The Impact of National Standards and Statewide Assessments on Curriculum:

A Proposed Mechanism for Subsystem Dynamics within the National Science Foundation "Driver 1"

David Gibson, Ed.D.

Abstract

A systems dynamics perspective and modeling approach is used to explore a proposed mechanism for the impact of national standards and statewide assessments on curriculum in large systems such as urban districts and states with 100,000 students or more. Behaviors of the model under various initial conditions and changes in key variables are observed under simulation, with brief discussion of each situation.

Introduction

Driver 1 of the National Science Foundation states that the implementation of standards-based curriculum and assessments can raise the performance of all students in science, mathematics and technology. At The Vermont Institute for Science, Mathematics and Technology (VISMT), we have begun to conceptualize some of the specific mechanisms that might bring this about. This paper describes a subsystem model within Driver 1 that deals with the relationships between “assessment” and a “curriculum gap” formed in an envisioned network of relationships at the heart of the driver.

The complete VISMT model of Driver 1 involves five subsystem components, beginning with 1. the impact of assessment, 2. a gap between the assessment and the curriculum, 3. the implications of the gap for professional development, 4. the provision of learning opportunities, and 5. student achievement as measured by the assessment. (Figure 1) The model assumes that the five subsystems are linked hierarchically with connections running both "up" and "down" the system. Signals running "down" the system feed the

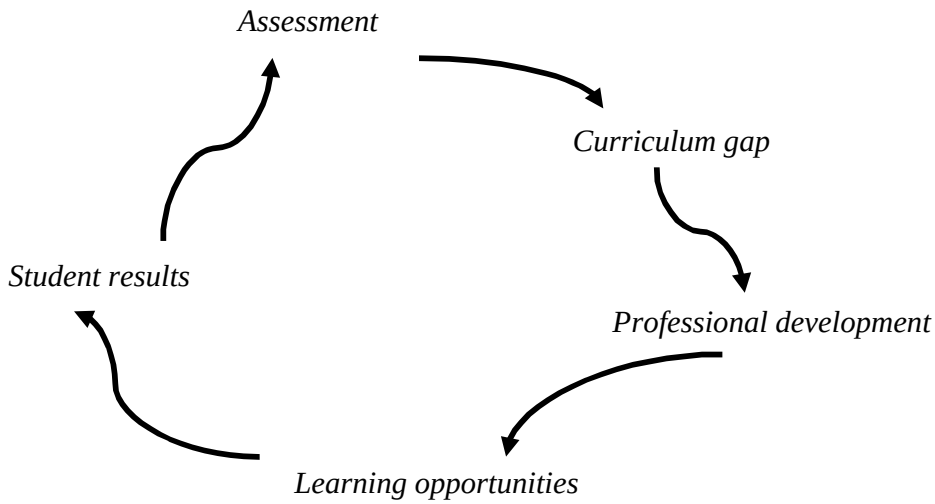


Figure 1. The five sector model of Driver 1

inflows that control the current states at the next level below, and signals running up the system feed the outflows of the current states of the level above it, forming balancing loops or limit cycles.

A balancing loop is a structure that constrains variables to operate as loosely coupled processes counterbalancing each other within a range. It is a complementary structure to a self-reinforcing loop, which behaves like a ball rolling down a hill, gaining speed with each passing moment. If the impact of a large-scale assessment were on the whole, self-reinforcing, then giving a single test would send the educational system into either a tailspin of ever worsening results or spinning off to ever-higher levels of performance. It seems much more likely that limit processes are at work within the system, which is why we've chosen the balancing loop structure as a generic pattern underlying the model.

Assessment in the model

Assessment is assumed to have an impact on curriculum in one or three ways: 1. by either "raising the bar of performance" for example, by introducing new concepts, 2. by having essentially no impact or a neutral effect, or 3. by "lowering the bar of performance," for example by underestimating, depressing, or missing entirely the performance of students. In addition, the impact of an assessment is assumed to have some relationship to the perceived value of the assessment by those administering and taking the assessment. For example, if people believe the assessment is a waste of time, then its administration may be poor or invalid and student performance may be low. The raising or lowering of the bar is represented in the model by the independent variable "content validity," and the perceived value is represented by the variable "face validity."

