



Measuring the Impact of State-Level College-Cost Initiatives

By Roland J. Cole

December 2006

1. Introduction

This report is one deliverable of a year-long project conducted by the Sagamore Institute for Policy Research (SIPR) for the Lumina Foundation for Education (Lumina). In this part of the project, Lumina asked SIPR to develop a scheme Lumina could use to monitor its ongoing effort to support state-level initiatives that are designed to serve the goals that Lumina is trying to serve.

Lumina and SIPR both hope that the scheme will provide:

- (i) ways to identify and categorize “good ideas” that might be initiatives to try,
- (ii) ways to select states in which to try initiatives, and
- (iii) ways to measure “how well” the initiatives did in dealing with the barrier of “costs”.

These steps are, of course, means to solutions, not solutions themselves, but they should be helpful to Lumina’s ongoing efforts in this area.

In a classic example of a “bell the cat” solution, the scheme itself is very simple to state, but can be very difficult to apply. The scheme SIPR developed has five steps:

1. Categorize the State (by capacity, opportunity, need, interest)
2. Categorize the Initiative (by goals, strategies, players, clients)
3. Identify and Measure the Inputs
4. Identify and Measure the Outputs
5. Estimate the Impact on Outcomes

The rest of this report sets out the background for those steps, explains some of the terminology, and provides example applications of some of the steps.

2. The Basic Problem

Lumina has stated the basic problem in several ways, but for the purpose of this report, SIPR has used the following statement (SIPR’s words, not Lumina’s):

Almost all residents of the United States need at least two years worth of post-secondary education, and many will need four or more years worth of such education, in order to be equipped to make the social and economic contributions to their communities and to the nation that are needed to keep the United States a place where social and economic conditions keep improving for most of its

residents. Unfortunately, the current system does not deliver such education to enough U.S. residents. It delivers less to students of lower income families, minority students, and students who have already assumed adult responsibilities. One of the barriers to delivering more is the cost of such education, meaning both the total social cost and the cost normally expected to be paid by the student or his or her family.

The following figure shows this problem in graphical form. The average cost per year, across all four-year, degree-granting institutions, is approximately one-third of the median income of a white family in the U.S., and one-half of the median income of a non-white family.

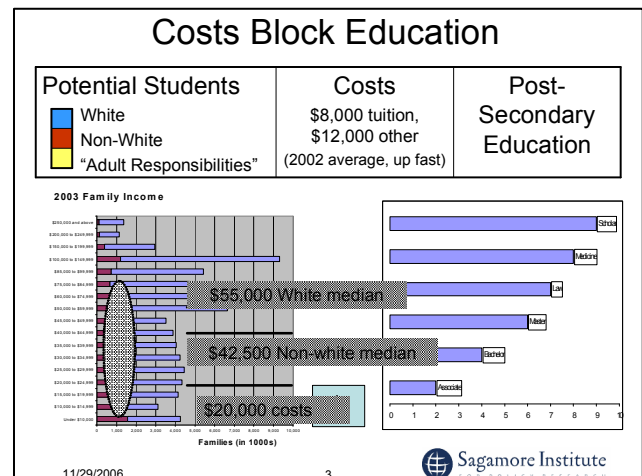


Figure 1. Student Costs Block Education

Note that “college costs” in Figure 1 are only the “student costs” – that is, the out-of-pocket costs that the student or his or her family are normally expected to pay, absent some form of scholarship. They do not include the portion of the social cost that the student normally does not pay. Nor do they include the opportunity costs the student (and his or her family) pays by being out of the full-time workforce for at least nine months of each school year. The costs quoted are a national average across public and private four-year institutions – tuition is much higher for private schools and out-of-state students attending public schools. Other costs are much higher in high-cost (usually urban or suburban) areas. Of course, both tuition and other costs can also be much lower when attending an institution without leaving home or one in a cheaper location or other such circumstances.

“College” in Lumina’s terminology is a short-hand term for all forms of post-secondary education – vocational school,

academic two and four year programs, graduate programs, resident and non-resident, full and part-time. However, as this report discusses below in more detail, the model most often assumed in discussions of “college costs” is “rite of passage” education for 18 to 29 year olds versus skills development and certification for professionals at two-year levels or more than four-year levels. The discussions assume that the “gold standard” is a program of around four years, where the students attend school full time at least nine months of the year, live on or near campus, participate in a variety of extracurricular activities as well as attend class, and are connected to an institution that conducts research and community service with the same people who serve as faculty for the students. The report discusses below the question of whether anything else is “good enough,” let alone equivalent or better for some students or programs.

Lumina lists three target groups of students for its activities – students from lower-income families, minority students, and students who have assumed adult responsibilities (“adult students.”) It has recently launched a project with SIPR that adds a fourth target group – immigrant students, which include both children of immigrant parents and students who themselves are immigrants. These groups do show substantial overlap – minority families are often low-income families, immigrant families are often low-income and minority, and adult students are often low-income or minority or immigrant. However, as discussed below, this overlap does not mean that a policy targeted at one group will also work well with the others, even within the overlap – policies for lower income students may be necessary, but not sufficient when the student is also minority or “adult.”

Note also that student costs are increasing faster than ability to pay, as shown in Figure 2.

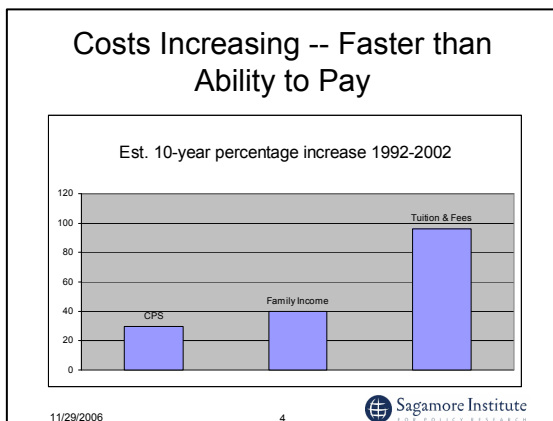


Figure 2. Student Costs Increasing

Costs shown are for 4-year institutions, but the pattern is consistent across all types of post-secondary education providers. Also, the value of educational completion versus non-completion is increasing, so time-cost of delaying completion is increasing as well. Various estimates of the value to the student of receiving a four-year degree suggest that the rate of return to the student who pays full tuition and other costs is about nine percent (9%), assuming the student receives the degree at age 21 or 22 or so, and has 40 plus years of working life after college to receive differential income.

Note also that many educational providers are being asked to play increasing other roles – research, help with technology transfer, entertainment, and other forms of economic and social development -- so that even when the “marginal” costs per student may not be increasing, the average costs are, because the institutions are performing these other roles

One result of rising student costs versus ability to pay is, as suggested above a systematic underinvestment in education from society’s point of view. The Lumina understanding is that we would all be better off if more U.S. residents got two or four or more years of post-secondary education than they do now. One confirmation of this understanding is that many other countries now have a higher percentage of their residents with four-year degrees than the U.S. does. Lumina has reported some very interesting figures about this comparison, as shown in Figure 3 below (from College Board, *Education Pays: Second Upgrade*, found at www.luminafoundation.org)

International Comparisons

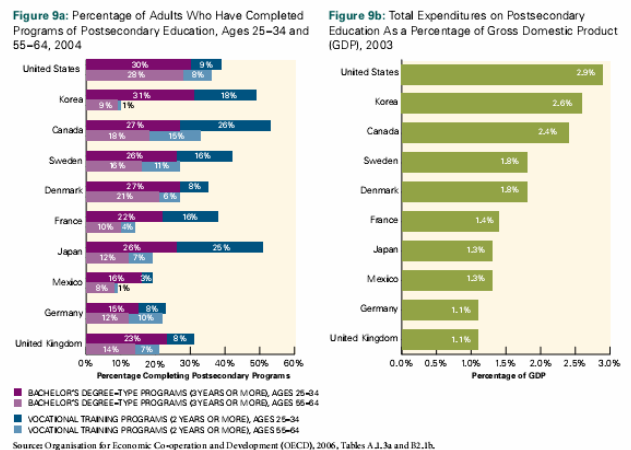


Figure 3. International Comparisons

One caveat about the international comparison should be kept in mind. Actually raising the percentage of those obtaining a four-year degree during the past 40 years or so is no mean accomplishment, given all the stresses of drugs, diversity, and economic ups and downs. The U.S. clearly wants to do better, to catch up to Korea, Japan, and Canada, and stay ahead of other countries that are much larger, such as China and India. But this record is no cause for heaping abuse on the present system. Unfortunately, the fact that it is not a complete failure actually makes change harder than if it had failed completely, as the section below on incentives discusses.

A second result of the high and increasing student costs is that education is regressively distributed – that is, more of it goes to higher income and majority students than to lower income and minority students. Figure 4 shows the numbers from Lumina.

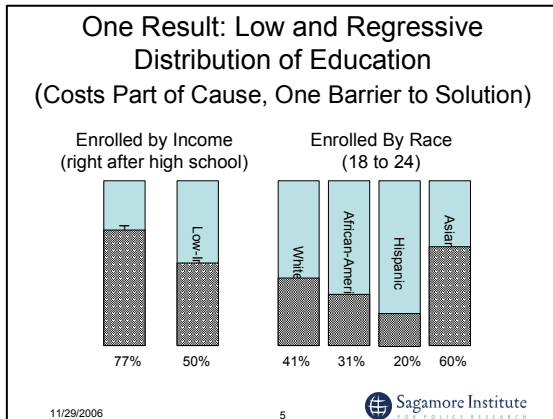


Figure 4. Regressive Distribution of Education

3. Tools for Categorizing College Cost Initiatives

This section discusses how different aspects of the problem outlined above lead to different ways to categorize college cost initiatives – whether state-level initiatives or other types.

