



Association for Public Policy Analysis and Management

**2005 Spencer Foundation Distinguished
Lecture in Education Policy and Management**

“Policy Analysis: Is It, or Could It Be, the Fifth Estate?”

By Eric A. Hanushek¹

© 2006 Association for Public Policy Analysis and Management.
All rights reserved.

¹ Hoover Institution of Stanford University, University of Texas at Dallas, and National Bureau of Economic Research. The ideas here benefited from extensive discussions with Macke Raymond along with interactions with the members of APPAM attending the lecture. This text is revised from the lecture originally delivered at the APPAM Fall Research Conference in Washington, DC, on November 3, 2005.

Policy Analysis: Is It, or Could It Be, the Fifth Estate?

By Eric A. Hanushek

The Association for Public Policy Analysis and Management was born over a quarter of a century ago in Chicago. The somewhat awkward first research meeting in 1979 has blossomed into a mature and vital organization. The first meeting was an interdisciplinary meeting but the attendees quite clearly wore their battalion insignia on their sleeves. And the main criteria for presenting a paper seemed to be a stated interest in some policy. Few would have thought at that first meeting that there would so soon be the wealth of insightful and well-crafted research papers and scholarly discussions that we have come to expect at these meetings.

The journey can also be charted by seeing the changes in the *Journal of Policy Analysis and Management*. Volume 1 appears to be a third cousin of the current journal. There is a slight family resemblance but the analyses are very different in sharpness of the question, in the skill of the execution, and in the implications for policy. I don't really want to talk about history, however. I want to talk more about the future.

It is also important to recognize the sponsorship of the Spencer Foundation. The Spencer Foundation has long had a mission of improving education in the U.S., but its involvement with APPAM represents a new and exciting thrust of the Foundation to open discussion of education policy beyond the traditional (and insular) research that has dominated the field of education. I will draw examples and lessons heavily from education. This focus results in large part from my greater familiarity with the area and not as much from a strategic choice based on substantive or analytical issues. But, I also think that the lessons from education generalize to other important policy fields.

This talk, while wandering around across some of my own experiences, has a simple but direct focus: If the research is so good, why don't the policies improve? This issue has been raised many times in the past, but I want to sketch some ideas about changing this. Specifically, I think that we in APPAM have avoided these issues and that it is time to contemplate how that

might be changed for the good of the policies about which we are concerned and for the good of society itself.

Fantasies of the Neophyte

I will begin with some observations reflecting my views when I came to the scholarly basis of public policies. I describe them as purely personal, but I suspect that they at least represented common views and perceptions of a large number of APPAM members at one time or another during their careers. I should also say parenthetically that they are views that I would still like to believe are true, even if experience has led me to some revision of them.

During graduate school, people follow a variety of paths denominated by discipline, area of interest, theoretical perspective, and the like. I want to focus particularly on the group of people with an interest in substantive public policy. Or, again, autobiographically, I focus on my own thoughts as someone quite interested in how economics could be related to public policy issues and, at least for a portion of my graduate work, to educational policy issues.

What did I believe about the world? First, graduate school convinces you that the world begins with a well-defined question. With the addition of a clever research design and a dose of data, it is possible to provide a direct answer to the question. And, the key underlying belief is that, given the answer, the policy adjusts to it.

In this ideal world, completely independent analysis leads to clear improvements in policies. The underlying assumption is that people (policy makers) want accurate, complete, and unbiased analysis. In other words, information carries the data – data, insight, and youth (?) will speak truth to power AND power will listen.

Realities of Politics, Interests, and Analysis

It is difficult to assess precisely how and when policy analysis enters into the decision making process. Casual evidence would suggest that the use of analysis runs the gamut from 0 to 100 in policy making. For example, I think there are many tie scores for analyses that have zero impact. One of my candidates in education for close to the 0 end would be class size reduction, although a variety of other things such as analyses of teacher salaries are rivals. At the other end of the spectrum, I would naturally go into other areas. Perhaps an example of near complete adoption of analytical suggestions would be the move to auction of the radio spectrum

– although there we have seen the needs for analysts to return to the problem to deal with difficulties of the original introduction. Maybe another policy that is in the upper register would be the development of new welfare policies, even though analysts may not be completely satisfied with the results for a variety of reasons.

