameworks and state assessment programs instituted in the past decade, or they may result from a distribution of federal funds to the states, such as the

---

[4] M. Fullan, 1991. *The New Meaning of Educational Change*, 2nd. edition. New York: Teacher's College Press.

[5] A much earlier, but similar, model is provided by Dewey in *The School and Society,* 1900.

NSF sponsored Statewide Systemic Initiative (SSI). In the latter program, the NSF has targeted individual state school systems as the level of intervention. How the state actually uses the funds to implement school reform—by carrying out state-wide activities or initiatives at the level of individual districts or schools—is determined by the proposals submitted to NSF by individual states.

Evaluation of any Proposed School Reform

Applying the model outlined above to the evaluation of school reform basic questions concerning any school intervention can be formulated:

    1) At what level(s) is the intervention targeted?

    2) What is the strategy for influencing this level?

    3) To what extent does the program influence the targeted level?

    4) What is the logical argument that an intervention at the targeted level will influence the education of children?

    5) Is there a strategy for influencing all levels?

All of these questions are significant for any intervention, and each requires careful attention. Program implementers must be clear about what level (or levels) they propose to influence by their actions; they must have an action plan, specific mechanisms and tactics for accomplishing what they propose to do; and they need to be clear about why they believe this particular set of activities has the potential to make a difference for the education of children.

But an evaluator's perspective is different. Not every step in the logic of the argument that is intended to result in improved education for children is a component of the evaluation of a specific program targeted at a particular audience. Program evaluators have two major concerns. One is a need to understand the logic of the program being evaluated. What is intended, how is it supposed to work? Although there have been suggestions that evaluation should be goal free and that evaluators might even intentionally be ignorant of program goals[6], most of us believe that unless we understand a program's

---

[6] Scriven, M. 1972. Pros and cons about goal-free evaluation. *Evaluation Comment 3*(4), 1-7.

basic assumptions, we cannot adequately evaluate it.[7] The rationale for a program is embodied in the answers to questions 1,2, 4 and 5 above.

The second concern of the evaluator is to document and evaluate the actual program activities and consequences. In order to carry out the evaluation, the evaluator needs to understand the logic behind the program, but the evaluation work itself focuses on the activities of the program and the outcomes associated with those activities. In carrying out the evaluation, the evaluator will address primarily question 3, above.

An example will illustrate the situation. <u>Goals 2000</u>, first an activity sponsored by the National Governors' Association and now a federal school reform package passed by Congress and signed into law by President Clinton in l994, is a *national* initiative intended to change the education of children. The intervention itself, however, is targeted at the level of the *states*—The Goals 2000 Program requires each state to submit statewide plans that include standards and assessments to improve schools. It's strategy for change is twofold, it distributes money to states so they can carry out their plans, and it lends moral force to the efforts around the national goals[8]. The argument that the funding strategy will effect the education of children will, presumably, be contained in the individual state plans. If history is any guide, this logic may be somewhat tenuous, relying heavily on assumptions about education.

But from an evaluator's standpoint, the evaluation of the program is straightforward. The first task would be to determine whether the actual program—the plan to distribute funds to the states in response to their

---

[7] In practice it is usually the case that program implementers chose evaluators who share their basic beliefs about education.

[8] The original Governors' plan consisted of six goals, the federal Goals 2000 legislation lists eight goals, and the program has recently been described as consisting of "10 elements" (ERIC Review, *Systemic Education Reform*, *3* (2) Fall, l994, p. 10, U. S. Department of Education.) The important point here is that most of the goals refer to student outcomes such as student readiness for school and student achievement in various subjects.

plans—has in fact occurred.[9] The second level would be to examine the state
plans and see if the funds had accomplished what the plans had proposed.
Most likely, the funds will be used for activities such as teacher education,
development of assessment systems, school building and program
improvements, rather than direct services to children. It remains to be seen
whether the plans call for direct state support of activities at the level of
districts, of individual schools, or of new entities that are, in turn, supposed to
carry out activities within the state.[10]

The logic of the Goals 2000 program as I have just described it is contained in
Figure 4. One task would be to evaluate process X, the mechanism for
distributing funds to the states. This corresponds to questions 1) and 2)
discussed above. Another task would be to examine the various state proposals
and then evaluate the processes they propose to achieve the Goals 2000. One
state (illustrated by arrow A) may propose to provide funds directly to
individual schools. Another (arrow B) may chose to intervene at the district
level, while a third (arrow C) might set up an intermediate entity with the
intention of influencing districts, schools or teachers. Again, the evaluator's
responsibility would be to assess the proposed programs, A, B, or C, and to
determine the extent to which they were able to carry out the activities
described by each. In addition, an evaluation of the Goals 2000 Program will
provide information on the relative effectiveness of the different strategies,
A, B, or C, in accomplishing project goals.