3.1. Different Costs by Payer

This report has already distinguished between “student costs” and total costs to society. One can actually distinguish at least five different payers, as shown in Figure 5.

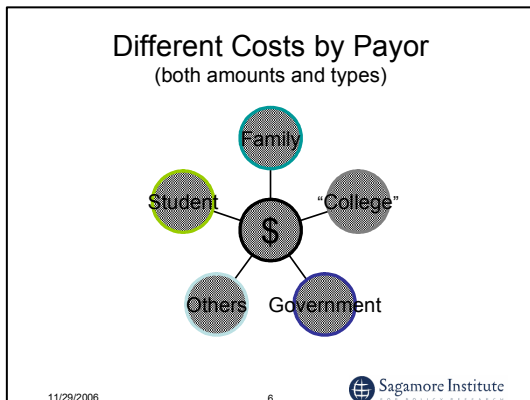


Figure 5. Different Costs by Payer

The “student costs” mentioned above are usually paid by the student’s family or by the student. One function of loans, as opposed to scholarships, is to shift the amount of the loan, not only in time, but usually from the student’s family to the student (and sometimes to the student’s spouse).

The institution providing the education (the “college” in shorthand), also pays a portion of the costs, since tuition is almost never equal to the full marginal cost to the college of each student. Indeed, to the extent it uses its funds to provide scholarships or below-market loans or above-market jobs, it is also paying a portion of the student costs as well.

Government often pays a per student fee to make sure in-state students pay less than out-of-state students, and often pays students costs on behalf of students via scholarships and below-market loans. Government also makes capital grants to colleges and research grants. The research grants often pay a

portion of faculty salaries and contain money for hiring current students to help with the research. Government may also be involved in dealing with the costs other than tuition through housing policies, discounts for current students, tax-free treatment of educational property, and other policies supportive of education in general.

Non-governmental entities, including alumni, industry, foundations, and others also make grants in support of educational institutions, including targeted support for buildings, research, and faculty as well as help in defraying student costs.

So two ways of categorizing a college cost initiative is by either the type of cost it targets or by the payer it targets.

3.2. Different Entities

An important insight about the problem of educational shortage is that the issue involves a number of entities. Figure 6 suggests a grouping of such entities.

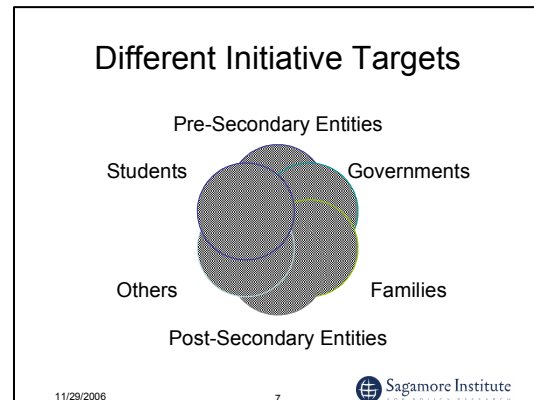


Figure 6. Different Entities That Could Be Targets

These entities include the students, or selected subsets of them, such as students from lower-income families, minority students, immigrant students, or adult students. They also include the families of students (often siblings, grandparents, aunts and uncles as well as parents). They include the post-secondary institutions providing the education. They include the pre-secondary institutions getting the students ready for post-secondary education, and often playing substantial roles in promoting student awareness and preparation. They include government at various levels, and various agencies with each level. They include a whole raft of others. Not just the potential funders listed above, but also suppliers of textbooks, housing, meals, and other supplies and equipment used by students and educational institutions.

Thus, another way one can categorize an initiative is by which one or more entities it targets.

3.3. Different Goals

Initiatives designed to deal with the problem of college costs can actually range across a number of sub-goals. A grouping of these potential sub-goals appears in Figure 7.

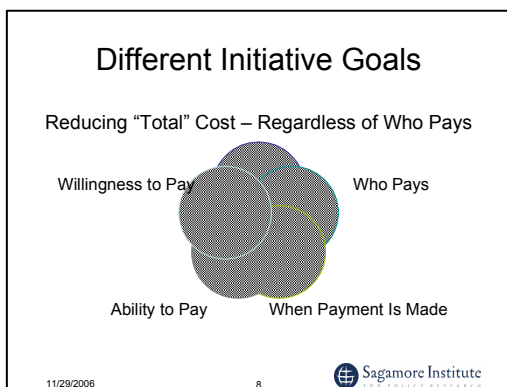


Figure 7. Different Initiative Goals

One major distinction is between those initiatives designed to reduce total cost and those designed to reduce student cost. In each case, the initiative may be targeted at a particular subset of those costs, such as textbooks.

Beyond that distinction, however, are the distinctions among (a) trying to reduce the costs (from the point of view of the person paying them); (b) trying to shift a set of costs from one payer to another (often from the student to the government); (c) trying to shift the timing of payment; (d) trying to increase the willingness to pay; and (e) trying to increase the ability to pay.

A single initiative may combine a number of these goals – for instance, a project to increase student awareness of and application for potential scholarships or loans may (i) help shift some student costs to the scholarship or loan grantor and (ii) increase the student’s willingness to pay the remaining portion. A project to increase awareness and use of tax-favored savings plans may increase both willingness and ability of a student and his or her family to pay student costs.

Lumina also sets out four elements of its efforts – (i) to increase awareness of the value of education to the student and to society; (ii) to increase preparation of potential students to receive post-secondary education; (iii) to deal with financial issues; and (iv) to increased institutional response to the problem. One can use these four elements to categorize initiatives as well.

Lumina has also provided another way to categorize its goals, and thus another way to consider for categorizing state-level initiatives. It involves three buckets, as set forth in Figure 8.

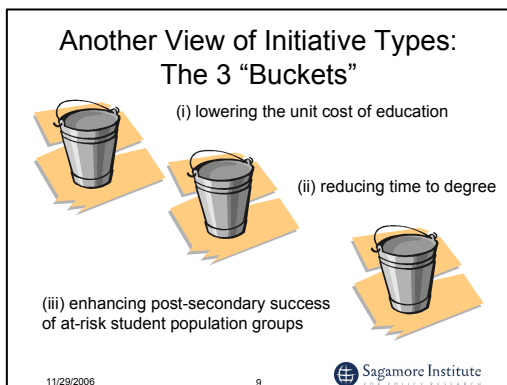


Figure 8. Three Buckets of Goals

Lowering the unit cost of education is the Lumina phrase for total social cost. The use of the term “unit cost” reflects that the overall goal does include providing more education – both to more students and more to each student. Reducing time to degree is a key technique for reducing both student cost and total cost. Enhancing post-secondary success is also key, because the US is close to equal on the percentage of the 18 to 24 age group entering post-secondary education, but is far behind other countries in the percentage that complete the program they start. As mentioned above, the key at-risk student population groups are lower-income, minority, adult, and immigrant.

4. Generic Policy Measurement Issues

Anyone seeking to measure a policy initiative encounters a number of generic issues, and state-level initiatives to deal with college costs are no exception. This section provides a brief reminder of the generic issues of particular concern in this area.

4.1. Inputs versus Outputs versus Outcomes

Most policy initiatives are undertaken, in large part, to bring about a change from one state of the target “world” (whether international, national, state, or local) to another. Policy analysts call this change in world-state an outcome of the policy. The outcome in this area is an increase in the number of U.S. residents equipped by post-secondary education to make social and economic contributions to themselves and to the communities of which they are a part, especially those potential students from middle and lower-income families, those from minority families, and those who have already assumed adult responsibilities.

Outcomes are notoriously difficult to measure directly. Often the appropriate data on the world-state is simply unavailable, whether prior to or after the operation of the initiative. Even when such data is available, sorting out the effect of the initiative from all the other activities in the world at that time is difficult. Also, initiatives generally do not pursue outcomes directly. Instead, they pursue intermediate steps that are deemed to be directly related to the desired outcomes. Policy analysts call these intermediate steps outputs.

Of course, initiatives also require resources to carry out. Money is one such resource, of course, but staff time, legal authority, and technology are all examples of other resources that a given initiative may require. Policy analysts call these required resources inputs.

The generic flow is a circle from inputs to outputs to outcomes, which then leads to new initiatives that keep the circle flowing. Figure 9 shows the three-step circle.

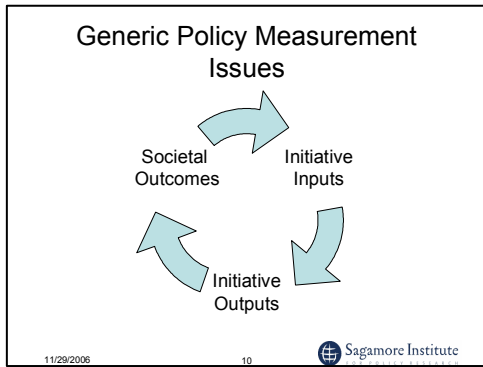


Figure 9. The Input/Output/Outcome Circle

The circle could have many more than three steps. Indeed, the Lumina four-step process itself suggests four types of outputs, each of which may be an input to a later output. *Awareness* of the value of and potential for post-secondary education leads to *preparation* by the student, his or her family and others, for dealing with *financial issues* and other matters, and these prepared students lead to *institutional responses* that hopefully provide more education than would have been provided otherwise.