Beyond the simple notions of whether analyses become policy or not, it is useful to tote up the how the elements of the naïve view of policy process appear to hold in reality.

The need for multidisciplinary approaches to many policies has long been a mantra of public policy analysis discussion, and seeing the interaction of analysis with policy discussions reinforces this idea. At the same time, a brief consideration of multidisciplinary research, in part informed by wandering through the pages of *JPAM*, suggest a reality that does not match the rhetoric. The attacks of different disciplines on a common problem, even when contained in a single article, tend to lack a common language, generally replete in fact with differing jargon. But, more importantly, the differing views of analytical structure and of standards of proof leave the audience and the clients of analysis confused.

The picture from the demand side also enters into the reality. In education, it is easily summarized by contemplating the question, “who was the last governor who did not run on the platform of being the ‘education governor?’” (While historians may be able to document this more precisely, my candidate is Thomas Dewey). But if we think about the direct implications of this, it is immediately apparent that it is hard to follow this political theme with a platform that involves a forceful statement such as “the centerpiece of my position is an aggressive effort to find out what works in education.” Indeed, from my vantage point in California, one needs only to think back to 1997 when Governor Pete Wilson, looking for an education policy that would appeal to parents and voters but that would not (he thought) all go to the teacher unions, announced his plans to reduce class size throughout the state. Casual analysis suggests that the immediate jump in his ratings in polls and not any newfound evidence of the effectiveness of this policy led 23 other governors within a couple of months to announce their class size reduction programs and led the President who oversaw the modest and restricted involvement of the federal government in education policy to follow suit. The follow-on was perhaps more interesting. An intensive effort was launched to backfill the evidence on class size reduction.

Indeed the prevailing interpretation of the evidence at the time of Governor Wilson's announcement was that it quite clearly pointed against any general class size reduction policies.

This policy atmosphere is not one that engenders a broad and unbiased search for the evidence. Indeed the California experience today provides an apt summary of the situation. State spending now approaches \$2 billion each year on its class size reduction program. Because of the universal introduction of the program throughout the state, the experiences with the program can provide virtually no useful information about the efficacy of the program. And, even though the program will, barring some unusual circumstance, likely stay in perpetuity, nobody can say whether it has *any* positive outcomes. The only performance measure is an accounting of how much money is spent annually on the program.

The search for evidence in this example highlights a frequently expressed view: The use of evidence is often very selective. In particular, a cynic might believe that a variety of studies are highlighted in the policy development process more for their answer than for their persuasive research design or their scientific merit. This possibility is regularly discussed in introductory policy analysis courses – and seldom in a positive way.

But another element of the control of information and the ability to evaluate programs is often overlooked. Analyzing the outcomes of programs generally takes the cooperation of the actors in the process. This issue comes up at every stage of the process. Think again of the California class size reduction program. Policy makers are fully vested in programs that they develop and often oversell. If policy makers or the governor were seriously interested in evaluating the effectiveness of the program, they would have thought about pilot programs or staggered introduction of various components of it. In contrast, one of the ultimate concerns about the program was that it was introduced very rapidly with virtually no warning, so that school districts were unprepared for an orderly development of programs – lacking not only physical classrooms but also qualified teachers. Moreover, this surprise introduction hurt urban districts who saw experienced teachers being bid away by suburban districts, even when the program was advertised as being aimed especially at improving schooling for disadvantaged students. But it did effectively preclude any evaluation.