Both variables combine to form a measure of assessment productivity, which in this discussion will stand for a "new" statewide or large-scale assessment entering the system at some point in time. (Figure 2) The productivity of assessment is a "moving average" of the content and face validity, meaning that their current values are added and then divided by two for the current point in time. The choice of mathematics at this point is justified because construct validity is an internal measure while face validity is an external measure of the assessment. They operate independently as terms, rather than as coefficients. Furthermore, the choice allows either of the terms to become neutral (zero) without completely destroying productivity.

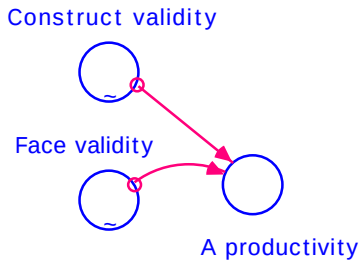


Figure 2. Assessment Productivity

The assessment subsystem of the model also presents national standards combining forces with the statewide or large-scale assessment to make a combined impact within the educational system. (Figure 3) The two then introduce either a positive, neutral or negative change (labeled "A downstream") on an impact of assessment already present in the educational system.

In this paper, the national standards will be kept at a constant between -1 and 1 over the period of time in each simulation. As the value approaches - 1, it will represent the strongest possible negative impact on a state's assessment system. Such might be the case if the national standards in a subject area ran counter to the state's directions in content and best practices for classrooms. In a similar manner, as the national standards approach zero, it would represent an increasingly insignificant role for the standards in the conduct of a state's assessment system. Finally, as the national standards approach 1, it will represent a "leading edge" of reform, where national standards are pointing the way in the "right" direction for curriculum.

When the signal from national standards and that from assessment productivity meet, a running average is formed. The choice here reflects similar thinking to that invoked for "A productivity." We have labeled that the "A downstream" influence of this sector within the larger hierarchy indicated in Figure 1. An “upstream” influence on Assessment Impact, from local instances of curriculum having responded to the impact of the assessment, is included in the model. The source of those upstream influences are visible in Figure 5.

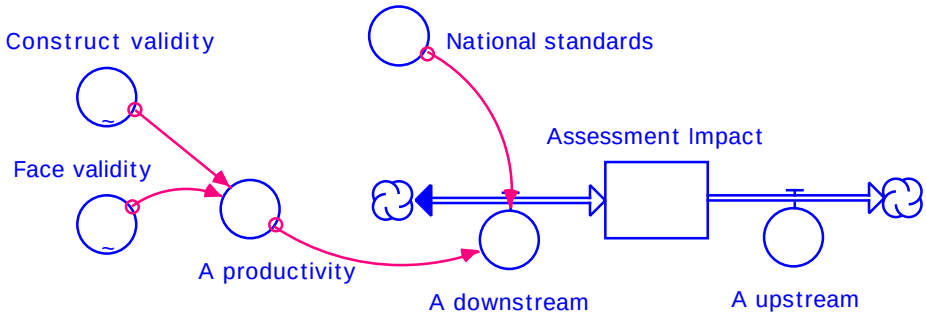


Figure 3. Assessment subsystem

This completes the description of the assessment portion of the model. Now, we'll describe the curriculum portion, then we'll begin to experiment with initial and boundary conditions and issues of change over time in the behavior of the two linked subsystems.

Curriculum Gap

The curriculum subsystem is built from two primary variables; the extent of "standards-based content" and "best practices"" in the curriculum. (Figure 4) Taken together, these two measures are an adequate representation of a productive relationship ("C productivity") that combines with the impact of the assessment from the level above to

form a gap between the assessment and the curriculum. As before, the term "C increase" has been expanded to mean, "increase, neutral or decrease."