4.2. Generic Cost Measurement Issues

Measurements of cost, we need to keep reminding ourselves, are a measure of specified resource or resources (usually dollars, but not always) to produce a given output, which in a policy context, we know (or hope or assume) is either a proxy for or a way station towards the outcome we hope to achieve.

To illustrate, some measure of class size might be compared to graduation rate, and indeed, according to most analyzes, having bachelor's degree institutions with smaller average class sizes usually have higher graduation rates than institutions with larger average class sizes. But...

(a) "Average" is a measure of central tendency that discloses very little information. A school that had huge lecture classes and small tutorials could have the same "average class size" as a school that had almost all moderate sized classes.

(b) Moreover, we know that certain incoming students are far more likely to graduate than others. So a school that fills itself with students highly likely to graduate is almost certain to have a higher graduation rate than a school that fills itself with students far less likely to graduate, regardless of the class size. Some have suggested that perhaps we should use the positive or negative deviation from "incoming expected graduation rate" to determine the impact of the input in question, as a way to account for the deviation in a related input, the nature of the students prior to entry.

The generic problem is one of making sure that the comparisons, such as between a state with the initiative and one without, are apples to apples comparisons. Figure 10 diagrams this issue.

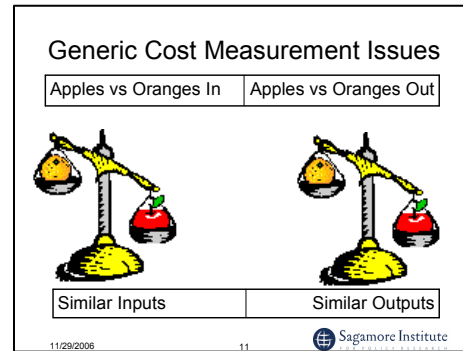


Figure 10. Apples to Apples Comparisons

This research has concluded that one of the two biggest steps forward in this area would be achieving greater and more fine-grained consensus on the outputs that should be involved. (The other step is to reduce time to completion, as this report discusses elsewhere.) One could hope for greater and more fine-grained consensus on the outcomes, but maybe that can follow consensus on outputs.

The report has already mentioned the "gold standard" of resident students at a research university that conducts a full panoply of research, community outreach, and extracurricular activities. One can certainly spend a lifetime working on reducing the total cost and the student cost of that package. But anyone who wants to explore alternatives to that package could really use good measures of educational outputs that could compare the alternatives to the gold standard. Indeed, even if working within the gold standard itself, it would be nice to compare outputs from one variant with those of another.

We do have some measures now. Graduation rates, graduate record exams (GRE), and some emerging tests (discussed below) purport to measure overall output. Advanced placement exams (AP) purport to measure first-level output in some 22 subject areas. Various more specialized readiness tests, such as the law school exam, the medical school exam, and the business school exam, purport to measure readiness for further education in those areas. Some effort is underway to do something similar in the engineering field. But greater and more fine-grained consensus along these lines would certainly help compare various approaches, both within the gold standard form and outside of it.

4.3. Generic Policy Distribution Issues

This policy area is concerned with groups of students, and with a variety of types of post-secondary education. Any policy area dealing with groups and with types of programs faces generic distribution issues, as illustrated by Figure 11.

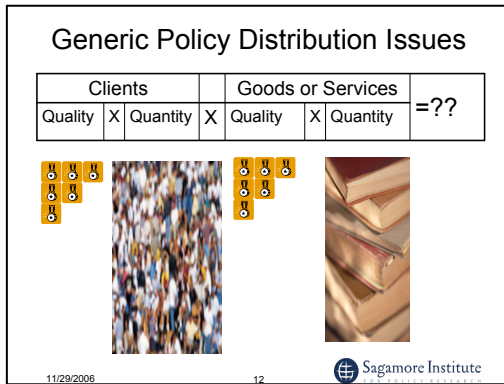


Figure 11. Generic Policy Distribution Issues

The potential initiatives usually involve groups of potential students or families – e.g., all those from families with less than the median income. They may also involve groups of entities of various types – pre-secondary, post-secondary, government, other. Does the initiative try to focus on the best, middle, or worst (however measured) in the group, and how many does it try to serve?

These initiatives may involve a range of forms of post-secondary education, currently delivered in “normal” increments of 2 to 9 years, with subsets of each (e.g., “introductory courses” or one medical specialty versus another). What quality and quantity of post-secondary education does the initiative try to promote?

Of course, some initiatives are very highly focused – e.g., a scholarship for a named student to attend a named institution, and others are not – e.g., a grant of money usable at any post-secondary institution where the recipient is chosen by lottery.

4.4. Generic Policy Implementation Issues

As mentioned above, perhaps the single greatest barrier to increasing the percentage of students getting post-secondary education is not that the system has stopped producing graduates. Rather, the system has fallen behind some countries in the percentage educated and behind others in the total number educated. To use a rough analogy, the car still runs; it just does not go as far or as fast as we think we need it to go, and other newer cars do exist that appear to go faster and farther. Moreover, short of failure so obvious that a school is shut down, the participants in the system other than the students and their families appear to have few incentives to lower either total costs or student costs. Students on need-based scholarship do not even have an incentive to wish student costs were lower, since they are not paying them.

Figure 12 diagrams the generic policy implementation issues that face most policy initiatives, including state-level college cost initiatives.

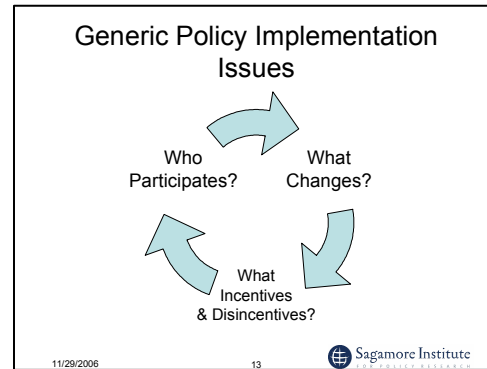


Figure 12. Generic Policy Implementation Issues

To understand almost any policy initiative, the policy analyst must answer the following questions.

4.4.1. Who Participates?

Initiative sponsors usually have a clear target in mind, but often fail to identify the other entities, groups, and individuals that must participate in one form or another in order for the initiative to be successful. One example is labor union leadership when dealing with a unionized workforce, even when the group is not called a union per se, such as a faculty senate.

4.4.2. What Changes Must the Participants Make?

Sometimes a little change is harder than a “clean break from the past;” sometimes a change that seems easy from the outside is hard from the inside and vice versa.

4.4.3. What Incentives Do the Participants Have, Whether Provided By the Initiative Or Not, To Make the Change Required?

For instance, it is not standard practice for most providers of education to reward teachers for teaching more students per class; indeed, they may simply have to grade more papers and tests than if the class were smaller. Textbooks are a counter example – since book publishers pay royalties, almost always in whole or in part to the textbook author, the more students that use the same textbook, the more the author is rewarded.

Note that rewards need not be directly economic; increased prestige, affection, more flexible scheduling, et al. can all be powerful incentives and reductions in them can be powerful disincentives

5. The Special Problem of Educational Outputs

This report has already mentioned the special role played by having to measure educational outputs that we hope are related to educational outcomes – that is, measures that help determine whether the students have the appropriate preparation to make social and economic contributions to themselves and their communities.

Unfortunately, we all know that a degree, at any level is not quite sufficient. Figure 13 illustrates this situation.

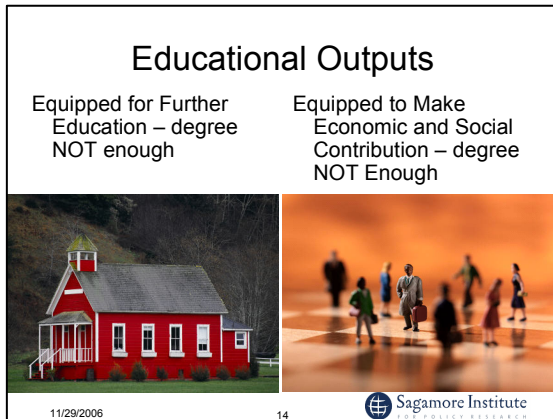


Figure 13. A Degree Is Not Enough Measurement

Note that even as a measure of fitness for further education the degree is not enough. The ACT and the SAT supplement the high school degree and the GRE, and similar tests for Law, Business, and Medicine supplement the college degree. Some states have a supplemental test for high-school graduates that is designed to measure “life skills” that high school should have taught, although some analysis suggests that the skills measured by these tests are similar to the skills measured by the ACT, albeit with a slightly different, less academic vocabulary.

There is some consensus or satisfaction with these “comprehensive measures” of being equipped for further education, even on national basis (although all of the tests have their critics). There is much less consensus/satisfactions with “comprehensive measures” of being equipped for economic and social contribution, even on state by state basis, although some examples, like state-by-state law exams, have been in operation for decades.

Similarly, there is some consensus or satisfaction concerning particular schools in particular areas (such as MIT or Caltech for Engineering), but these tend to be very selective, very high cost-per-student schools, in part because students are full time, resident, with related extracurricular activities and teachers and other members of the institutions are highly involved and recognized in research and publication and other related to but non-teaching activities.

