Special APPAM Teaching Workshop:
Teaching Policy Analysis and Management in Cross-National Settings
Session B: "Teaching the Skills of Policy Research and Evaluation."
November 10, 2007

Edited transcript of remarks prepared by Doug Besharov and Justus Myers at the University of Maryland/American Enterprise Institute and Erik Devereux at the APPAM office from a transcription of the session tape. Because some remarks were made at a distance from the tape recorder or without use of amplification, those that could not be discerned on the tape have been deleted. Certain remarks have been edited for clarity and flow in written format.

Citation: Association for Public Policy Analysis and Management, "Workshop: Teaching Policy Analysis and Management in Cross National Settings," November 10, 2007, available at http://www.appam.org/conferences/workshops/pdf/APPAMworkshop_20071110_SessionB.pdf.

Chair: Rebecca A. Maynard, University of Pennsylvania.

Panelists: Alberto Martini, Universitá del Piemonte Orientale (Italy); Christiane Spiel, University of Vienna (Austria); Richard Berk, University of Pennsylvania; Robert Walker, University of Oxford (United Kingdom).

Themes: *Research and evaluation are intended to help guide policy. What are the various research and evaluation tools used in different countries? How are they similar and how are they different? Can we learn from each other? What materials and teaching methods work best in cross-national and multi-national settings (lectures, Socratic method, and case method)?*

Rebecca Maynard: Okay, we are going to get started with the second session. I think this session will build nicely on the discussion we just had. We are going to move course just a little bit and talk about "Teaching the Skills of Policy Research and Evaluation." What is it that we need to teach students in order to be able to actually analytically address the questions that come up with reasonably solid answers?

Instead of going into a lot of detail about our speakers, what I have done is ask each of them to talk a little bit about their background and the context of what they are going to tell us about this issue of teaching skills and policy evaluation. So I'm just going to tell you a little bit about my humbling experience, sort of getting into this international teaching arena.

My first experience was about seven years ago when I came into a classroom of 60 students, about 30 of whom, much to my surprise, were public management students who came from working in developing countries. And I'm teaching what I always teach, which is program and policy evaluation and I know how to do this, right? Wrong. These students were so unnerved because I was picking up the culture of the U.S. where we have data, where we have that—you know, this issue of going to administrators talking to them. We have got research to build on. So that was my lesson number one.

Lesson number two was working on comparative education projects with six or eight other nations and thinking that we were going to come up with a single research product that will make sense across countries, and these are Asian countries, European countries, U.S. And, again, we have differences in methods, we have differences in data availability, we have differences in what is important for questions and contexts. Boy, what a lesson that was.
And my third experience was having a room this size full of Chinese doctoral students. My job is to teach them how to think about program and policy evaluation. Number one, they speak Chinese; I speak English. So I have a translator next door.

And then number two, we have not only that but we have all these other cultural issues and data issues that we talked about before, so very, very important and challenging issues. Yet, the fact that we are all having these experiences means that we have got to understand this better.

So we have four exceptional people here today who are going to help us think about this, share their experiences. The first speaker is Alberto Martini from the Universitá del Piemonte Orientale in Italy. We have Christiane Spiel from the University of Vienna in Austria, Robert Walker from Oxford University in the U.K. and Dick Berk from the University of Pennsylvania. So I'm going to just give the time to our expert panelists here. They will tell you about their experiences and their thoughts for where we go from here. So Alberto is first.

Alberto Martini: Good afternoon. Well, I spent half of my adult life in the United States working as a policy analyst, then the other half in Italy. So I grew up with split personality because they are really at the end of the spectrum in terms of teaching policy research and evaluation skills. There is a missing subtitle, which is "Mission Impossible." Seriously, the question is: Is it possible to teach policy research and evaluation, which I abbreviate as PRE, in a country that lacks the fundamentals? By fundamentals, I mean what I got used to in the United States. I have a very polarized view of this: either Italy, an intellectually deprived country or the United States which is the model. I admit I have this polarized view.

What I mean by "intellectually deprived country" is a country where there is no tradition of social science in government and I have told [people in government] this. Take the Italian Courts of Account, which should be the equivalent for the General Accountability Office - the top layer of the Courts of Accounts are not social scientists. You might think they are accountants; no, they are not accountants. So you might think they are lawyers; no, they are not even lawyers. They are judges. So this is funny—goes from not having a real tradition of social science in government.