Program introduction is not the only area where information control is relevant. In order to analyze on-going programs, researchers must have access to information about the operations

and outcomes of the programs. Actors who like specific programs and who would not want to see reductions in support may strategically prefer that the program is never evaluated. Similarly, withholding information may be an effective way to prevent "bad" things. For example, the federal accountability legislation (No Child Left Behind, or NCLB) mandated annual testing of students in grades 3-10. This assessment system would permit direct investigation of the impacts of teachers, programs, and schools. Faced with this possibility, the teacher unions in California and other states have set out to block efforts to assign unique identifiers to students, thus making it impossible to track student performance over time and across districts. This action was initially successful in California, but, under considerable pressure, the legislature reversed itself and required development of a system of student identifiers. The state department then began a multiyear planning process to decide how to develop such a data system, even though a majority of the people attending this Spencer lecture could complete most of the task before the lecture ends.

In other states, considerable effort is made to control the extent and form of any data releases. While seldom stated explicitly that this is done to block any research and evaluation, it is very consistent with such a hypothesis about its purpose. Such things as limiting any details of the distribution of student achievement, not releasing any data in usable electronic form across school districts, withholding programmatic detail, and the like appear in many cases to be related to preventing any independent analyses of schools and programs. And, while there are complicated issues surrounding privacy and confidentiality – a matter of federal law under the Family Educational Rights and Privacy Act, or FERPA – at least of portion of the time when it is invoked as the reason for not releasing data appears again related to a more narrow interest of preventing outside evaluation.

The implication of these issues about gaining access to information and to gain analytical opportunities is that quite generally it is necessary for researchers to work closely with the states, districts, and program personnel – who may have a stake in the analytical results. In other words, researchers interested in working on some of the most interesting projects must become involved in the political process. But, the closer one gets to the political process, the more likely the work is to be judged as advocacy with potential biases.

The interplay between analysis and data access introduces a kind of selection bias that we normally do not consider very much. In all of our empirical work, we are now trained to sniff out possible sample selection biases. Since the early observation by Jim Heckman that that analyses of working women might be affected by the special nature of the sample and might not generalize to statements about nonworking women, most analyses of individual behavior appearing in *JPAM* now have explicit consideration of how the sample came into being and of any possible biases of that.

The issues of data access create what might be thought of as a first cousin of the Heckman selection problem. Consider the meta-selection problem in evaluating schools that comes from data access issues. In education research, a substantial proportion of the new and interesting findings have come in the past decade, owing to the availability of large scale panel data on student achievement. A few states – Texas, North Carolina, Florida, and, to some extent, New York – have allowed researchers access to panel data covering most if not all students in the state. A few cities – Chicago, Los Angeles, and New York City – have also participated in varying degrees in providing access to data.

These locations are not a random selection from the nation. We know immediately that they currently differ in some ways from the vast majority of the remaining states. First, they have the capacity to produce useful data for the researchers. Second, they are interested enough in the potential research that they permit it to go on. Third, they are willing to take the potential risks that some of the research may be politically embarrassing to them. Fourth, they tend to differ in observable ways from other locations through programmatic and population differences such as having a long history of school accountability or having high percentages of minority students. In what other ways might they differ? And, are the results of the research affected by the special nature of these states?

This meta-selection problem comes in when we consider how to generalize any results. For most purposes, we are not interested in, say, an evaluation that applies just to a single school. If some school, say a magnet school or a charter school, invites a researcher in to assess its programs, the question is always whether the results of this evaluation might apply to any other schools. But the same presumably applies to states or cities that invite evaluations. These are

issues that have not received much attention and that are quite difficult, especially as the research enters into new areas where there is little experience or past results to build upon.

Let me summarize this walk through the realities of educational policy. Reality is a distant cousin of the ideal that many of us had at entry into the profession.

The message is "suck it up and deal with it." If any contribution is to be made, we must adjust to the world. Reality is not going to change to us, at least in my lifetime. It is a wasted vigil waiting for the world to change for us. It is not going to happen through waiting for truth to be seen.

If we want to see something different, we must make it happen. We must adjust, because the world will not adjust to us.

The Fourth Estate: The Media

Before considering our business, let's look to the mass media. The mass media has been given special status as the "fourth estate." The media thinks of itself, and wants to be thought of, as an agent to keep the system moving in the right direction.