Lumina has sponsored and reported work on developing more and better measures (see the Carey article listed at the end), such as the Collegiate Learning Assessment for broad readiness, various industry-related tests for depth readiness – ABET (Engineering), TEAC (Teaching), and the state-by-state exams for law, medicine, psychology, accounting. At the lower end, individual states are developing “employee readiness tests,” largely targeted at the high-school graduate level.

At the subject level, as mentioned above, the College Board offers 37 tests in 22 subjects that are considered satisfactory measures of college-level achievement by four-year institutions all across the country, including the most elite. Some of these same schools will also use College Board Achievement tests in some subjects, such as languages, to perform the same function. In other words, a student who obtained score of N or above on a College Board Language Achievement Test, such as Spanish or German, would be

deemed to have met that school’s language requirement, regardless of how many or how few years of the language had been taken.

Having such measures, at the course level if not the comprehensive degree level, would go a long way to solving the apples and oranges measurement problem. Imagine a post-secondary institution that rejected many of the major premises of the gold standard. That is, where the gold standard offers four big elements:

- (a) 4 year right of passage for 18 to 24 year-olds (includes much more than academics)
- (b) lecture, readings, two exams and/or papers format (sometimes one exam and/or paper)
- (c) subject-by-subject approach
- (d) every student on his or her own

A given alternative might offer:

- (a) outcome test, independent of input and timing and “non-academic activities”
- (b) practice, feedback, active engagement
- (c) multidisciplinary “problems”
- (d) collaboration/leadership/followership

How would one be able to compare graduates of one program versus the other? If the goal is to get the students ready for law or medical school, maybe the existing law and medical readiness tests would suffice. If for graduate education, maybe the GRE would suffice. But what if the goal is men and women equipped to make social contributions? Would a CLA-type test be acceptable? Would it be helpful to have the college level equivalent of the GED, for those who attended alternative programs or studied on their own?

What if the comparison were at the course or subject level? Would a College Board Achievement test suffice? The purely professional schools do offer some alternative measures of how well they do by their students, as suggested by Figure 14.



Figure 14. Measures of Quality Offered by Purely Professional Schools

Two additional factors compound the problem of measuring educational outputs. One is that we want our graduates to have combinations of broad and narrow skills. To this end, colleges usually require that the student pick a major for the narrow skills and satisfy distribution requirements for the

broad skills. Some call the goal to produce “T-shaped” individuals, as illustrated in Figure 15.

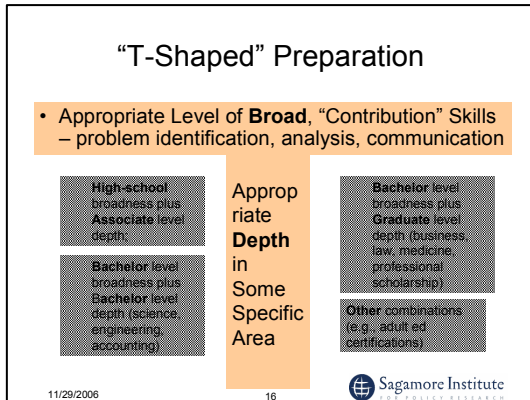


Figure 15. T-Shaped Preparation.

The desire for T-shaped preparation means that we really want T-shaped measures – one or more for the broad skills, as the CLA tries to measure – and one or more for the narrow skills, as a College Board Achievement Test tries to measure. Moreover, as suggested by Figure 16, we want varieties of T-shaped preparation.

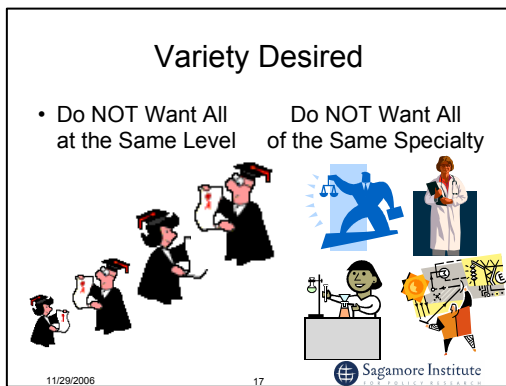


Figure 16. Variety of Preparation Desired

As any lawyer knows, we certainly do not want all lawyers. But we also do not want all engineers, or all scientists, or all doctors, or even all with advanced degrees. Some should probably stop at the associate level, some at the bachelor level, some at the master level, and some at the doctoral or professional level.

Some states are actually trying for “outcome” measures. The concept is that if one were to measure actual contributions over the graduate’s next 30 or 40 years, that would provide some indication of how well equipped they were at graduation. Figure 17 diagrams this concept.

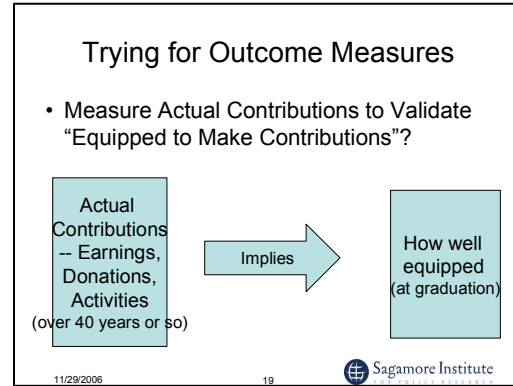


Figure 17. Trying for Outcome Measures

Some states (e.g. Florida) are putting in “universal/lifelong” student IDs so that they can better tie “actual contributions” (earnings, taxes paid, donations made, activities volunteered) to institutions attended and perhaps, with CLA or similar test, to measures of “educationally equipped.” Even the *US News and World Report* rating system includes some measure of average income and contributions from alumni as measures of economic and social contribution. Note that a lifelong student ID raises all sorts of privacy concerns. Although theoretically the individual could be protected by an anonymizing process, one would still have to trust that the process was consistently applied over a long period of time. Nonetheless, some states have opted to take that risk for the benefits the data can provide.

Of course, this requires waiting decades to see the long-run effect. It would be nice to have predictors other than just graduation from an elite institution. But perhaps once analysts can tie the record of actual contributions to institution attended and perhaps other measures of educational output, we will begin to get these predictors.

As mentioned above, we do have some efforts underway to develop more national measures of both educational inputs and outputs that would be useful in this policy area. Figure 18 diagrams some of these.

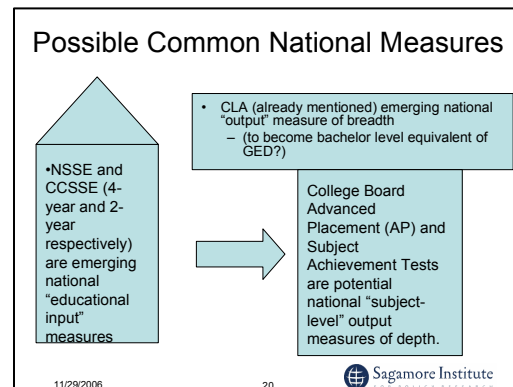


Figure 18. Possible Common National Measures

See the Carey article referenced at the end of this report for more discussion of the NSSE (the National Survey of Student Engagement), the CCSSE (the Community College Survey of Student Engagement) and the CLA (the College Learning Assessment). Also see www.cae.org/content/pro_collegiate.htm and

www.cic.edu/projects_services/coops/cla.asp. National measures are more helpful the more the alternative is “outside the box” of the traditional “gold standard” education, as suggested by Figure 19.

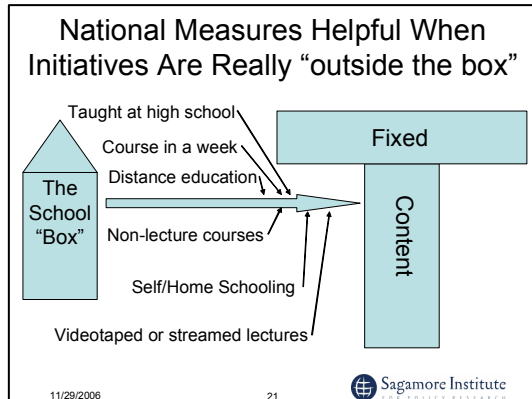


Figure 19. National Measures for Outside the Box Initiatives

For instance, the School of Public and Environmental Affairs at Indiana University helps meet the needs of some of its adult students by offering otherwise traditional courses in one calendar week of 8-hour classes. In that case, it can use the same test it uses for the same class in a more traditional format. But if another school wanted to try that approach, it would be especially reassuring if the test used were the same test the new school would use – a multi-school, if not national, test in that subject.

Lumina is also sponsoring work on “articulation agreements,” by which one school agrees to accept students and grades from another. Compare “articulation agreements” (receiving school accepts grades of sending school) versus “test agreements.” In some states, graduates of in-state law school do NOT have to take bar exam; everyone else does. Early college programs have “articulation agreements” such that the HS students have college credit from the signed up college; the Advanced Placement system grants college credit for scores on AP tests, regardless of the school/other inputs involved. Articulation agreements are certainly more traditional, but test systems may allow for more innovation – although if the receiving school “really trusted” the sending school, innovation would be possible in that system as well.

Even when an agreed upon test (or other measurement device) exists, the statistics question still remains, since the goal is to educate groups of students. Figure 20 shows some of the possibilities.