It should be incumbent on program staff to articulate why they believe that
the activities they carry out—whether teacher enhancement, local school
restructuring, or new state mandates that apply to all school systems—can be
expected to lead to changes in students' performance. That is, they need to

---

[9] In this brief discussion we are not distinguishing between *formative evaluation*, which
would be used to improve the funding mechanism in response to state plans, and
summative evaluation, which would report on the results of the funding program.

[10] An example of such an organizational structure would be regional centers (as utilized
by some SSI Programs) to facilitate implementation of their programs in different regions
of a state.

address questions 1, 4, and 5 above. But it is not logical to expect that program evaluations of the state programs will directly answer this question.

The reason there is a logical leap between program evaluation and assessment of children's progress in school is that there are a host of local, environmental factors that may intervene between a program's activities and student outcomes. *Although a valid rationale for any program's impact on children is a <u>necessary</u> component to justify an intervention at any level, it is not a <u>sufficient </u>one to assure an observable outcome at the level of the child.* In any particular instance, a program that could, in principle, benefit children may be overwhelmed in a state that has cut overall support for schools; another program may be badly implemented because of local teacher-administration labor disputes; and the positive effects of a third program may be overshadowed by a larger school reform effort that occurs simultaneously. Program evaluation that focuses at education levels other than those targeted for intervention may miss the significant features of that particular reform effort. It must examine implementation where it happens, not indirectly through its (possible) effect on children, who may be several levels removed from the "action."

An analogous situation would be provided by an evaluation of a particular intervention to treat tuberculosis victims with an anti-TB medication at a hospital outpatient clinic. In this instance, the clinic can be compared to a school setting, and the program to distribute the drug to a school-based intervention. The actual efficacy of a particular chemical for the treatment of a disease, will be determined by:

a) The evidence (logic of the argument) that connects that particular chemical structure with an effect on the course of the disease, and

b) The environmental factors that determine its value in any specific situation. A drug that is effective in principle may still not be effective in practice because of expense, difficulty in administration, public antipathy towards the treatment or a host of local confounding conditions. The appropriate level of analysis for this program would be the treatment program

at the clinic. Evaluation questions could focus on such issues as, was the
correct patient group targeted for treatment? Did the staff follow up on
treatment? Were patients satisfied with their care? Did the patients stay with
the regimen?

If the evaluation focused primarily on an analysis of the changes in the
incidence of TB in the community, then it would be impossible to ascertain
whether any changes (or lack of them) were caused by the outpatient
treatment program, by to the efficacy of that particular chemical, the
incidence of a new drug resistant form of TB, the fact that patients were or
were not taking their medication, or any one of a number of other
environmental and economic factors influencing the course of the disease. In
order to carry out a study to cover all these factors, it would be necessary to
follow the lives of individual patients and be able to isolate the various factors
that influence their health.[11]

The Model Applied to Local Systemic Reform
The NSF Local Systemic Initiative Program (LSI) is intended to bring about
reform through intervention at the school district level. The intention is
similar to the goals of the Urban Systemic Initiative (USI), and, indeed, the LSI
program solicitation cross-references the parallel USI Program. The
solicitation goes farther. It not only specifies the unit within which reform
should be implemented, but also points out that systemic change involves an
entire organizational unit[12] and then identifies school districts as the target

---

[11] The analogy is of limited applicability because in clinical drug evaluation no
substance can be tested in community settings (analogous to a school intervention) before
both the safety and efficacy of the drug have been demonstrated in controlled, laboratory
settings. In education we sometimes initiate new processes at various levels without
compelling evidence for their value at *any* level.