Its status comes from balancing the self-interested positions of politicians in the other three estates. (We now think of the other three states as the executive, legislative, and judicial branches, but the origins of this idea identified the clergy, aristocracy, and commoners). The power of the fourth estate comes from shining a bright light on excesses and wrong-doing, thus offering some restraint on unbridled self-interest.

The fourth estate is of course increasingly subject to a variety of pressures and shortcomings. In today's world, one of the most obvious pressures comes directly from competition among different information sources. To grab an audience in the nonstop world of work and entertainment, there is an elevated importance of the sound bite. In-depth reporting only seems to exist in cities with long commutes on public transportation – and even there the Ipod is making considerable inroads into available time. A related aspect engulfing significant portions of print and electronic media is a quest for sensationalism.

A quite different element entering the equation is the complexity of many of the substantive issues of the other three estates. This complexity implies a steep learning curve and limits the field of vision and insights of the media "generalist." Indeed, career pressures on media personnel often call for moving across posts. When development of expertise centers on

on-the-job learning, the analytical capacity of the media then devolves to trying to identify experts with specific knowledge and hoping that the main messages can be extracted on the fly. Education and a number of other social policy areas are generally viewed as low status in the media, leading education reporters to seek other assignments quickly and leaving little experience in the area.

Finally, there are issues of editorial focus (e.g., the New York Times and the Wall Street Journal or Fox News and CNN). Of course, the news and editorial portions are supposed to remain separated, but there is, I think, some bleed across the barriers.

These are not arguments against having media or against supporting their watchdog role. But instead I would argue that the media is increasingly less central to "keeping the system moving in right direction."

As government pursues evermore complex programs and policies, societal welfare is increasingly dependent on making the right policy choices. Scandals of self-interest and graft happen, but they are secondary to the decisions that are being made.

Education provides a clear set of examples. There are occasional scandals – a superintendent on Long Island or a union leader in Washington who pilfer funds. But, the much more important issues revolve around policy decisions that affect what student learn. And, here, self-interest is permitted, and reinforced, by the media that is apparently unable or unwilling to distinguish among policies that improve student learning and policies that enhance the personnel in the schools. Policy is affected by legal influence buying – campaign contributions and direct involvement in low turnout elections – but these are not scandals. The closest media reporting comes to this issue is a perennial story about the magnitude of political contributions of the teachers unions. This story, generally by the political writers and reporters, is almost never linked to discussions of educational policy initiatives and decisions. Clearly, there is not a direct trade-off in the dimensions of student interest and adult school personnel interest, but just as clearly the best policies in either dimension are not obviously the best policy in the other.

On the substantive side, the mass media struggles to understand different policy issues, but usually the best the reporter/analyst can do is follow the editor's dictate to "find the other side." This doesn't typically lead to a very satisfactory discussion, given that the reporter lacks any real ability to judge the evidence.

Policy Analysis: The Fifth Estate

Let me return to my main message. *As a society, we are not solving the most pressing problems of the day. This will not happen until we (the people in this room) have a more central role. And, that won't happen unless we work to make it so.*

Let me parse this message. First, there is an assertion about the state of policy outcomes today. This assertion is perhaps clearer in education than other areas. I will not attempt to document it in detail, but the flat performance of students on our own national tests (NAEP), the embarrassingly bad comparisons to students in other countries on international tests, and the obvious mess right here in the schools of the nation's capitol make a compelling case. Second, the message turns to what some might take as an arrogant statement about our potential role in helping solve some of the pressing policy problems. The foundation for this is my belief that the things we do come closest to some notion of "working in the public interest." Third, and the key element in this message, is a call for some real introspection on our work. In simplest terms, I think there is some rebalancing needed if we are to make it work, that is, if we are to have a substantial positive impact on outcomes.

In my opinion, the right way to phrase this is "what would it take for policy analysis to become the fifth estate?"

The fourth estate is built on the idea that providing raw information is the way to improve the operation of society. But I currently see little shortage of raw information. Information is plentiful. Witness the capitalized value of Google: \$90B or some nine times that of the New York Times. Indeed by the publication date of this talk, the disparity is likely to be even larger.