There are no think-tanks. A policy research and evaluation organization tends to be small and weak; that is, again—the opposite is the United States. Public debate is dominated by ideology and petty politics. Yes, it is much worse than the United States. My friends, you do not know what you are talking about if you think that you have a low-quality policy debate. Try Italy. Try Italian television. Of course, you do not speak Italian, so you are never going to be exposed to that. [Laughter] But it is the worst and believe me, I understand both.

So the answer is—it is possible to teach policy research and evaluation skills in Italy? The answer is yes, it is possible. But the real question is: Is it worth it? We need more research on this question. [Laughter then clapping of hands] Now, ignoring the absence of the three basic ingredients, in the last 10 years there have been some attempts to teach a U.S. style policy research and evaluation skills, and I have been part of these attempts. I have been back in Italy for the last nine years.

We have a Master in Analysis. The four main focuses or foci for the curriculum are the traditional U.S. style: policy design agenda-setting, problem definition, performance management process analysis, and impact analysis. So these are the four pillars of the curriculum. It is a one-year curriculum.

We use a fairly traditional teaching style. I agree with what Sonja said before [in Session A of the Workshop] - it is difficult to use and find good cases in the Italian context and we are not used to that type of teaching style. We produce about a hundred fifty policy analysts - U.S. style policy analysts; about half do public policy-related work in a very broad sense.

There are other experiences, other master programs not so focused on policy analysis but more on public management like the University of Bocconi. Then another experience I have been involved in, which is small but significant, is the summer school in France for policy evaluation. We had it every year since 2002; it is a 10-day long course aimed not at undergraduate or graduate students but at the junior researchers from the government in the Policy Evaluation Organization; or the equivalent institutions. There is high demand for that. We had 80 applications for 20 positions and we formed—we trained about a hundred researchers.

The third experience I have been involved with is training courses in policy analysis and evaluation for the purpose of legislative oversight. This is a very peculiar experience. These courses are aimed at the staff of the state legislatures. Italy has regions which are fairly equivalent to U.S. states; it has 20 of them. Sicily is a region and is famous for inventing Mafia. Tuscany is another region. So they are very much like U.S. states and they have legislatures. We push them to do legislative oversight. In support of that, we taught endlessly courses in policy analysis and evaluation to their staff.

Now, the question is there is a real demand for oversight for politicians or are the staff driving that? Anyway, it has been an experience in what you call "capacity building and evaluation." Now, the question is: Why train new policy analysts if there is little demand for them? I have been asking this question myself for the last nine years.

There are two motivation hypotheses. One is the supply can stimulate its own demand. It is not really happening but [government personnel] certainly like the driving idea: Let us create the supply and it is going to stimulate the demand. The other theory for motivation is being prepared for a demand shock. If politicians realize one day they need it, we are prepared with the supply of policy analysts.

Now, there is an example I have been told; I do not know the story—it concerns Germany and the Hartz Labor Market Reform. The Hartz Reform was triggered by a scandal in the German equivalent to the U.S. Bureau of Labor Statistics. So there was a scandal in the way they were preparing an employment statistic; the Germans created reforms which, in turn, triggered a wave of evaluation. This demand “shock” for evaluation was not expected. Even Germans did not expect these—neither the Hartz reform nor the heavy investment in evaluation.

Now, I have a question for you: Wouldn’t a European APPAM with national chapters help in developing policy research and evaluation in Italy? And the question is in Europe. Of course, my thinking is European-centered. Now, one issue to think about—its focus should be on policy research rather than on evaluation for the following reasons - to avoid endless debates, and by the evaluation, fear is in paradigm wars. Americans like paradigm wars. But Europeans do, too, particularly second-hand paradigm wars. So we can avoid that.

The purpose will be to focus the attention on evidence-based policy and what works rather than accountability. Money spent on time - that is the European Union bias in talking about devaluation. They really think about accountability and not learning what works. And to open a debate on the standards of evidence - how to produce credible evidence. That was really lacking in the European Union. They remind you of APPAM’s mission. APPAM is dedicated to improving public policy and management by fostering excellence in research, analysis and education; there is no mention made explicit of evaluation.