[12] "Systemic change projects are characterized by a shift in focus from the professional
development of the individual teacher to targeting the professional development of the
teacher within the whole school organization." This and subsequent quotes are from the
NSF Publication, *Local Systemic Change Trough Teacher Enhancement, Grades K-8,*
Program Solicitation and Guidelines, Draft, 1994., p. 1.

for the solicitation, emphasizing that they should focus on teacher enhancement for teachers from an entire district.[13]

The LSI Program has an added specification, as is evident from its placement within the category of teacher enhancement projects. It specifies that the school reform should take place through "a focus on teacher enhancement with the attention to the implementation of exemplary instructional materials." (*ob. cit.,* p. 1) The solicitation does not say anything, for example, either about how school districts are organized within a state or how they relate to state initiatives (a "higher"[14] level of school structure than is included in this program), nor about how the districts need to connect their teacher enhancement activities to activities undertaken by children (a "lower" organizational level than the one addressed by this solicitation.) All the discussion in the solicitation refers to the school district as the target for the intervention.

Using our model, the LSI Program can be illustrated as indicated in Figure 5. The program will fund a series of projects that will operate either by mechanism X or Y to carry out activities in districts. All the projects will have some impact on teachers, either directly (path A), through schools (path B), or some combination of these two mechanisms with the expected outcome to be some form of teacher enhancement.

Evaluation of any LSI project can reasonably be expected to determine the extent to which these activities have been effective in bringing about the proposed teacher enhancements.[15] It is reasonable to expect evaluations to

---

[13] The actual language is more specific, indicating that a subset of teachers may be chosen, but stressing that a cross-school effort is expected, "Projects supported by this solicitation could be designed to impact all teachers of mathematics, science and technology or a subset such as: all teachers of mathematics and science at certain grade levels all teachers of mathematics or science; or all teachers in a particular set of schools within a large system." (*ob. cit.,* p. 1).

[14] see footnote 2.

[15] Again, this is a reference to the summative evaluations. Evaluators should be able to provide guidance and advice to projects based on data gathered during the course of the evaluation so that the projects can become more successful at achieving their goals.

document changes in teacher beliefs and behaviors as a result of participation in the LSI projects. If both paths A and B are used, information may be generated on the relative efficacy of the two paths in fostering these changes. In the course of their evaluation activities, evaluators are likely to obtain evidence concerning environmental factors that influence the extent to which the teacher enhancement activities are likely to have an impact at any level. A dysfunctional school, in which teachers and the principal are feuding, may not be able to show the results of the enhancement activities on the teachers, and certainly not on the students one level further removed.

Impact on students, (but are the kids learning anything?)

Despite the argument given above, that program evaluation must be focused at the level of the intervention, all of us concerned about education still need to worry about the most important members of the school structure, the children. Activities proposed under the LSI initiative are only worth carrying out if there is reason to believe that they have the possibility of improving the education of children. An answer to Question #4 above, is necessary both to justify the project and to permit the evaluator to carry out an appropriate evaluation.

 The RFP for the LSI Program makes a series of statements that assume a relationship between teacher behavior and student outcomes. It argues (*ob. cit.,* p. 1) that the following attributes of school systems are of value:

• Creation of professional communities where teachers are empowered to bring about change and encouraged to reflect on their own teaching and learning.

• New beliefs, new skills and new behaviors must be learned and explored [by teachers] within a supportive school culture which itself is engaged in renewal.

• Just as students should learn mathematics, science and technology through the process of inquiry, so too should their teachers.

These teacher behaviors are linked to "making significant progress towards reaching national goals for the teaching of mathematics, science and

technology education." (*ob. cit.,* p. 1).   The rationale for the connection between these teacher behaviors and student outcomes comes from a whole series of research studies and policy papers that delineate the appropriate teacher behaviors that lead to desired student outcomes.[16] These studies fall into two general categories:

1. Research findings based on some form of controlled setting. Whether the research involved a treatment and control group or whether it was more naturalistic, the essence of any research study that tries to make a connection between teacher behaviors and children's learning necessarily requires a rather narrow focus. The researcher specifies a set of conditions and then notes what happens.

2. Descriptions of classroom life. Over the past twenty years we have been provided with a wide range of descriptions of "real" school life.[17] Some of these have been horrendous examples of the worst of our classrooms; others have been glowing accounts of stellar efforts to educate children.