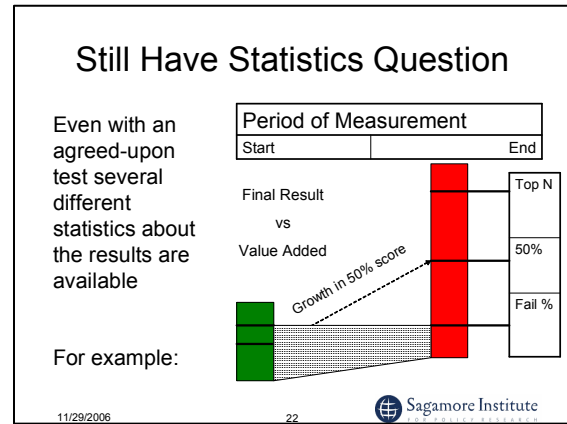


Figure 20. The Statistics Question

In some areas, the goal is to get everybody above failing, and the process can be measured on that statistic. In others, the goal is to add value, so some measure of before scores versus after scores can be calculated. In still others, one may want to award prizes to the highest score, or most improved score. Or maybe an entire class gets a prize, based on average improvement, total improvement, 10 highest scores, or some other measure.

Note that the quality control literature, however, suggests that offering prizes for one out of 20 does not improve the overall performance – it just makes one person happy and 19 unhappy. In many Army boot camp competitions, the prize goes to the team who has all its members finish first – in other words, it rewards the team with the highest minimum score.

6. “Best” Need Not Be the Enemy of the “Good”

Much policy analysis reporting begins with what some have called “the obligatory data whine.” That is, many analysts begin their reports with a complaint about how the data is missing or misleading. Even so, the profession as a whole is still wedded to the idea that data and analysis can help, so they proceed anyway, trying to do the best they can.

The situation is similar in this context. While, as the section above discusses, having generally accepted, fine-grained measures of educational output, even of partial outputs, would be extremely helpful, much work (and analysis) can be done even before those measures are available. Figure 21 suggests some of the ways policy initiatives have dealt with the lack of output measures.

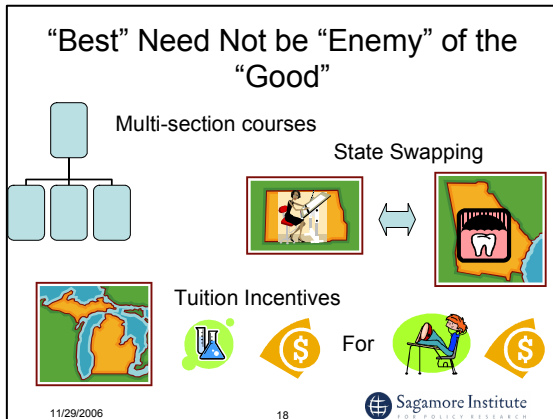


Figure 21. Dealing with Lack of Output Data

In the Twigg project (see the reference at the end of this report), the cost-reduction activities involved multi-section courses, so all the students took the same test. In that context, the test was the “accepted measure of educational output.”

In the Kansas state-swapping initiative, Kansas swapped slots so that Kansas residents paid in-state fees at Missouri dental schools; Missouri residents paid in-state fees at Kansas architectural schools; or the state paid down tuition for the Kansas student at a Nebraska school, so the student paid in-state fees. Since the goal was largely to reduce student costs, the education was held constant. To the extent it saved Kansas from the overhead of building its own equivalent schools, it was assuming the Missouri or Nebraska offerings were equivalent to what Kansas might offer.

Michigan designed its initiative to counteract the pressure community colleges felt to reach budgets by steadily increasing student costs. In this case, if the institution held down tuition increases, it got some additional state money. Here again, the educational output was deemed to be the same, no matter how much of the budget came from tuition.

The conclusion for a scheme to categorize and measure state-level initiatives is that much can be done before we have generally-accepted, fine-grained measures of educational output, let alone similar measures of outcomes.

7. States as Unit of Analysis

The scheme in this report is to help Lumina categorize and measure state-level initiatives, because Lumina has launched a major project, with large grants to three organizations, all focusing on the state level. The Foundation will focus on three strategies to achieve these objectives:

- Best practices. Document and publicize new approaches that improve the quality of teaching and learning and deliver education at a lower cost.
- State- and system-based reform efforts. Identify and support states or systems ready to change their ways of doing business to reduce costs, raise quality, and serve more students.
- Public will building. Increase public awareness of the issue and build the will to achieve the initiative’s objectives.

Three national organizations will help carry out the next phase of the initiative: Boston-based [Jobs for the Future](#), the [National Center for Public Policy and Higher Education](#) (San

Jose, CA), and the [National Center for Higher Education Management Systems](#) (Boulder, CO). All have a solid track record of achievement in higher education research, analysis, and innovative program development and implementation. In addition to those named above, other organizations will be added as specific areas of work are identified

Many might question the use of the state as a unit of analysis, along the lines suggested in Figure 22.

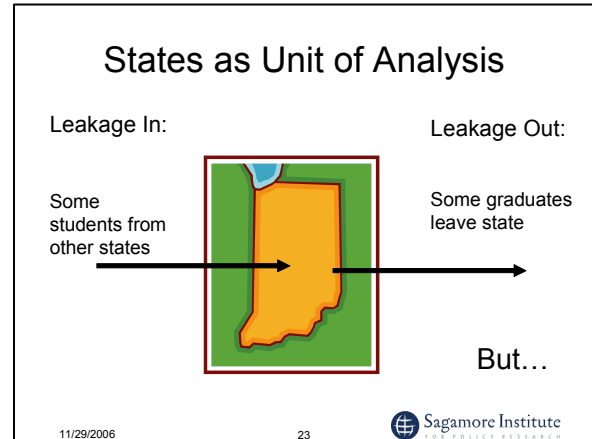


Figure 22. States as Unit of Analysis #1

The problems are generic to measuring activities in any geographic area – some “colleges” are local, some are statewide, but many are national or international. Similarly graduates may stay in area, in state, or in nation; but may not. If the percentage of graduates in a given area is high, the question remains were they “produced” or “attracted?” If an area feels a need for more college graduates, such as Kokomo, Indiana, how much of its effort should be on attraction and how much on production?

Yet countervailing considerations do exist, as suggested by Figure 23.

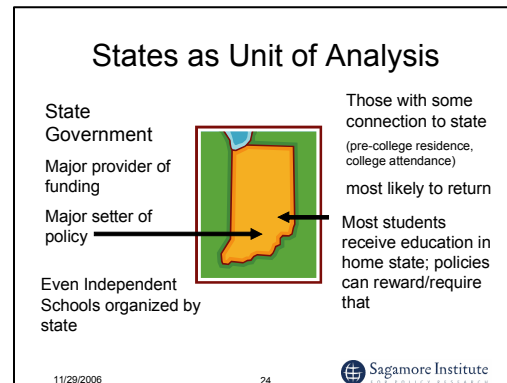


Figure 23. States as Unit of Analysis #2

The state government is a major provider of funding, especially to public post-secondary institutions. It is a major setter of policy, including taxes, articulation agreements, acceptable degrees, and other matters. Even independent colleges tend to organize by state, in part because the state is a major setter of policy.

Although students do migrate some, both into and out of post-secondary education, a large percentage attend

institutions in their state and stay in the state afterwards. Even those who leave are more likely to return than graduates with no previous time in the state. Moreover, states (and institutions) within them can certainly set policies to reinforce this process – such as lower tuition for pre-college residents, loan forgiveness for graduates remaining in the state, and more job placement services dealing with in-state employers.

Those who call states “laboratories of democracy” are certainly on point in this policy area. Each state provides a slightly to greatly different setting for college cost initiatives, as suggested by Figure 24.

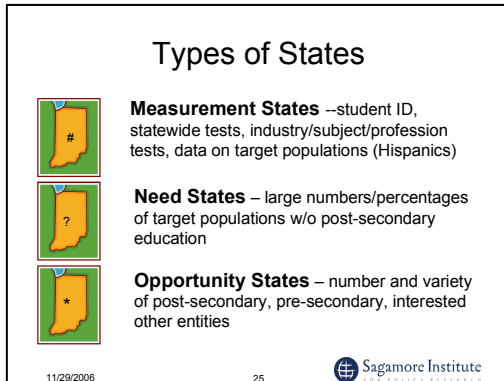


Figure 24 Types of States

Figure 24 suggests three different ways to categorize states: measurement states, need states and opportunity states. As mentioned above, some states, like Florida, have put in place, or are about to put in place, measurement aids, such as statewide “employment readiness tests” and student IDs, that are very helpful in measuring policy initiatives. Other states have special needs because they are behind in graduates or ahead in target groups of potential students, such as minority or immigrant populations. Still other states pose special opportunities because the state government or the educational institutions have special interest and capacity to work on the proposed initiative.

Of course, one state may fit in all three categories. Indiana, for instance, is an emerging measurement state, with an emerging immigrant population, and some institutions already hard at work in the subject area, and Lumina has recognized this with a special grant to work on immigrant students in Indiana.

8. The Scheme Revisited

The introduction has already set out the “bell the cat” statement of the scheme, as shown again in Figure 25.

By now, however, this report has put some flesh on the bones of the scheme. Section 7 has discussed the variety among states and how one might categorize them. Section 3 has discussed the various ways to categorize the initiatives. Sections 4 and 5 have discussed the measurement issues, with special attention to educational outputs and the connection to educational outcomes.