Now, the question is: Should we reproduce the dualist APPAM-AEA where AEA is the American Evaluation Association? Is it worth it? European countries and Europe as a whole have already national evaluation associations. Should we reproduce the dualism, which is typical of the United States between APPAM and AEA? Is it fruitful or should we confront ourselves as policy analyst with evaluation community? Thank you.

Christiane Spiel: Welcome everybody. It is a pleasure for me to be here and I would like to thank Doug [Besharov] for inviting me to come. It is the first time that I attend the APPAM Conference and I would like to continue what Alberto said. I’m the President of the German Evaluation Society—now, not German because also the other German-speaking countries are included. So we dropped the term “German” and it is now Evaluation Society for the German-Speaking Countries or DeGEval. But I think, what we are doing in Germany is much more closer what APPAM is doing than, for example, what is the European Evaluation Society and, also, the American Evaluation Society is doing. I feel more in a similar environment here than when I attend the European Evaluation Society, for example. So maybe—

Alberto Martini: Me too.

Christiane Spiel: Oh, fine. So we agree. I would like to introduce myself a little bit. I’m a trained psychologist. My research topics are, on the one hand, in educational psychology but also in evaluation. That means when I speak now, I am mostly focused on evaluation; I am not trained in quality analysis and management.

We are doing research on topics which are very close to politics, that is, for example, on bullying. And I’m now preparing a general strategy for violence prevention in Austrian kindergarten schools; it was amended from the Federal Ministry of Education. And also some other research, and also projects with the ministries. And as I told you before, since four years, President DeGEval, and I am elected for additional two years.

I will present, very briefly, two topics. The first, I will present you some ideas on the DeGEval and then I will present you learning environment we developed to teach more efficiently research methods and evaluation at the University of Vienna.

Today, we have more than 90 million inhabitants which are German-speaking, but I think it is very important and similar also, as Alberto mentioned—there is no German equivalent for the English term “policy.” I think this is very important and so we also have no tradition like that in policy analysis and management.

Also, in 1980s to 1990s there were increased evaluation activities in various fields of application but there was a professional fragmentation. There was an increased need for professionalization in the 1997 so the DeGEval was founded. And one year before, the Evaluation Society in Switzerland was founded; that is very close together. The aims and work are similar to the APPAM, that is, professionalization, consolidation of different perspective, promotion of information and communication. We have also annual meetings and we have also a journal; this is the *German Journal of Evaluation*.

I think this is a very important topic of our work. We have also various working groups as you can see and an executive committee. But we try to publish booklets on standards for evaluation and, also, which is very important for us, guidelines for implementation of the standards in the fields of self-evaluation.

The problem is, for example, in the educational section, that going to school and have evaluation on teachers - assessment of teachers - is impossible in Germany and Austria and I think also similarities with the situation in Italy. That means we have to start with self-evaluation and it is impossible to do real evaluation.

We have also no real employment market for teachers because they were not selected from the schools, yes? This is important to know. We also published guidelines for education and training in evaluation, and guidelines for people who commission evaluation. These education and training guidelines should define in terms of a fundamental environment profile what knowledge and competence are needed and those are the five fields: theory and history, methodological competences, organization and knowledge, social and personal confidence, and evaluation practice.

But I will now finish this first part and the problem is that so far, there is only one Master on Evaluation established in the German-speaking countries. This is in Saarbruecken - University of Saarbruecken in the Saarland. And the second one is now started at the University of Bonn; that means no real training. But the problem is that most of the people who are working in the public administration are trained in social sciences.

I think it is also similar than in Italy but—and here we have the problem and this is the reason why we developed this Vienna E-Lecturing. The problem is that Vienna E-Lecturing is an internet-supported teaching concept that should lead to increased commitment amongst the students by guiding them to grasp the material at a deeper level. So far, it is implemented in a public administration course on research methodology and evaluation.

But the problem is that students in social sciences in Europe and, especially, in Germany and Austria are not very much interested in the topic. That mean they have anxiety in statistics and they try to avoid these courses.

So we developed this program with the idea, as I mentioned it before, that they should learn more about this methodology and more about evaluation. But we also have some other challenges. Because, you know, self-regulated learning, I think, is really a challenge for the future but the research on that shows that there are a lot of deficits.

Also, another challenge is working in teams and, also, the research in our country showed that there are also a lot of deficits to working very efficiently in teams. Efficiently—not to say I cannot work together with some other colleagues but working efficiently. Also, we think it is important to successfully use new media. These are the reasons we developed this concept with the idea to promote factual knowledge in research methodology and evaluation, learning competence, self-regulated learning competencies, cooperative learning and e-competence.