What is in short supply? First, it is making sure the right question gets asked. Second, it is organizing information in meaningful ways. In other words, the thing in short supply is analysis. Interpretation not raw information is the deficit item. And that is commodity needed to "keep system moving."

Government is clearly attempting to do more things in, arguably, a more complicated world. What are some of the things surging through the "policy" headlines of the last year? Stem cell research. Ubiquitous and ever-changing communications technology. A military deploying drones, advanced electronic counter measures, and facing sophisticated suicide bombers. Hybrid cars that may or may not reduce fuel consumption.

This partial list makes a clear case that policy cannot be made "on the fly." Neither can reporting of these items by the fourth estate lead us systematically to good policies. It may even be possible to make the case that normal reporting is as likely to lead us in incorrect directions.

Let me return to education, because there the examples and illustrations are plentiful. To begin with, everybody is expert. (Documenting this is easy for somebody who has questions about the efficacy of class size reduction). It is also an area where self-interest is pervasive and where there is a clear need for balance. Match legislators and campaign contributions or school board elections and union support. The Fall 2005 referendum in California on teacher tenure allows some quantification of the issue. Voters were asked whether the tenure decision for K-12 teachers should be moved from the current *two* years to five years. In three months, the California Teacher Association raised over \$50m for an advertising blitz that was sufficient to kill the initiative. The news media reported the spending (as if anybody who watched television could miss it). The substantive stories on the referendum centered on the potential impact of employment uncertainty on exacerbating the impending teacher shortage and harming schools through reducing teacher quality – a bizarre mixture of fact, fancy, and self-interested thoughts planted by the anti-referendum advertising campaign.

It is nonetheless necessary to go deeper into the issue. My explanation for why the politics play out as they do is that science has not been central to the education policy discussion. Parents, voters, and the media allow the self-interested discussions go unchecked precisely because there is no widely available evidence about what impact any actions would have on students. Parents, focusing on their children, might side with a scientific argument about the impact on student outcomes if it were available. Yet, even where solid evidence exists, it is easily overwhelmed by counterarguments based on thin or faulty evidence, given that there is no consensus on what constitutes good evidence.

Experimentation has been viewed as a per se evil idea in education. After all, the argument goes, do we really want to use our children as guinea pigs? This commonly held position of course ignores that fact that schools experiment on children all the time. They just do not in a nonsystematic way and seldom evaluate the experiments so that we can learn from them.

Nor is there replication of any program evaluations. Almost never is there an attempt to observe any program impacts in different settings. The prevailing view has been that, if something has been documented, there is no need to redo the work again.

Ultimately, however, the fundamental problem is the lack of metrics for evaluating research. This strong statement probably holds closest for education, where there is a large universe of people working on evaluations but lacking any firm underpinnings of methodology. I suspect, however, that it also applies to other policy areas, ones that have their own distinct historical development.

Education may be on the verge of a substantial structural change, largely as a result of the introduction of broader accountability through NCLB. To begin with, providing clear information about student performance obviously affects the debate on programs. Arbitrary choice of programs and policies produce disappointing outcome results that are now much more obvious. In education, it is often stated that it is necessary to guard against political decision making, such as principals showing favoritism toward some teachers on the basis of things unrelated to performance. But, favoritism that does not promote to student performance faces increased scrutiny when attention is focused on student performance and its measurement. It is not surprising that, for example, we do not worry about political decision making by managers in the steel industry, because – while possible – it is checked by information and accountability based on output and performance.

One other direct element of the legislation was a call for use of programs that were supported by scientific research. In related legislation, the restructuring of educational research under the Institute for Education Sciences spells out a vision of scientific research that places heavy emphasis on randomized experiments. Because of the newness of the idea of randomization in educational research, these efforts are still being refined and balanced against alternative approaches (and against a budget constraint that is a miniscule fraction of the NIH's budget). Nonetheless, they have changed the rules of debate.