As SB Content and Best Practices approach 1, they are aligned with the impact of assessment. As they approach zero, they are insignificant in the dynamics, and as they approach -1, they are misaligned with the assessment. It is instructive to stop for a moment at this point to examine how these sets of relationships work in the dynamical system. If the system were static the alignment of content and best practices with the impact of assessment would seemingly lead to an ever-widening curriculum gap - just the opposite of what we would expect in the real system! In addition, if content and best practice were perfectly "misaligned" with assessment, then no curriculum gap would be created![1]

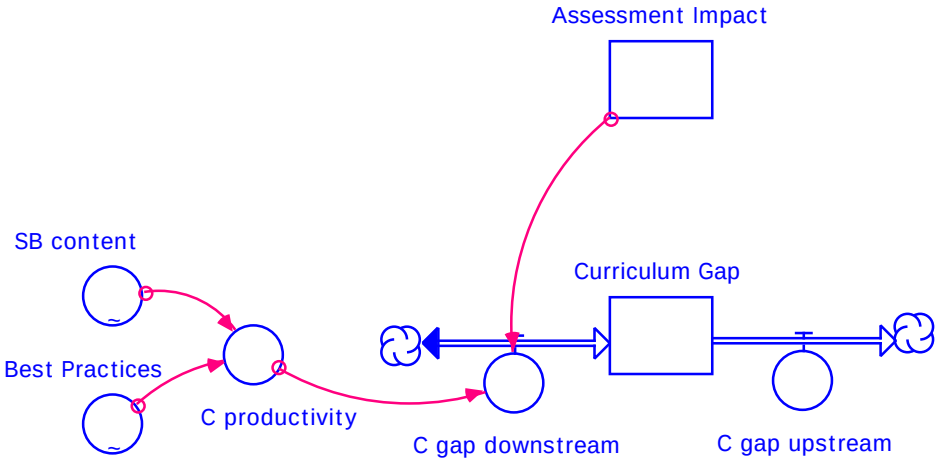


Figure 4. Curriculum gap

These results, which would be easy to find with linear methods, fail to take into account several dynamic features, the most important of which is the passage of time between giving an assessment and its having an impact on curriculum. When a time lag is introduced into the model, the odd results cannot ever occur and instead of "adding up" to an odd result, the two terms act as attractors for each other. As one term increases, the other term follows in the same direction. This is indeed what we expect to happen in the real system. When an assessment includes items that test a new concept like problem-solving in mathematics, at first there is a gap in the curriculum, and then in certain circumstances, the curriculum "catches up" with the assessment over time.

In addition to the introduction of time, a dynamic view of the system includes feedback, which in this model, balances or slows down the impact by "pushing back" on the level that created the impact. Staying with the example from the previous paragraph, once the impact of the assessment has helped cause the curriculum to respond, its shock value on the old curriculum is spent. Once the curriculum "catches up" to the implications of the impact of the assessment (i.e. by teaching student how to problem-solve in mathematics), the impact of that new item on the assessment falls.

The complete model

At this point, we can define the remaining connections that link the assessment and curriculum subsystems together into a complete model capable of simulation under various experimental conditions. (Figure 5) As the curriculum gap changes (in any direction - positive, negative or neutral) it sends a positively correlated signal upward in the hierarchy to "A decrease" causing an outflow to alter the impact of the assessment by lowering, raising or remaining neutral, (respectively).