For this discussion, what we can generalize from both kinds of studies is that any proposal to develop evidence concerning the connection between teacher behavior and student learning is major activity by itself. To attempt it within the framework of the LSI projects would be well beyond the scope of most program evaluations associated with NSF funded teacher enhancement projects. If the LSI projects attempt to carry out such research, they will have neither time nor funds left to carry out what they are intended to do: bring about systemic changes in teacher beliefs and behavior. Within the NSF there are other programs specifically committed to funding research on student learning and how it relates to teacher behavior. Our knowledge base in this

---

[16] The list is too long to enumerate here. For work in science education, a useful summary of much of the relevant research can be found in several chapters in D. L. Gabel, editor, 1994. *Handbook of Research on Science Teaching and Learning,* New York: Macmillan.

[17] The Easley and Stake Case Studies in Science Education (Stake, R. W. And J. Easley, 1978. *Case Studies In Science Education* , Champaign, IL: U. of Illinois, CIRCE) are an outstanding example, as are the more recent case studies of California Mathematics teachers, (*Educational Evaluation and Policy Analysis*, 12, #1, Fall, 1990.)

area may be inadequate, but it will not be enhanced if program evaluators for LSI projects are asked to carry out studies beyond the means available to them, and studies that are primarily late extensions and additions to projects that were conceptually developed for another purpose.

Conclusion

The evaluation of Teacher Enhancement efforts must focus on observable outcomes for teachers. Projects must explain how they expect these changes to impact students. The research base that links teacher behavior and student learning is still inadequate. Additional research studies are necessary to provide more evidence for this connection and to describe more fully the ecological circumstances that influence it. But it is important to make the distinction between the capacity of project evaluators to provide information about implementation efforts—and to assist in improving these through formative evaluation—and studies that attempt to link teacher behavior with student outcomes.

If evaluators are asked to assess teacher enhancement programs by the student outcomes that may or may not be associated with them, they will produce primarily uninterpretable data without being able to develop a better understanding of the factors that encourage or hinder teacher enhancement.
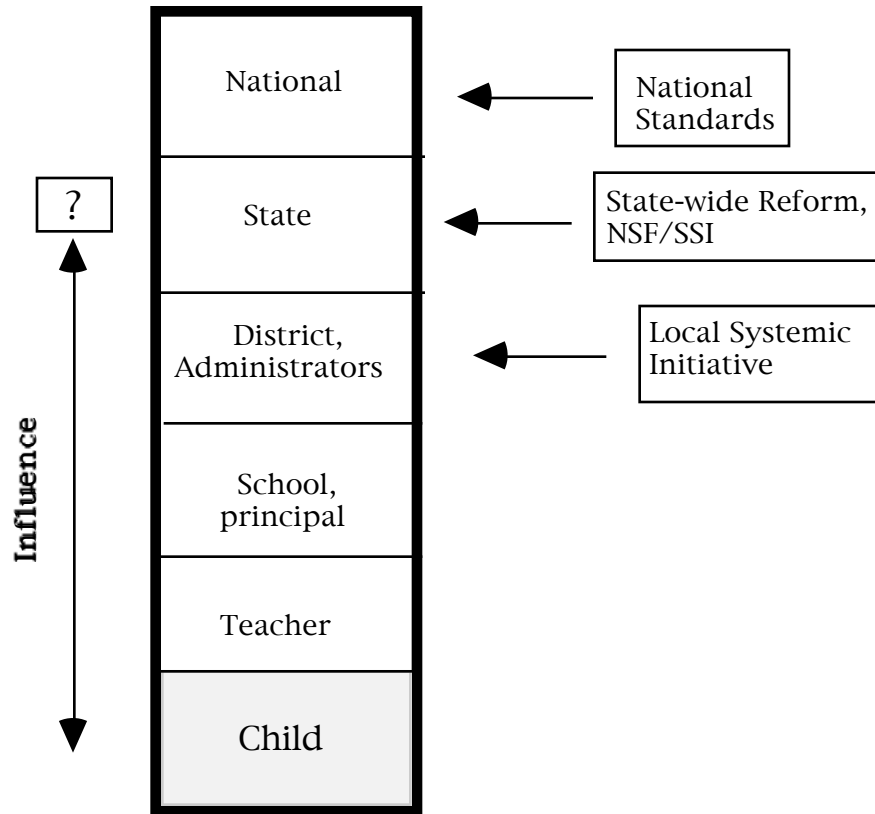
Figure 1

```
                                    ┌──────────────┐
            ┌───────────────┐◄──────│ National     │
            │   National    │       │ Standards    │
            │               │       └──────────────┘
            ├───────────────┤       ┌──────────────┐
    ┌───┐   │    State      │◄──────│ State-wide   │
    │ ? │   │               │       │ Reform,      │
    └───┘   ├───────────────┤       │ NSF/SSI      │
      ▲     │  District,    │       └──────────────┘
      │     │ Administrators│◄──────┐ ┌──────────────┐
   Influence│               │       └─│ Local Systemic│
      │     ├───────────────┤         │ Initiative    │
      │     │   School,     │         └──────────────┘
      │     │  principal    │
      │     ├───────────────┤
      │     │   Teacher     │
      │     ├───────────────┤
      ▼     │    Child      │
            └───────────────┘
```