And it is based on a model of self-regulation. As you can see, starting with the four phases—that means starting with your goals, then coming to action and monitoring your action and finishing as self-reflection phase. That means you have also to judge your own results. Based on this model, we have four didactic principles. One is blended learning; that means we combine online courses, online sections with face-to-face units. The next concept is the interlocking of learning goals; that means we combine factual exercises with also self-regulated exercises. And the inclusion of various types of knowledge; that means we not only focus on this factual knowledge in the sense of declaration knowledge but also how to proceed with this knowledge and under what conditions you should apply it.

Last but not least, we try to focus on both explicit and implicit forms on imparting. That means to see—they have to study literature; they have to work on tasks that is explicit but implicit also; they have to discuss this in small groups and put it—the results—on the discussion boards. That means to have real checks on their learning platform and so on. We have also face-to-face units where the explicit parts are factual knowledge but more on a deeper level than usual courses, usual lectures.

The program structure is as follows: we are starting with a kick off event. Then we have so-called "meet the experts" units; that means this is not similar as a usual lecture. We do it in that way that the students have in advance to read the literature and during this lecture we discuss it on a more deeper level. Then we have tutorials to promote learning competence, teamwork and so on and, in addition, self-regulated learning units; this is very important. We have online modules; we have 10 modules for the whole year.

All of these modules are structured in the same way. That means it is not easy for you to see all of them structured. The first is what is the content of the module; the second is what are the tasks you have to do; then what is the literature—where you can find the literature. But the literature is not put on the platform. They have to read books; they have to read articles and so on. This is different and depends—and then, they had to work together in a group on tasks and to look whether they are able to cope with the content. They have self tasks and spend also self-regulated exercises.

We evaluated this concept and we used a matched control cohort because it was not possible to randomly assign the two groups. I will show you some of our results. If you look at the factual knowledge, we are much more interested—they come to a deeper level and here you can see, especially on this deeper level, they are real effective in both semesters. But we are also—that means you will see they are much better than before.

But we are also successful on our other goals; that means promotion of self-regulated learning competencies on collaborative learning and then e-competence. That means we hope that we are able to create a culture of evaluation. If you are interested to get more information, we published a lot about this concept. And also, we developed a new matching procedure on that and we also published this matching procedure. Thank you very much.

Robert Walker: APPAM provides a wonderful window on American policy and American research. It is why I have been coming to APPAM for 12 years. Throughout those 12 years, I have been persuading colleagues here to look back through that window on the rest of the world. So I'm very delighted that we have this opportunity after all that time to begin this process of looking both ways through the window.

I said earlier that I was here as a result of accident; it is not quite true. I was here because I already was here at APPAM, and Doug [Besharov] did not have resources to fly in any other experts. I fulfilled the main qualification, which was that I was not American, and that is why I'm here talking to you. So I'm not talking from a position of knowledge but a position of quite a degree of humility, I hope.

Oxford [University] does not teach public policy. Lots of researchers there have some interests in policy and they are currently having a quite heated discussion as to whether we should spend the £212 million endowment on setting up a public policy program in school. And most of the disciplines are fighting against it; it soils their hands, engages with the real world. And what we do in Oxford is most certainly not that.

So I do not know you whether you are half-competition but at the moment, we have money looking for a cause. I profess social policy and we run two master's programs a year - one-year programs, not two. I'm responsible for the one on comparative social policy; I may talk about the other one in a few moment's time. We take 60 students a year. Ten percent of those are British, 20 percent are American, and the rest are from, in fact, 38 different countries according to our records this year. Our big problem is not to shut the Americans up, although that is quite a challenge, but to get our American students to realize that the American way of doing things is not the only way of doing things. And that is a profound challenge.

Rebecca Maynard asked us to talk about what should be taught, how it is taught, and the strengths and limitations. Context is important and I wanted to talk, first of all, about context first. I obviously speak as a subject of Her Majesty the Queen, not as a citizen. There are no British citizens. But we do have a long history of research-informed policy, which goes back to the Royal Commissions.