An ancillary and unintended element of the standards and accountability movement of the 1990s and its expansion nationwide under NCLB the development of a system of regular assessments of student performance. The administrative data thus created make it possible to do rich new analyses of how teachers, schools, and programs affect student performance. It is now

possible to follow students as they move through different classrooms, schools, and districts and to disentangle the impact of each.

While the availability of these administrative data dramatically expand the research potential, the politics have not changed. Access to these data is not assured. In fact, if anything, the balance swings further toward those who want to limit access to data. In order to use the data effectively, the students must be assigned unique id’s so that they can be tracked over time. But, of course, the decision to develop id’s and to match students with programs is recognized as a move toward providing assessments that can be linked to schools, teachers, and programs, so the first defense – witnessed in California – is resistance to such a move that would make the data useful for research and evaluation. This political control of information is also made more difficult by federal laws on privacy and confidentiality of educational records (the Family Educational Records and Privacy Act, or FERPA). By some extreme interpretations of this, any use of accountability data for research would be effectively stopped. This on-going struggle has yet to be fully resolved.

My simple summary is that the same politics of data, research, and evaluation are present, but there are distinctly new opportunities. By themselves, these new forces seem unlikely to fix problems of the disjuncture of analysis and policies. Yet, they signal opportunities that, if capitalized on, could lead to radically different policy relationships.

Some Initial Ideas about Improving Our Impact

It is commonly stated that politicians just look for a study that supports their position. Without judging the prevalence of this, I think it is possible to find a number of examples supporting the assertion – but I believe this possibility is not independent of our actions. (I take the underlying assertion seriously even if it also appears just as a convenient excuse for why our work has so little apparent influence on policy).

We should accept the fact that we contribute to this situation, and move on. We can, and must, think about changing this.

I maintain an old fashioned, naïve view that the number one issue is providing good policy information and analysis. If we want to enter into the policy process effectively, we simply have to have something useful to say that would help the policy. But, good policy analysis is somewhat complex. We know that in general the standard two paragraphs on “policy

implications” tacked on to the end of every journal article would not be counted as useful information.

Part of generating useful policy information clearly involves relating what we have to say to the interests and perspectives of users. Again, bemoaning the fact that we have a different perspective from various users does not take us very far. Taking the perspectives of users into account, particularly when there are different perspectives, takes work. Most of us do not automatically know what different users are looking for, and we often tend to think that our job is research, not trying to figure out how various users might approach the policy process (particularly when their basic framework on the surface appears very different).

Past having useful policy information and analysis, we need to consider how to communicate it effectively. A first part of communication is providing the results and implications in terms that are understandable to users – not in the convenient jargon that we might use to communicate with our peers. We obviously cannot be upset to learn that policy makers and their staffs do not fully appreciate the nuances of technique and methodology that we have worked many years to understand and to perfect.

However, there is a more fundamental issue wrapped up in communication/information. An element of the communication process is that we are seldom the only researcher in an area – and the other researchers may have “better” answers from the perspective of particular policy users.

We have to communicate why our answers are superior – and that requires having clear standards for judging superiority. I am not sure that we currently have such clear standards that are broadly accepted. I participated in the National Research Council panel to develop scientific principles for educational research (Shavelson and Towne (2002)). This project was a necessary first step, but it is clear from reading the report that there are many different conceptions of what is scientific and little in the way of general guidance about what research may produce better answers.

If we do not have a clear set of standards and ways to judge quality, we cannot look down upon politicians who choose the answer they like from an array of possible answers by different researchers.

So let me return to some things that I believe could improve our ability to improve policies.

First, I think much more attention has to be given to *framing the right questions*. Policy makers typically look to research to inform (or support) their conception of the right policy question. While I indicated previously that we must deal within their framework if we are to move policies successfully, I think it is important to realize that the questions asked are endogenous. Researchers and policy analysts can influence the questions, and that is perhaps the largest and most important role they can assume.