Figure 2

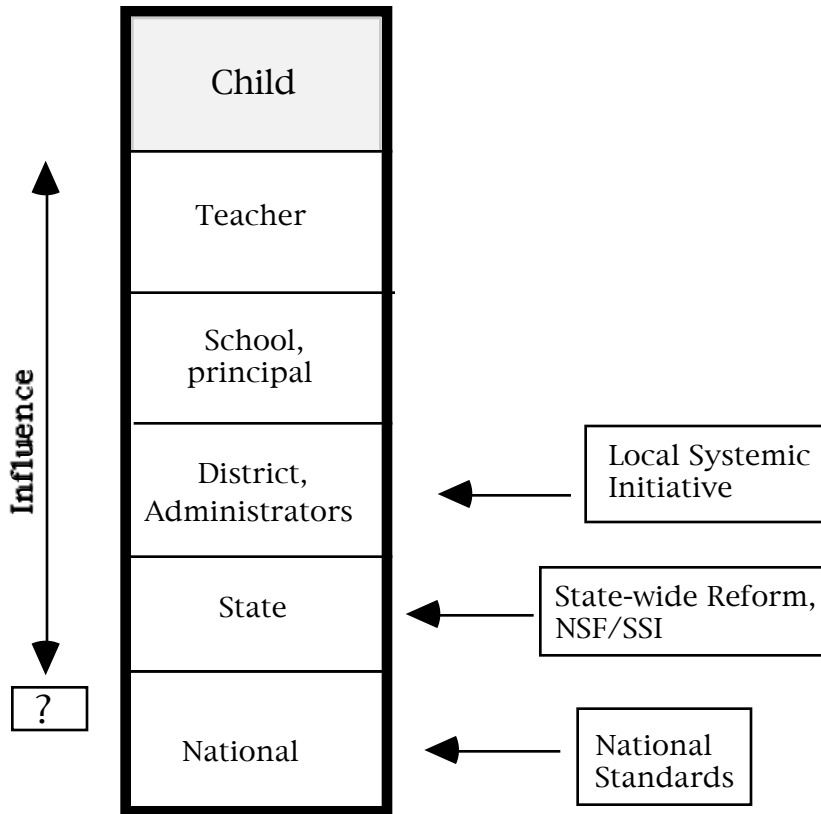## School   Structure
(Child   Centered)
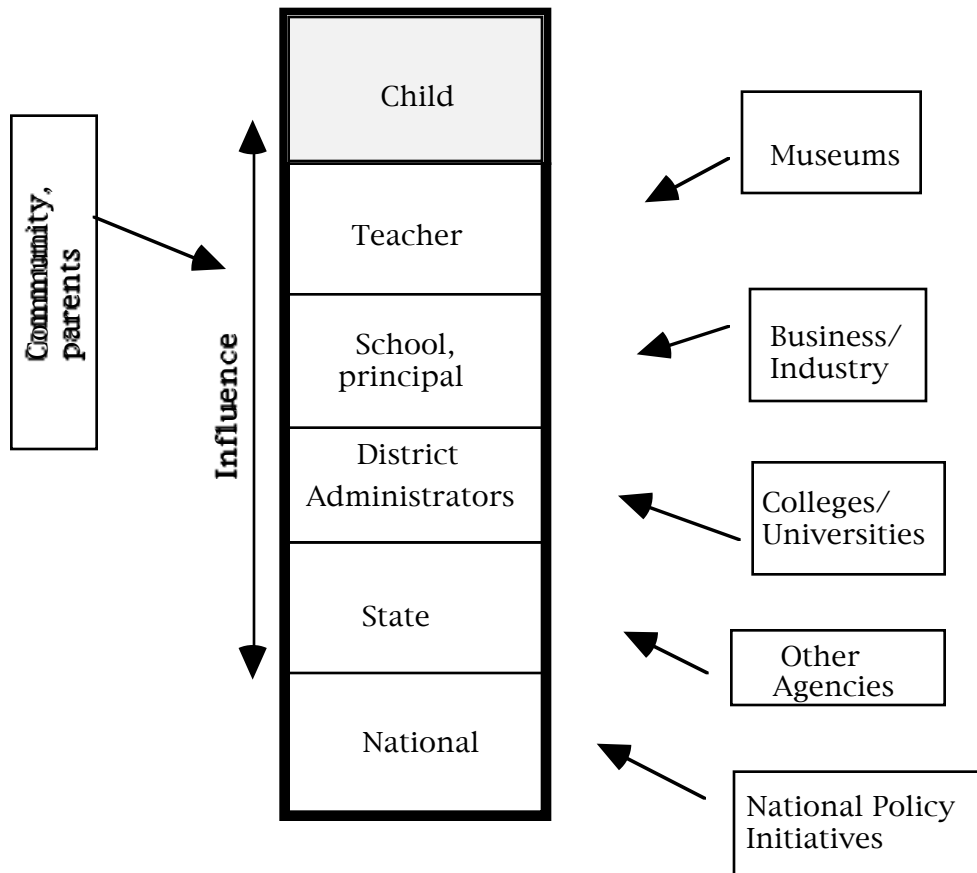