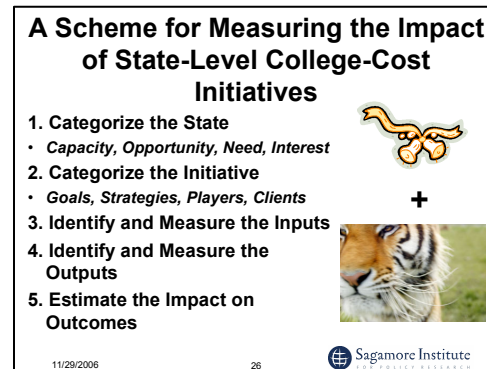


Figure 25. A Scheme for Measuring the Impact of State-Level College Cost Initiatives

9. Some Final Sagamore Lessons

One of the insightful requirements that Lumina imposes on its grantees is a call for lessons learned, whether directly related to the subject matter of the project or not. SIPR imposes this same requirement on itself and conducts regular “all-researcher” sessions to help share these lessons across disciplines and subject areas. This section sets forth the major lessons stemming from this part of SIPR’s project for Lumina. Since SIPR is a think tank, its lessons take the form of targets for further examination rather than final conclusions.

9.1. Is Time of the Essence?

Time to completion is an absolutely crucial component of student cost, and is usually a large component of total cost as well. Other things being equal, the earlier and faster in calendar time a given student can complete his or her post-secondary education, the lower the student costs and probably also the lower the total costs. Moreover, the student can begin making economic and social contributions earlier than otherwise.

The question is, can other things be equal? Do students still get the depth and breadth required to be “being appropriately equipped to make economic and social contributions at the specified level”? Is the learning curve the forgetting curve? What about the acknowledged value of “time on task”? Can a bachelor’s program finished in less than four years ever be the equal of one that took four (or more) years?

Time may be of the essence in another sense as well. The “gold standard” contemplates that students will be close to 18 when they enter and close to 22 when they leave. What if the student is considerably younger or older than that? Does the same education produce an inferior graduate? Put another way, do students outside the “gold standard” age require a different program in order to receive equivalent education?

SIPR does not have answers to these questions. But its research clearly reveals how central they can be to increasing the number and percentage of residents equipped to make economic and social contributions.

9.2. Is Less More?

If time is a key component of cost, could less of it, better spent, produce better outcomes? If subject coverage is a key

component of cost, could less of it, better learned, produce better outcomes? Should the T-shape of broad and narrow skills be applied on a course-by-course basis, with more attention to core concepts and less attention to broad ones? Robert H. Frank, full Professor of Economics at Cornell, is one who argues strongly for this point of view, at least in his discipline, and especially for students not majoring in Economics.

Should we ask/allow instructional personnel to spend less time per student, but better time, by freeing them from non-instructional duties, such as scheduling students and perhaps even grading papers and exams when such procedures can be automated or delegated to non-instructional personnel?

Bill Coplin (see reference at end of report) argues that we should give credit towards educational completion of activities we think are educationally valuable, even if now considered “student” or “extracurricular” activities. This might include participation in art, music, drama, journalism, athletics, or politics, with or without an accompanying “academic” course.

Potential lawyers taking the bar exam in most states are NOT required to take a bar review course. In some states, they are not even required to have graduated from an “accredited law school,” or even a law school at all. Especially at the introductory level, should we trust a “core concepts test” without any prerequisites? Can an AP-type system work, at least for some subjects, in the absence of any other articulation agreement?

Assuming a residence experience really helps, what about shorter calendar time but perhaps more intense experiences. For instance, some graduate schools of public policy start each entering class with a one-week program that runs from before breakfast to well after dinner in order to provide a short-term version of what residence in a dormitory provides over a longer period of time. Can short but intense residence ever be the equal or better than or almost as good as the more traditional nine months in a dorm?

Note that better measures of educational output would be especially helpful in these questions. One could measure output after a “less is more” approach and compare that to a more traditional approach.

Less may be more in the sense of reform as well. Educational reform is full of proposals to start completely over that die with a whimper, or at most very marginal tweaking. Sometimes, starting over does help – charter high schools are a way to innovate at the pre-secondary level that at least in their first few years of modern operation have found innovation much easier than inside existing traditional public or private schools. But often, low-level, reversible, cheap, low-risk even at the expense of low-gain, innovations are more successful in the long run than “throw it all out and start over.” Thus, trying these types of innovations at the course level and seeing if they will spread may be more likely to succeed than starting on a school-wide or nation-wide level.

9.3. Does Practice Make Perfect?

One of the fashionable educational phrases these days is “time on task.” Students who spend more time on a subject seem to learn it better than students who spend less time. Note however, that the “time on task” is not necessarily

correlated with calendar time per se; some of these advocates actually favor short-term intense activities like a day-long lab session over the same number of hours spread over several weeks.

Another way to think about this question is what might be called “the free-throw model.” Imagine we taught people to shoot free throws the way we teach them to perform economic analyses, for instance, with a traditional lectures/readings/hour exam/final exam process.

First we would have lots of lectures and readings about free throws – maybe including stories of people shooting them, maybe including pictures or even videos, with the teacher lecturing or assigning readings that critique these performances. We might – or might not – be encouraged to try shooting a few ourselves, but of course we could expect little or no feedback from the instructor. He might, in office hours, listen to our description of what we did and offer some suggestions. We would at least have a clear picture of the “output” we were seeking – the ball should go in the hoop, and maybe even some quantitative measure of what percentage of success was deemed “good.”

Then we would have a mid-term, where we were required to shoot 3 free throws. At a minimum, we would be graded on how many we made. We might – repeat might – be graded on our form. The grade might – repeat might – include a description of what we did wrong and how we might try to improve.

Then we would have a final, where we would shoot 10 free throws (3 times the hour exam, right?). Once again, we might be graded on how many we made and maybe on form, with or without comments – but of course the comments would arrive too late for us to improve our performance and/or our grade. It could be even worse – we might spend the mid-term and the final critiquing descriptions or depictions of others shooting free throws without ever being asked or required to shoot any ourselves.

Of course this is an exaggeration to make a point – creative writing classes are famous for demanding lots of writing, with lots of feedback by students and faculty. But it is not that much of an exaggeration, especially for introductory courses.

Now imagine the reverse – teaching students to perform economic analyses the way Gene Keady (long-time basketball coach at Purdue), for instance, teaches his players how to shoot free throws.

First, you would start the class with a detailed knowledge of the final exam – not just the subjects that would be covered, but quite a bit about exactly how they would be presented and how they would be graded. Moreover, you could administer a “mock” final exam to yourself as many times as you wish.

Second, although you would have lectures and live demonstrations and maybe even readings about economic analysis, you would be expected to practice actually doing economic analysis. You would be expected to practice a lot.

The teacher would not expect you to be good at it from day one (but might excuse you from class if you were). He would watch you practice, and give you personally detailed feedback on what you were doing right and wrong and how

you might improve. He might also have you watch others practice and listen and watch his feedback to them and their responses. At the end, you would still shoot the 10 free throws yourself, (do the economic analysis), but you might be allowed to do it over and over again until you and your teacher were satisfied with your performance – your ability to perform an economic analysis.

But does it cost too much? A basketball coach has at most a dozen or so “students,” and most of them are involved because they are already quite good at activities like shooting free throws. An introductory Economics course may have hundreds or more students, most of whom have had little or no prior exposure to formal economic analysis, let alone any chance to try doing one, and nowhere near one teacher per 12 students. Moreover, one can shoot 10 or even 100 free throws or more in the time it takes to do one economic analysis.

All true – but perhaps the administration and scoring of the tests can be automated and/or delegated to junior personnel in part because the effort to do so can be justified by the fact the tests will be re-used during a given course and during next sessions of the course. Also, they do not have to be created new each time, because they are no longer secret. At least during part of the course, students will be actually encouraged to help each other identify the correct answers and score each other’s tests.

If accompanied by more focused attention on key core concepts and less breadth, perhaps the approach need not cost more than its traditional alternative. Indeed experiments studied by Carol Twigg (see reference at end of report) found a variant of the “free-throw” approach, in standard introductory courses, actually rarely cost the institution more and often cost it substantially less.

9.4. Can High-Stakes Tests Be Turned Into Low-Stakes Ones?

The word “tests” in this lesson is a short-hand for “a measure of educational output.” The report has mentioned often the value of having generally-accepted, fine-grained, measures of educational output that could be used to compare alternatives to the “gold standard” and to compare cost reductions even within the gold standard.

But if such “tests” are what is known as “high-stakes tests,” they can create all sorts of counter-productive behavior. Cheating (before, during or after the test), and neglecting other aspects of education to “teach to the test” are just two of them. The current experience with “no child left behind” testing supports this concern.

Many academic tests are a one-time partial sample of the material the course was designed to cover, in large part because tests are deemed expensive to develop, expensive to administer, expensive to take, and expensive to score.

They can be very high-stakes tests because that single measurement, with no chance to learn from it, sets eligibility and or probability of further academic advancement or other substantial reward. Getting one shot at anything increases the psychological stakes tremendously, in part because all know that humans are not that reliable – even the best of us have “off times” and even the worst of us occasionally “get lucky” at least to some degree.

If that one shot is based on a partial sample of the material that was covered, the need of the test administrator to keep the test secret until administered, and the value to the student of having advance notice of which part of the material will be included, also increase the stakes. The incentive to cheat also increases, because the value of having advance notice or having chosen to study the “right portion” exacerbates whatever differential among students would have existed anyway, and lowers the amount of information that needs to be conveyed from one student to another to make cheating worthwhile.

Many question whether any single-time measurement of any human activity is a reliable indicator of anything. One does not pick a free-throw shooter by having him shoot the ball once, figuring that once is a reliable sample of how he will do time after time in the future. One is reminded of the aphorism quoted in current NFL ads – “the amateur practices until he can do it right; the professional practices until he cannot do it wrong.”