Certainly, I have been chasing them back to the 19th century; they go further back but I have not been able to look at their records. Royal Statistical Society, 1832—it was given a Royal Countenance in 1887. We had researchers in the 19th century and much of our policy can be sourced to their work. I think of people like Seebohm Rowntree who investigated research; lots of other people who investigated public health and placed it on the agenda.

The Government Economic Service has been in existence since the 1930s, but we did have that demand shock that Alberto was talking about in 1997. It was a demand shock brought about by the incoming Labour administration. [The new Prime Minister] said to play with palettes and evaluate everything before they put it in place. Skeptics like I and many of my colleagues thought that was merely a device to prevent the opposition from casting the programs before they were elected. But in practice, they have sought to do precisely that—to evaluate policies before they put them in place. The demand shock has benefited quite a lot of organizations, MDRC being one.

Evaluation, which was [indiscernible] as Germany and the U.K. from the centrists, an incredibly centralized society obviously with a size of a medium-sized state. But we are run from the center. We have the elected dictatorship that was referred to in relation in Korea. We have been running with it for quite some time.

But things changed in 1976 when we had a sterling crisis. You won't remember, but it changed the whole nature of policy making in Britain. IMF came in, sniffed around; we got scared and we did things rather differently. Prior to that, we decided what we wanted. So if we wanted hospitals, we added up a number of hospitals we wanted, and planned on that basis. We did not ask how much they would cost.

From that shift, from volume to cost, came about a whole different approach. That approach was systematized in 1997—again, coincidentally, this happened before the Labour Party was elected —into a Green Book. A Green Book is a bible, a how-to-do-it, a recipe book, a cookery book, which every self-respecting British civil servant should be able to repeat. Typically, they repeat it backwards but they mentally repeat it forwards and go from beginning to end.

It is built around that acronym - ROAMEF—thinking about policy, lots of Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback. What that document does—and it is quite a significant thing and this is the before-after Labour. But it has been updated. It is on the website. And when the sites get posted you will be able to visit. Is it a distinguished appraisal, which is a process of assessing whether a policy should be put in place from evaluation. But it is approximately—it says just the same thing as appraisal except it is done at the beginning of the policy rather than at the end. And it is not putting it at the beginning as other parts of government say, putting it at the heart of the policy making process, which has shifted the basis of making policy in Britain.

We have had what you might call an experiment since 1997 in evidence-based policy built at the center. The fact that you had visiting APPAM this week, the Chief Social Research Officer, whose responsibility it is for running this program, is significant. Perhaps more significant is that from the end of December, have posters being abolished. It may be the end of the experiment in evidence-based policy making. Watch this space. Coincident and alongside this movement towards evidence-based policy, we need to think about how we train. I'm making a distinction between training by government because that is where the demand is and that is who does it and training in a moment in academe.

We have a National School of Governance and I just looked on their website and pulled out those causes. That cause is aimed at senior civil servants while middle-ranking civil servants—both analysts, but much more particularly, the policy makers. Over the last 10 years, the analysts and the policy makers have been bedded out together, so that you now have analysts working, sitting on the shoulders of the people shaping the policy.

Alongside that, we have this online resource, the Treasury's Policy Hub, where we have the Green Book as I said, and I have described that, but also the Magenta Book. On the Magenta Book—we will come back to the colors of the rainbow. I will explain the boring story if you want me to later. We have the Magenta Book which is about perspective policy evaluation.

The National Science Foundation funds most of the social science research. It set up a Center for Evidence Based Policy and Practice at one of the colleges in London. They in turn have established a network and recently, I think a year and a half ago, a new journal *Evidence and Policy*. ESRC has also set up another center called an epicenter which is about building links, explicitly building links between the research community and government, making, essentially, a portal, a portal which has lots of worked examples of policy evaluation.

And they also provide the only NSC that I can find in the UK which has a—sort of tighter around that evidence—public policy and practice. It contains quite a lot of evaluation studies in there because I was one of the people who were consulted on when they were putting it together. I actually was not very complimentary about it and I have not looked whether my comments have been taken abroad—probably not.

And finally, there is a center of a micro data methods and practice, which every now and again, addresses its—a mind to the process of evaluation and spending their expertise. And here, we have a three-day course and to respond to Rebecca's—her question about what the contents should be, there is a content of the course about evaluation as perceived by British economists who are trying in a sense to spread the word. And in many respects, those sorts of divisions may well be particularly familiar to you all.