Framing the questions differs from micro aspects of research design, specific articles, and like. Good researchers know how in specific research to define the question so that it is one they can answer with their theory and empirical analysis. In fact, the art of publication often involves camouflaging the weaknesses of a study and overstating the generalizations that can be made. Questions framed in this way for the specific purpose of developing a publishable article very seldom influence the way policy makers think about general issues, and rightfully so.

In educational research, I think of the impact that the original Coleman Report (Coleman et al. (1966)) had by pointing attention to student performance and away from just the inputs to schools. This framed the way subsequent people thought about schools and their impacts.

In general, individual articles cannot effectively frame the larger questions (although sets of articles combined may). In general, journals will not allocate much space for this. Instead, I see this more of a group activity – through conferences, NRC panels, and the like. And, I see a clear role for catalyst organizations such as APPAM to develop and support these activities.

Second, we need *clearer standards* of what constitutes good research and evaluation. The Coleman Report can be used to illustrate the second point except in a negative way. The inappropriate statistical methods of the Coleman Report led to a series of widely disseminated conclusions to the effect that schools have little impact on achievement. These conclusions, while clearly very wrong, survived and persisted because of confusion about the standards of such research and statistical analysis.

If we agree that “anything goes” in research, we will in fact find that anything goes in policy also. The place where this has most bite is when we consider issues of causality. Most research-based policy assumes that taking some action will lead to an achievement outcome that

matches a research finding. To be true, the research must have identified a causal effect, as opposed just to a correlation of the action and achievement in a study situation. Identifying causal effects, however, requires use of appropriate methods, ranging from random assignment experiments to a variety of other statistical approaches designed to emulate random assignment.

Indeed, we frequently hear of the problem of "going to scale." Specifically, a wide range of research findings from very specific evaluations are not reproduced when taken to another setting. My interpretation of this is simply that the original analysis typically did not really identify a causal effect but instead merely labeled a correlation with a specified factor that was related to what influenced achievement as the relevant policy factor. The easiest examples come from evaluation of specific programs. In the analytical sample, the program looks great, but then flops when transported elsewhere. Very often, I believe, the missing element is information about the quality of personnel in the analytical sample, who are not transported to the next site for application. We cannot turn away from these issues and hope for our analysis to be influential in improving student outcomes.

It is of course Pollyannaish to believe that all researchers are going to agree on the one best approach. We will continue to have academic squabbles about specific research and conclusions. That is healthy, and it undoubtedly leads to better research. For example, the attention over the past decade or two to sample selection in a wide range of social science research and to the issues of causality previously mentioned has involved a healthy debate that has led to better research.

Nevertheless, the debates about specific research findings and approaches often are small relative to the overall policy issues and framing of the questions, even if these debates tend to confuse policy makers. APPAM and other professional associations can in fact provide leadership on issues of standards. It is, I believe, possible to have better standards about what counts as good evidence and what does not count without resolving all of the micro questions of research within every policy area. In particularly important policy areas, we may even think of organized ways to evaluate and to summarize evidence for consumers. Again, if we cannot resolve these issues among ourselves, we can hardly expect policy makers to "make the right choice" when confronted with conflicting research.

Third, and perhaps more controversial within the academic community, we have to think more about the *incentives for policy involvement*. Universities contribute to some professional conflicts that tend to work against developing better public policies. The rewards in universities, with some real justification, go to professional publications. Publications have the possibility of a lasting impact, through changing the academic and policy debates in the longer run. They also clearly link ideas and work to individuals, thus creating the basis of individual rewards in the academic community.

The academic reward system does not provide much if any incentive for individuals to move beyond a journal article into the policy process. If on the other hand policy related research can have its impact only if it is effectively translated and communicated to policy makers and their staffs, there is an obvious problem. The incentive system itself points to relatively underappreciated and relatively ineffective research when judged from its impact on policy.

I realize the potential difficulties of entering into this area. Universities have endlessly debated the under-rewarding of teaching relative to research. But in teaching there is at least information from teaching evaluations, from indirect performance of students in follow-on courses, and from student research, theses, and placement. If we move to “policy interactions,” the measures of performance and outcomes become even more indirect and imprecise.