Also the modern industrial practice of statistical process control (SPC) originated in the auto industry, whereby Demming and others discovered that one could cost-effectively learn enough about a process to improve it by examining a sample of the parts or assemblies it put out, rather than every single one. This works because of the relatively low variation in the nature of the material being processed and the application of process to that material. No one would argue that students and the application of teaching methods to them are similar enough that one could test a few students and make reliable, bounded inferences about the rest of the class.

Even SPC rarely assumes that sampling or even exhaustive measurement of the results of one process yields valuable data about the results of other processes. Most people do not assume a student’s grade in course #1 tells us much of anything about her likely grade in course #2 – certainly not enough to rely on it. Yet a test will often make the assumption that if the student does not know in advance which of ten parts of the course will be on a test, one can reliably infer what her performance would have been based on testing her on two or three of them. Even though that assumption is often made in that context, it is rarely made in others. For instance, one would never assume a car is safe if automakers were told the following: “10 systems in each car need to perform at an acceptable level in order for a car to be deemed ready for the road. Therefore, the safety inspector may pick any two of those systems to inspect before declaring the car safe, and will assume that the other 8 will be up to the same standard as the two he inspects.” The car driver would insist that if all 10 are needed for the car to be road-ready, all 10 should be inspected.

Similarly, the car manufacturer would not accept a one-shot inspection. If the car fails, he wants a chance to fix the unacceptable system or systems, and have the car inspected again. Note also what computer quality control people call “regression testing.” You do not want to re-test only the failed system, since fixing it may have introduced new problems somewhere else. You want to re-test all systems, even those that passed before.

If we can make tests less expensive to develop, administer, take and score, then we can make them repeatable and dramatically lower the stakes.

First, if we can get beyond the “secret subsample” (i.e., study all 10 chapters because I am going to test a sample of 2 or 3 but you will not know in advance which 2 or 3), we can lower many costs. The expense of re-developing the test to cover a new subsample each time will go away. The expense of keeping secret the knowledge of which subsample is involved will go away. If we can get beyond the “secret” test we can (a) avoid the costs of keeping the test secret and (b) perhaps avoid creating a new test for each course.

One of the reasons for keeping an “all 10 chapters” test secret is that each question might actually be its own secret partial subsample of that chapter’s material. Another is that the student would learn mechanical rather than substantive answers to the questions – e.g., the answer to question #10 is A or “Mr. Green in the parlor is always the guilty party.”

Automation might provide some relief with the first problem, as it might lower costs enough to allow developing enough questions (or enough complications in a given question) to avoid or acceptably minimize the subsample problem. It can definitely help with the second problem by scrambling the order of questions and answers on each test administration, so the question answer combination that was 10A on administration #1 is now 7B on administration #2 and (ii) scrambling the names of characters or variables or the values of variables so the guilty character is now Ms. Blue in the lounge without changing the substance of the question. Note: Technology should allow for both intra-student scrambling (between the first and second times student A takes the test) and inter-student scrambling (between student A and student B taking the same test).

Another advantage of re-taking a test is that the test designer does not have to worry so much about misinterpretations by the student. If the student knew the item, but answered the question wrong because he or she misunderstood the question, the explanation that comes with the first administration, plus the chance to take it again, should help minimize the problem. It also helps with computer scoring, which can be more “fussy” about precise answers in fill in the blanks questions than humans. A human might give equal credit to both “reproduction” and “copying,” while a computer would not. But if the first administration of the test tells the student which word the computer needs to see, he or she can use that word the next time.

One SIPR researcher has seen this phenomenon work three times already. First, when his daughter was studying for her AP test in biology, the online accompaniment to the textbook worked just this way. As parent, he told her to repeat the tests until she could show him a printout with all of them right. Note that the test results were more than a score. They identified which questions were answered incorrectly, what the correct answer was, and why it was correct. He did not have to stand over her at any step in the process, nor inquire how many times she took the test. When she took her [very similar] high-stakes AP test, she was very relaxed and did get a score that qualified for college credit.

Second, SIPR knows a firm that develops simulation software for very expensive CNC machines (costing millions

of dollars to over \$100 million). Downtime on the machines is very expensive, as are repairs, and the machines operate at such high speeds on such hard materials that they can be extremely dangerous if operated incorrectly. For all those reasons, management insisted on a relatively high score on the simulation test for each and every operator who would touch the machines. The operators are highly-paid and highly-regarded and unionized, so the union leaders and the rank and file were very leery of the test until management agreed (a) each potential operator could take the test as many times as needed in order to pass it; and (b) management need never know how many times a given operator took the test. Opposition went away, and operation has been much safer and more cost-effective ever since.

Third, a number of online “learn a language” web sites work this way. However, the counter example in this case was that the teacher insisted she wanted to see only the results of the first time the student took the test. That requirement converted a low-stakes activity into a high stakes one (for the class of very “competitive” lawyers) and created an incentive to cheat/shade the truth that would have been entirely absent if she had adopted the “test until perfect” standard used in the previous two cases.

Back to the question of costs in the free-throw model: If the student tries to re-test until perfect and cannot get there, with all the online help that can be built into the system, he or she has a very focused and demonstrated need to interact with instructional personnel. It is much easier and faster (and cost-effective) to do such interactions, and thus a given faculty member can actually handle far more students far more effectively than under a “come see me during office hours if you have questions about any of the readings or lectures” system.

Some have raised the concern that increased “measurement” could lead to a “tracking system” that would further divide students by educational achievement than they are already. Especially since current student achievement is highly correlated with the stability and income of the student’s family, such further division might exacerbate separation along racial or income lines.

There is no question that it could be used this way. But it need not be. If such measurement is used to identify students who can be moved on, and those who need help, and help is available, then it can actually reduce the tensions that stem from providing one size fits all service to a group of very different sizes.

Another concern often raised is that teachers will “teach to the test” and neglect other material. The problem is particularly acute in “secret partial subsample” tests where the teacher, but not the students, knows the secret partial subsample and teaches only that – say only the vocabulary words on the test. However, if the test is comprehensive, rather than a partial subsample, the material should be the same. Also, if the test is not secret, the advantage gained by teaching to the exact form of the questions goes away. But the test does have to be a good enough measure of output that teaching to it is the same as teaching to the course objectives. If the test is not that good, the solution is not to rail against teaching to it, but to change the test.

Here again, there is some hope that automation and reuse can make tests easier to design, administer, and grade, so that changing the test is feasible.

9.5. How Do We “Incent” Cost Reduction per “Learned Unit” per Student?

In some cases the goal is to reduce “total cost,” not just student cost or cost to any other specified party – in other words, to lower the lump, not just push it under the rug. In those cases, SIPR has learned that all of the parties currently involved, except perhaps for Lumina and SIPR, face weak incentives to lower such total cost and may face strong counter-incentives.

Consider students and their families. Scholarships, loans, and perhaps calendar reduction may be enough for them. They might possibly think about brothers and sisters or future generations, but those would ordinarily be weaker considerations.

Colleges and other educational providers are currently rewarded for high resources per student – small class sizes, exalted faculty, the full panoply of non-direct student activities. In addition, most assume that such a full panoply contributes to quality of student preparation. The Michigan state government’s program to reward tuition “decreases” with research grants may be a small counter-example. Individual teachers are not generally or directly rewarded for teaching more students per class or otherwise reducing the cost per student taught.

Governments and other interested parties are often “buying” much more than student learning units, such as research, economic development, entertainment and thus while not opposed to reduction in total cost, it is not their first priority. Kansas “buying down to in-state tuition” for non-Kansas facilities may be a small counter-example, as may be grants that lead students to academic 2-year institutions for their first two years rather than directly to academic 4-year institutions

Suppliers to students or providers (e.g. textbook publishers) might have incentives already – but hard-to-find clear examples of learned-unit related (a) price-elastic demand; (b) increasing returns to scale (printing one run of a paper textbook is small counter-example); or (c) network effects. For instance, before a specific effort in California focusing on textbook cost, professors lacked both the incentives and the information to choose a cheaper textbook over a more expensive one, or an older version over the newest. There are few, if any, mechanisms in place to get a textbook cheaper if two professors adopt the same one.

SIPR did challenge itself to suggest some examples of how to provide such incentives and came up with the following.

9.5.1. Bid Out AP Courses

In cases where we have (a) an agreed upon test, (b) an agreed upon method of administration, and (c) an agreed-upon method of scoring, one could imagine issuing bids for entities promising to produce students passing the test. We do have such a system provided by the College Board – it is, after all, 37 Advanced Placement (AP) exams across 22 subject areas, with more coming. Each has a collection of study aids, practice tests (based on former tests) and other study aids,

and existing credibility or even explicitly college credit standing with a range of institutions, including some of the most highly-rated. (See <http://www.collegeboard.com/student/testing/ap/about.html>)

The system at present does not follow the full “free-throw model.” It keeps the actual test secret. It does not provide any feedback other than the 5-point cumulative scoring system. As the web site says “Sub-scores are not available (except for Calculus BC and Music Theory) nor are scores analyzed to determine strengths and weaknesses.” (See <http://www.collegeboard.com/student/testing/ap/exgrd.html>.) It does not allow re-testing. It currently charges \$83 per test taken (the school administering the test may require additional proctoring fees), and it uses computer scoring on its multiple choice sections but uses a panel of college professors for its free response section.