It is very—and I just mentioned the UK Evaluation Society, not the European; there is the UK one. We have our national ones. This was founded, I think, in the very late 1980s; it has a journal called Evaluation. And I mentioned the disciplines; I've already indicated what the disciplines think of policy evaluation in Oxford. I think that pretty well-shared. It is quite difficult to know precisely where it is that evaluation and policy analysis might be taught. Social policy—my discipline does a bit.

Public policy—well, that rather comes out of politics and public administration. It is not quite the same sort of thing you might think of in terms of public policy. The economists were again—again, government is trade and dirty. It is not the sort of the thing that a self-respecting economist would want to do and they certainly would not get published in the top journals. And since top journals are the only thing that British economists are worried about, not many people engage with the government process and thinking about policy evaluation.

So where do I go for the course? Well, I went for the course that I set up when I was at Nottingham. Talking about internationalism, University of Nottingham is a moderate sized university center of England. It has four campuses. It has two in Nottingham; it has one in China, and it has one in Malaysia. And essentially, it is setting up a franchise—and franchise is the clock tower which was originally built on tobacco money in the center of Nottingham. Nice white marble. You can now find copies of it as I have indicated in Malaysia and China and the list is growing; negotiations are underway—this process of globalization.

Doctorates in Britain are very different from yours. We basically teach people to go away for three years and think. They end up knowing a great deal about very little. We try to break the mold of that, taking people after their Masters, putting them into a Ph.D. program or a doctoral program which add another year which was thinking about methods and was explicitly engaging them as a vocational degree to enable them to respond to the demand that was coming through the center for people who can engage with the policy process from a research framework.

You can see the sorts of content of the structure. Just skip through all of this. I'm not going to talk about the policy analysis module because I do not know anything about it. I will talk about the advanced quantitative analysis; we use "advanced" because it sounds good not because it is. As I go back—oops, I cannot go backwards. Sorry, I wanted to draw your attention to on that. Was it the statistics? The quantitative component was focused very much on evaluation. And so that in a sense differs.

The design of it—basically, it was built around a group project which actually led not to publications but to a piece of funded research in the ESRC project which is actually looking at the multidimensionality of poverty using structure equation modeling. So that came out from that particular year, led to a £150,000 research contract. And that was complemented with a set of group seminars which will help equipping the students with the techniques to undertake that piece of research.

The model was thinking about your evaluation question for the perspective of when in time you ask it. It seems to me that the evaluation question and the consequent methodology changes. Hence, the questions run from what worked to what will work; alongside that, a range of different methodologies all of which were taught. And you will note the bottom one is only sort of a small subsection which is a section about program evaluation—was embedded within this much broad sets [sounds like] of research evaluative techniques.

We have taught research management tech skills and techniques which—not very often spoken and taught about it in British universities. But the major reason why so much of British research government money has been wasted—because the research has not actually delivered the project that was expected. It is partly designed. It is much more often the botched management of those processes.

We also taught qualitative techniques; picking up again on the European model, it is the inability of most Europeans to be able to count and certainly not wanting to be seen to count. I guess it is much more focused on process informative work and we integrated the research management with the qualitative techniques by running a research project where we put the two things together. It was a piece of work which was a policy evaluation in terms of government policies for higher education. And essentially, it was an attempt to make the research real by giving the students the opportunity to publish which indeed is what they did do. Saw the project down the middle.

I would just skip on. I was going to talk about the Oxford evidence-based and ASEAN social intervention. The reason that I'm not going to talk about it is I do not know a great deal about it, but I was persuaded to tell you about it because it is perhaps relevant. It is one intriguing difference from much of what I have had and of what I have learned here in the States is a much greater emphasis on meta-analysis [sounds like], on systematic review, of using the results of evaluations creatively to connect with policy decisions. And that is I think its unique twist.

It leads to real risk evaluations. That is one piece of project which was looking at effectiveness of an anti-AIDS program in South Africa. So it generates again real research. And I reached my conclusion—well, a conclusion of course, to the questions that I was asked to talk about at the beginning. Why on earth should you expect me to be able to answer them since my only qualification is not to be American?

Richard Berk: I'm going to be brief. I'm anxious to hear what your reactions to our thoughts might be. I have some overheads but the advantage of going last is that I can alter what I was going to say, to maybe respond to some things that were not mentioned already, or perhaps, need to be tweaked a little bit.