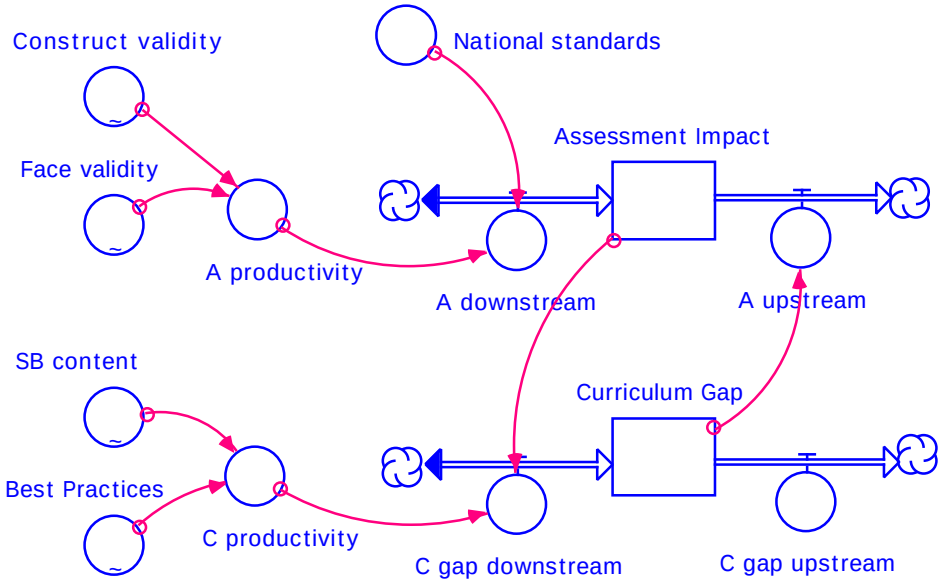


Figure 5. The complete model of impact of assessment on curriculum

Simulation experiments

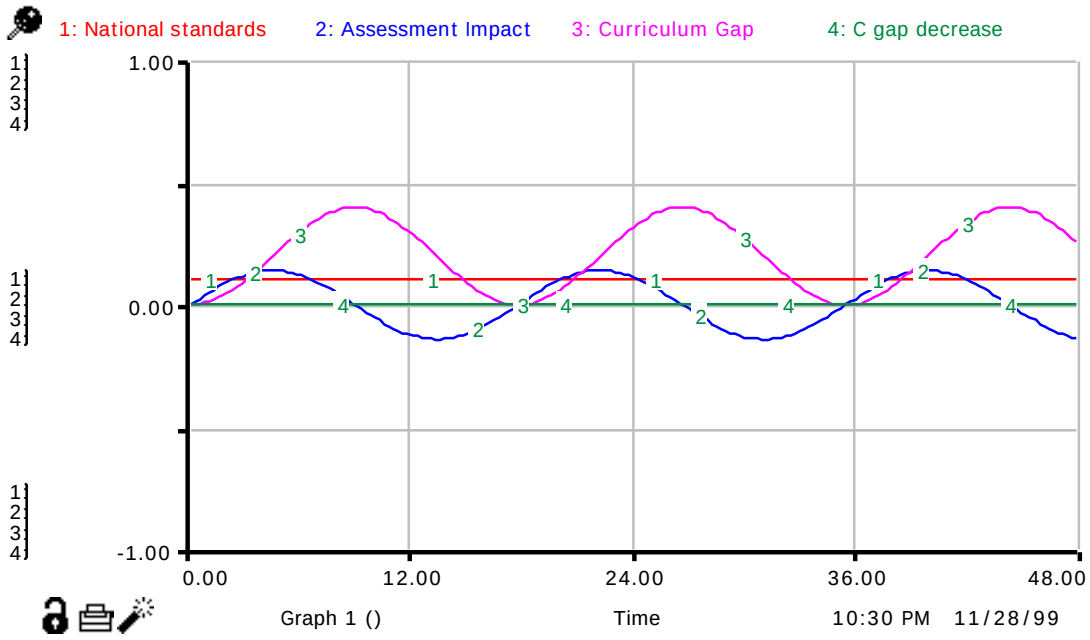
The resting state of this system is a dynamic equilibrium, an oscillating set of relationships. This means that assessment and curriculum are linked in a dynamic balance with one another. Whatever happens in assessment will have an impact on curriculum and whatever happens in curriculum will have an impact on assessment. A negative influence from assessment acts as a ceiling on curriculum; a positive influence acts like a pump, pushing curriculum higher. But at nearly the same time, curriculum pushes back. As curriculum peaks, assessment impacts drop. To learn more about how the systems interact, we can submit the system to a variety of initial conditions and hypothetical histories of the variables. To begin, let's review the independent and computed variables by looking at Table 1.