Although there could be a much more free-throw like “test until passed” system, this system is in place, and imposes few if any “course requirements” (home schooling and online courses, as well as independent study, are all explicitly allowed). It may well suffice for a smaller step forward. So one could imagine (if it does not exist already), a post-secondary provider or educational funder (state government, federal government, local government, foundation) putting out bids for AP courses, conditioning payment on so much for each 3, so much for each 4, so much for each 5, etc., with or without a base amount per student involved regardless of grade. It could award bids via cost – either lowest bid per student grade, or largest number of students per total amount, etc. The bidder might share revenue with the students in a number of ways – offering discounts if they get higher grades to encourage effort, or since each grade that qualifies for college credit is worth several thousand dollars in reduced costs to the student, perhaps offering refunds to students who fail to achieve the higher grades or requiring a “bonus payment” from the student if they in fact get the higher grades.

9.5.2. Pay Premiums per Learning Unit per Student

Educational providers generally require faculty to teach a certain number of courses, paying per student taught (so larger courses would pay more) is generally not done. Nor is the faculty member held to any more than a very implicit standard of how much or how well his or her students learn; indeed, although the provider often gives explicit guidance as to the grade curve it strongly suggests, it rarely gives guidance on how to evaluate whether the students learned the material or not. Occasionally multiple teachers will use the same test, and some degree of comparison between classes is made, but since students are rarely randomly or carefully assigned between the sections, the information gained is less than might be hoped and is viewed with those major caveats in mind. Such comparisons are generally not given enough weight to explicitly reward the faculty member whose section did best by some measure of group-to-group performance.

Online/distance education and other “out of the mainstream” courses or courses taught by adjunct and other “out of the mainstream” faculty are sometimes an exception in that the faculty member is paid per student involved, but here again, without any measure of “learned unit per student.”

Faculty textbook authors can earn royalties per textbook bought, or other teaching materials, such as software programs, tests, teacher aides, et al. but since they have little control over which teachers adopt what and how many students each of them have, plus the items tend to be price inelastic; neither the student nor the faculty member considers the price very much when determining to buy the item, except when the price is extremely above or below the expected price.

9.5.3. Allow Cross-Bids with Kickbacks

Even in states without articulation agreements, one could imagine a two-year institution in the same city or town as a four-year institution (or nearby) offering to the four-year institution or to the common funder of both (say the state government) to take over teaching of some first and second year courses at a cost per student less than that of the 4-year institution. A major constraint may be demonstrating that the students do not thereby suffer a diminished learning experience, and an agreed-upon test (with agreed-upon administration and scoring) may be a key ingredient in reaching agreement. Another may be some form of cost-reduction sharing. If led by an educational funder, it may simply prefer to pay less per student to the 2-year institution rather than more to the 4-year one. But politics are rarely that straightforward, and one might hope for situations where the two institutions would agree among themselves. In that case, the cost-reductions might be split in some fashion. At one extreme, the higher-cost institution keeps the difference to apply to other purposes; in the other, the lower-cost institution does it instead. In between could be a range of splits in money or other resources. The goal is to explicitly reward somebody for achieving the cost reductions.

Universities have worked out such splits with regard to revenue from licensing of university inventions. Some portion goes directly to the faculty and other individuals involved, some portion to the department and/or school which houses the individuals, some to the technology licensing office, and some to the university's general fund. One could certainly imagine the same with regard to cost reductions.

9.5.4. Block Grants

One could imagine educational funders including funding from central administration to sub-units or from one sub-unit to another, structured to encourage per learning unit per student cost reductions. One such structure would be block grants, rather than grants divided into so much per student, so much for capital, so much for equipment, so much for academic staff, so much for non-academic staff, etc.

Here again, agreed upon "output measures" could greatly facilitate the trust by the transferor in the transferee. If one could demonstrate the nature and amount of the learning absorbed by the students, most likely by using agreed upon tests with agreed-upon administration and scoring, it would help a great deal.

9.5.5. Involvement of Students

The college cost area is full of initiatives designed to shift costs away from students, in the form of lower tuition and fees (make up the difference from other sources), loans and

grants, and other such mechanisms that have little or no effect upon total costs. But students can be involved in ways that do reduce total costs, although perhaps at the expense of greater intellectual effort by the students. One of the indirect ways is that many universities, including Indiana University, charge so much per credit hour per semester, but have a flat range (at IU, 12 to 17 credit hours per term). So it costs the student no more in credit-hours fees to take 17 hours than it does to take 12 credit hours (it may well take more effort, and more costs for books and other such items). If this flat range in fact represents a diminished marginal cost to the university, than a program to encourage/facilitate students taking closer to the 17 hours than the 12 hours would be a student-involved cost reduction activity.

Another way suggested by some of the Lumina-funded analysts is to increase student displacement of university employees. Students have long served in dining halls, in dorm cleanup crews, and even in research projects. The proposal is to employ upperclass undergraduates as course assistants, tutors, and graders and in other appropriate academic roles, perhaps using "the best way to learn is to teach" approach to award a combination of credit and pay that would lower costs over using non-student employees, also taking advantage of the fact that students already have many of the health coverage benefits by virtue of being a student.

9.5.6. Involvement of Future Employers

Technical and professional schools have leaned on future employers of their students for decades. By providing paid and unpaid internships, "real" research problems or data or access to costly equipment or supplies, unpaid or quite low-paid adjunct faculty, thesis advisors, all subsidized by the value the future employer gains in access to potential employees, actual and potential clients, these outsiders make substantial contributions to student learning, at little or no explicit cost to the educational provider. One need not imagine the expansion of this phenomenon; it is expanding already, with growing involvement of government agencies with related schools and departments, nonprofits, and even arts, music, and athletic organizations. The challenge is to run these relationships in a way that lowers the total cost, rather than simply letting the outside agency pay a portion of the existing and usually increasing costs, and in this instance, to figure how to encourage the parties involved to work on that goal instead of merely cost-shifting.

9.5.7. One Final Point about Incentives

One of the hardest changes for an institution to make, especially one that has been relatively stable since around 800 AD, is to drop existing activities. It is much easier to add programs or even schools than to drop old ones. Ordinarily, it takes a real financial crisis, or an extremely long time, or both, to do so. The same is true of ways of organizing the work that gets done.

As a consequence, even in the most modern universities, the "craft of teaching" usually bears more resemblance to a cottage industry than to a modern factory. Each faculty member, with perhaps a secretary and a graduate assistant or two, is usually expected to oversee if not conduct all aspects of the operation – designing the course, giving the lectures,

designing the tests and other assignments, administering and grading them, and dealing with all personnel issues associated with the students. Of course, just like the cottage workers of old, the fellow craft workers swap tips, occasionally help each other out, and the like, but stay a far distance from anything resembling a production line with specified processes and specified outputs.

The issues are whether a college cost policy initiative can provide incentives to move to a more modern form of organization, at least in some cases, and whether the resulting educational outputs will be acceptable.

9.6. Can We “Open Source” Education

The final lesson asks about an analogy to the open source software movement, where such software is a viable alternative, although not a replacement for commercial software. The concept is that a number of people would contribute to a body of teaching tools – course syllabi, assignments, tests, lectures, and readings – that anyone would be able to use, perhaps with some requirements for attribution, but without paying fees.

The past few years have seen several emerging activities that constitute steps in this direction. For instance, MIT has put a large percentage of its course syllabi online, with no restrictions on use by others. Berkeley is putting video files of its lectures online. Some websites that are available for free use come close to approaching the body of material that would be found in a textbook – cases, questions, short essays. Others offer the same plus various sorts of online tests.

Dewayne Matthews of Lumina (see reference at end) has suggested that content is no longer a way to differentiate schools, at least at the introductory level – in large part, we want the same material covered in introductory Economics or Psychology or Chemistry regardless of the institution providing the course. In the medieval church, the organ master was expected to compose something for each Sunday. Now we expect them to play material already composed (of course, a few do still compose, but they are the exceptions). Although we still expect the preacher to compose each Sunday, that too may change someday. Certain schools now hire faculty to teach courses that are already designed, sometimes complete with assignments and exams.

The possibility is that, with open source materials, the institutions could compete on new measures of excellence, just like providers of open source software have to compete on service, rather than the software itself. As in software, no one should expect that “open source” education would replace “closed source” education. But it might provide a very useful supplement.

10. Selected References

Twigg, Carol A., “Improving Quality and Reducing Costs: The Case for Redesign,” pp 32-49 in *Course Corrections: Experts Offer Solutions to the College Cost Crisis*, (Indianapolis IN: Lumina Foundation for Education, Inc., October 2005)

Matthews, Dewayne, “Toward a New Way of Thinking: Quality, Productivity, and College Costs,” pp. 104-111 in *Course Corrections: Experts Offer Solutions to the College Cost Crisis*, (Indianapolis IN: Lumina Foundation for Education, Inc., October 2005)

Coplin, Bill, “Seven Steps: Ways to Reduce Instructional Costs and Improve Undergraduate and Graduate Education,” pp 20-31 in *Course Corrections: Experts Offer Solutions to the College Cost Crisis*, (Indianapolis IN: Lumina Foundation for Education, Inc., October 2005)

Carey, Kevin, *College Rankings Reformed: The Case for a New Order in Higher Education* (Washington DC: The Education Sector, Inc., 2006)

Dr. Roland Cole (rollie@sipr.org) is a senior fellow and director of technology policy at Sagamore Institute for Policy Research.