If, however, we believe that the proper outcome of policy related research is better policy, we are explicitly biasing the rewards against the appropriate outcomes and toward a series of potentially misleading proxies. There are of course other arguments towards weighting publications very heavily, but we must realize the biases against policy impact that are implied.

It is a delicate balance. As a long-serving department chairman, I was always reluctant to think of rewarding people for “schmoozing” as opposed to doing “real work.” Nonetheless, I think it is incumbent upon universities – and particularly policy programs – to think more about how to provide incentives for people to follow through on research by effectively introducing it into the policy process.

Can we do this through specialization? Having grown up on a diet of Adam Smith, I at one time thought that the best organization was to have some people who did the research and some people who transported and translated it to the policy process. If that were efficient, it

might be clearer about how to develop the incentive system. For example, if these different roles were naturally correlated with age – young people do the research and older, over-the-hill researchers do the communication – the existing tenure and pay policy may be alright. I am, however, no longer so sure about the possibilities of specialization. Indeed, I am increasingly skeptical about the ability to separate roles like this. Old masters may both have the stature and the wisdom to communicate with policy makers, but they may lack the detailed knowledge and understanding of the research that is needed to get it right. Moreover, this separation of activities reinforces existing incentives for answering more narrow and academic questions in the journals and for neglecting the tuning of research and policy.

Part of the difficulty in changing any incentives, however, arises from the structure of universities. Policy programs co-exist with other parts of the university. How does one convince the physical scientists that “soft activities” such as being more directly involved with the policy process are a legitimate thing to reward?

Currently, some efforts are made to take research to policy makers, but it is largely driven by the intrinsic motivation of researchers (which might be a fancy way of describing ego). I have little reason to believe that this is sufficient to draw out the optimum efforts of researchers.

My own view is that these issues of standards and of rewards are ones that individual universities should take up but moreover that APPAM as an organization should take up. APPAM has now achieved sufficient standing as a professional organization (even if a little suspect because of its multidisciplinary nature) that it can now take on some of the larger professional issues and be influential in the discussion. The institutional members should feel that this is in their interest.

Finally, I note that these considerations point to a classic externality problem. Individual researchers have too little incentive to push general policy development. Moreover, funding organizations also appear too little interested in these issues. Many foundations, for example, do not want to fund infrastructure or basic knowledge, preferring instead to concentrate on funding specific research projects at their marginal costs. The perspective here suggests that such an approach is quite myopic. This approach bypasses addressing the fundamental issues considered here. Yet clearly bringing about some of the more basic discussions suggested here needs the

catalytic role that foundations with their interest in multiplying the value of their investments can undertake.

Conclusions

The starting point of this talk was a simple question: If our research has gotten so much better, why hasn't policy improved? My answer is rather simply that policy has not improved because we have not had a sufficiently large role.

I think policy would be noticeably better if indeed we became the *fifth estate*. In that role we would provide the framing and interpretation of research necessary to the development of good policy. The fourth estate is simply incapable of doing this, particularly as more and more policy decisions are dependent upon technical information.

Establishing our position as the fifth estate will not, however, simply happen. It will take concerted effort on our part. We cannot simply sit back and presume that the beauty of our research will be apparent to everybody. We must think more about how our work can be useful for the policy process, and this effort undoubtedly includes clearer standards for judging the quality of evidence. Finally, if we are interested in efforts to improve policy, we have to be clearer about the incentives for participation. This will likely require some new thinking by policy institutions and universities.

I personally think this effort is worth it. At the same time, it is clear that this is not a freebie. It will require substantial effort. It is much easier to bemoan the biases and distortions of the policy world than it is to see the world from their vantage point and to figure out how we can be more effective in that world.

This brings me to a concluding question (perhaps a question for next Spencer Lecturer, who might have some empirical data). Can we do these things while remaining true to our analytical principles? In other words, can we retain our analytical soul while amplifying our impact?

References

- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Shavelson, Richard J., and Lisa Towne, eds. 2002. *Scientific Research in Education*. Washington, DC: National Academy Press.