

A REFLECTION ON A DAY SPENT DISCUSSING EVALUATION

David Shulenburg
Provost
University of Kansas

Evaluation, our topic for this retreat, is part of a much larger activity in which universities are engaged. Initially, the focus was on assessment of graduate and undergraduate programs. Now essentially everything in university life is subject to extensive evaluation.

There was a time when the academy was a more self-contained environment and informal evaluation of its activities was sufficient. This began to change with the advent of significant federal funding as society began to look to universities as vehicles for improving the world. Once this expectation was placed upon us, we were obliged to demonstrate what we were accomplishing.

Because the missions of universities are amorphous, no clear measures have emerged. Our situation contrasts sharply with the business world where evaluation is easy. The standard there is whether or not you make a profit. There are many who quest for a simple, straightforward evaluative model like that of the business world for universities. They yearn for an evaluation of university effectiveness so simple and clear that it cannot be argued or debated.

A storm of controversy engulfed academia as a result of the release of the National Research Council (NRC) evaluations early in the 1990s. In spite of all the adjustment and tinkering that will occur before NRC measures us again, I suspect that the uproar will be no less this time than last. What does that tell us? We don't agree among ourselves about what quality is. We don't agree about the objective function of evaluation. I don't mean to criticize the NRC process. What I suggest is that we are a long way from having an absolute criterion for reaching the gold standard of university evaluation, a simple non-debatable measure like the business world's profit measure.

We've talked about market models of evaluation at this conference and they surely should be considered as they begin to approximate the accepted model of the business world. Examples of "market measures" are the quantity of external funding the university attracts or the number of students who choose to attend. These models are often toyed with, but the simple problem is that the market model doesn't work unless you can demonstrate that you are covering the full cost. Essentially all of our activities are subsidized. Unless we can

assure the public that we are not “buying” our market with our subsidy, the market measures we develop will not be accorded respect by the public.

We should be aware that some of the measures we use are not persuasive to the public and for good reason. It does not persuade the public when we say that our graduates have higher incomes because we educated them and they consequently pay sufficiently higher taxes such that the state is now earning a surplus because it funded higher education. That argument fails not because the public refuses to believe that higher education raises income; it fails because the public knows that if we did not provide higher education to citizens of our state someone else would. For decades New Jersey chose to have almost all its students go to private universities or go out of state for higher education because of the dominance of this logic. Similarly, a claim that we have significantly increased revenues for farmers by developing a high yield variety of wheat does not work unless we are measuring only the difference in yield between our variety and that of the next best producing variety. The public is sophisticated enough to credit us only with the unique gains that we produce.

Leverage is another kind of argument that we make in support of our impact. We argue that if a certain amount is allocated to the University of Kansas, the institution will return to the state’s economy a multiple number of dollars. Is leverage not persuasive? It is, but the public and legislators know that the leverage we create ought to be compared with leverage that might be derived if the funds were spent on some other function of government. Is our leverage really better than that derived from money spent on traffic safety? On early childhood education? On expansion of the lottery?

Suzanne Ortega said that she lost more personal capital talking with faculty about assessment than any other activity. Why is this? I think it is because faculty do not really believe that assessment measures what education produces. The faculty believe assessment is something we have to do to satisfy those who accredit us, but they don’t believe that assessment measures whether the education we provide makes a difference. They see assessment as spending lots of time with activities that aren’t of value and which produce measures that won’t be convincing to others.

Our discussions have included the great importance we accord to evaluation of faculty based on publications in refereed journals. This is an important issue to me. I’ve talked with many academics, librarians and commercial publishers about this topic. Some believe that universities created the problem of rapidly increasing journal costs by rewarding faculty based on the number of articles they publish in refereed journals. Not everyone is convinced, as I am, that journals contain important research findings and new insights. Much of the rest of the world thinks we’re playing games with the articles we write, and the commercial publishers believe that as long as we are stupid enough to

continue the game, they are going to make money from it. Cynicism is raised to exponential levels by this cycle.

We risk being self-congratulatory by putting time into evaluations of our own construction that the rest of the world does not accept. How many times have you heard someone outside academe or the federal government talk about NRC rankings? The silence on this issue from the rest of the world reflects the fact that NRC rankings have zero meaning to them. What ranking does mean something to the public? For the people on the street, the most credible ranking is by *U.S. News and World Report*. Interestingly, academics tend to think it is meaningless, but the public demonstrates belief by continuing to make the rankings issues of *U.S. News and World Report* best sellers.

“Aunt Emma” stories are also effective. The credible information most people use to evaluate our institutions involves the stories about specific students whose lives were changed because they came to our universities.

If our evaluation schemes do not measure up to the high ideal of quality held by academe, we are endangering ourselves by engaging in them. The public will see us as phonies if we use measures of our effectiveness that even we cannot accept as legitimate. The public is willing to tell us that we are wearing no clothes if we insist on parading about naked.

There are reasons for measuring external funding—to encourage the faculty to seek it, and because we don’t have enough money to fund their research unless they seek it. There are good reasons to fund research competitively—because awarding funds objectively puts scarce resources to the most valuable use. If we use external funding as a measure, the reason should be that we want to encourage our faculty to put effort into successfully acquiring external funds. We should be very careful about arguing that we are a better university because we bring in more external funds or that society is better off because we are successful in the external funding game.

Susan Kemper expressed a healthy skepticism about the process of evaluation. It is important to keep her message in front of us. Evaluation can be a useful activity, but we must keep our perspective clear and claim no more for the activity than it merits.