As might be expected, with all independent variables set to zero, the system produces zeros for all computed variables. But with the introduction of any quantity into any independent variable (with a range from -1 to 1 to set the parameters), the model begins to oscillate. For example, let's begin by setting one of the independent variables to ".10." We might think of this setting as 10% of the positive potential allowed for the variable.

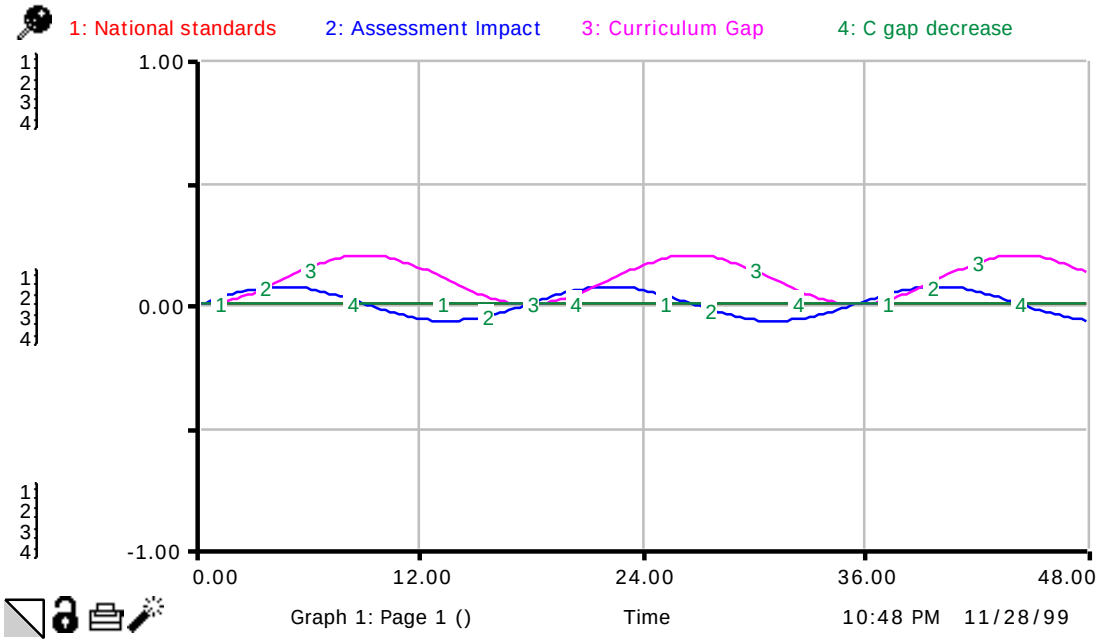
Table 1. Variables in the model

Independent Variables	**Computed Process Variables**	**Computed State Variables**
National Standards	A productivity	Assessment Impact
Construct Validity	A increase	
Face Validity	A decrease	
SB Content	C productivity	Curriculum Gap
Best Practices	C gap increase	
C gap decrease		

When national standards are set at .10, leaving all other variables set to zero, we see the following graph as the output of a simulation over 48 time periods. (Graph 1) This illustrates the "push" on curriculum from the positive impact of national standards. Curriculum is at worst, near zero, and it averages in the positive.