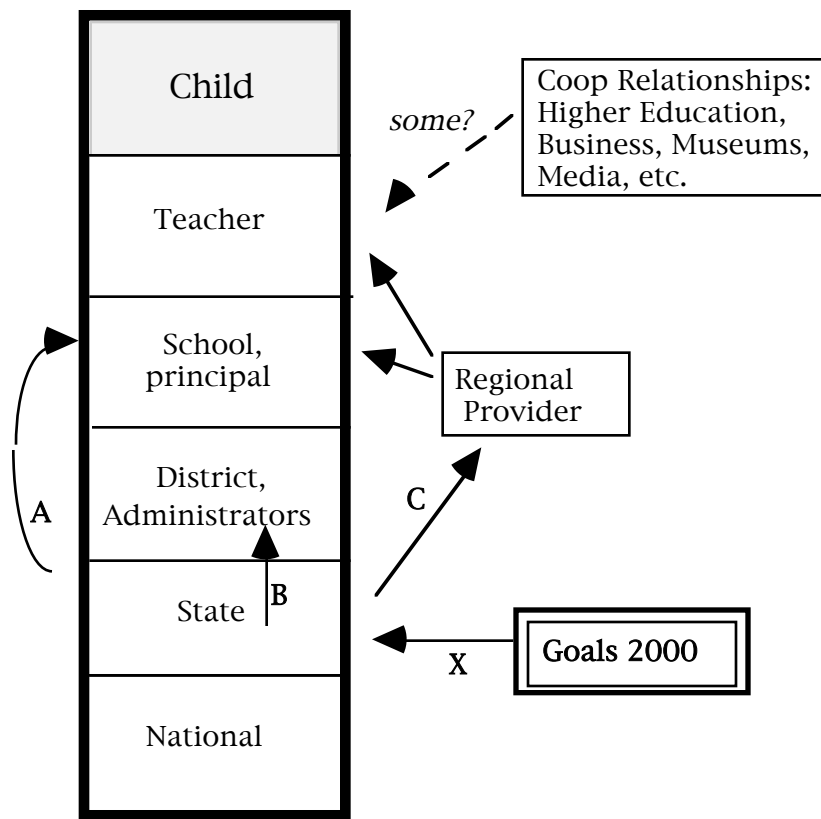
Figure 3

Figure 4
Hypothetical Goals 2000 Structure to Guide an Evaluation Plan

Figure 5
Schematic View of LSI Projects