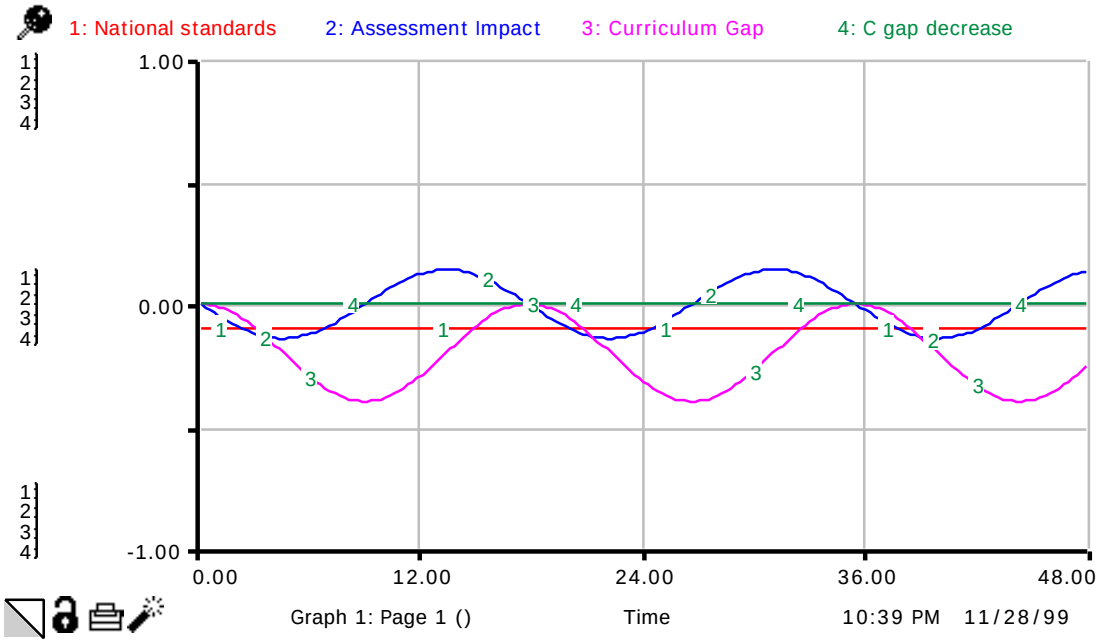


Returning national standards to zero and moving to either Construct of Face Validity set at the same .10, the following is produced. (Graph 2) Note the same general shape and wave period, but the reduced height or amplitude, which is caused by those variables being averaged before meeting the national standards at "A increase."



Graph 2.

The negative impact of national standards, with a setting of -.10, produces the following result. (Graph 3) Note the "ceiling" effect of the negative impact on the curriculum. Even though the dynamics are still working, that is, the impact of assessment still drives curriculum, now the maximum height of curriculum is near zero, or in other words, curriculum is at best, neutral but on average stays negative. A similar negative graph is produced by either Construct or Face Validity set to -.10 with lowered amplitude as in Graph 2.



Preliminary Conclusions

As the research team moves toward a more complete model, this early structural foundation will be refined by historical data from two large-scale sites (Vermont and Puerto Rico) and by expert teams who will contribute to the model based on their understanding and experience in large-scale reform. The model at this stage is capable of complex and surprising behavior within the basic parameters defined in this technical brief, and in general behaves as we expect large-scale systems to behave. The research team will need to operationally define metrics for each of the independent variables.

---

[1] The system would be (or become) static if there were no feedback loops, no forward movement of time, or there was a point attractor. If SB content = 1; Best Practices = 1 and impact of assessment = 1, then the increase of curriculum gap would be 1. If SB content = -1; Best Practices = -1 and impact of assessment = 1, then the increase of curriculum gap would be 0. The table below gives a picture of the relationships.

SB Content	Best Practices	Impact of Assessment	Result on Increase of Gap
-1	-1	-1	-1
-1	0	-1	-0.75
-1	-1	0	-0.5
-1	1	-1	-0.5
0	0	-1	-0.5
-1	0	0	-0.25
0	1	-1	-0.25
-1	-1	1	0
-1	1	0	0
0	0	0	0
1	1	-1	0
-1	0	1	0.25
0	1	0	0.25
-1	1	1	0.5
0	0	1	0.5
1	1	0	0.5
0	1	1	0.75
1	1	1	1