I was at UCLA for a number of years teaching in the Statistics Department and I taught people who are going to be statisticians. And so I bring a particular perspective on it, very different from econometrics that we are now using to work with social science students in program evaluation. I should mention that the vast majority of my students over those 15 years were not American. About half were from various Asian countries; we had a lot of Central Europeans. I'm used to being a minority in my own classroom; I do not see whole lot of problems with that and the students did not seem too, either. But maybe that is because of the subject matter.

I did notice, however, one important difference in the way the students approach subject matter and it harkens back to some things we have heard earlier, perhaps, with a little different spin. I found that the Central Europeans and the Asian students were much more top-down on how they learned statistics. You taught first principles, you proved theorems then they start worrying about what you used it for. I found the Americans could learn those first principles, in part, very well but they had to work bottom-up from case study material. They understood the context and then were able to do the math that was required.

I'm in the process of trying three things that are a little bit different. I'm at the University of Pennsylvania now in the Criminology Department as well as the Statistics Department. And right now, almost as we speak, I'm in the middle of a course, teaching a program evaluation on related topics to students who are basically social science-trained. I have some lawyers; I have some journalists, but the focus is on criminal justice evidence-based work.

I want to just talk briefly about three things that we are doing, three techniques that I think had some promise; perhaps, they are old hat to you folks and that is why I'm going to be brief. The first is that—I mentioned already. I find greater success with American students, but I should stress now it also seems to work pretty well for many of the foreign students in my class, teaching statistics in a case based way. So for example, if you want to teach about sampling, we could certainly start from first principles.

In fact, what I do, though, is I take a case study; for example, human trafficking where a key issue is defining a population as you all probably appreciate, and then figuring out how to draw appropriate samples to infer the number of such folks from which one could then establish baselines and do evaluations. So within a couple of weeks, we were doing things like recapture sampling. The kids got it really quickly in that context. I do not think they would have gotten it if I taught it from first principles.

On the matter of teaching internal validity, we looked at, for example, the war crimes trials based on activities in Bosnia and tried to figure out what fraction of the deaths one could attribute to NATO bombing versus ethnic cleansing. It was kind of gruesome topics but it certainly drew them to the material and we could address all the matters of counterfactuals and other issues you would discuss in internal validity from the bottom-up. And it seemed to work pretty well.

And just as a final example, if one is interested in construct validity, we are looking now at racial profiling where a key issue in that is how you measure ethnicity or race in a place like the United States where such a large fraction of the population are of mixed ethnic background if you even choose to define things that way. Again, it was working from case studies up to the technical stuff and, as necessary, up to the more mathematical.

Second thing that I'm trying which I think we have some success with is much, much greater use of simulation—many of you probably know that there is now readily available software that is very easy to use. And I use the simulation in several different ways. The first is I use it to characterize process. You can present students with a cross-sectional database and it is kind of boring and flat and it does not move; you can take the same phenomenon and simulate it and students then have a better sense for the complexity of what they are trying to analyze—the data are in some sense overly artificial. Even though that is what we work from, it does not really characterize what is going on underneath.

And one interesting example, for instance, there is some fascinating work by an anthropologist on crime who looks at auto theft in the same way that he studies hunters and gatherers in primitive societies. And he has an agent-based model in which prospective auto thieves and police officers move around the space with certain kind of simple rules, and he is able to reproduce interesting patterns of crime. Now, who knows if that is really the way it works? But it does give the students a sense of the complexity, the dynamics of what they are trying to study if they are going to try to evaluate the impact on such a process.

Similarly, we have been fooling around with various definitions of crime; what we call crime regimes. Crimes are, of course, are multidimensional; you can see how crimes move in multidimensional space. It is not just the number of drug transactions; it is also the associated violence that is with it and whether they are indoors or outdoors. And all these things can be done in the simulation and you can show the student the complexity of what they are studying before they get to the statistical analyses that is going to try to extract the essentials.

Another place where we use simulation is on the property of the estimates; this is old stuff in statistics, of course, but we do things that I think are a little bit interesting beyond what has been done in the past. First, we often show bad practice. So you all know that if you fish through data long enough, you will find something. So you can set up a situation in which you ask them to run some regression analysis like stepwise 50 or 60 times. You do not tell them when you give them the data that everything is noise. And of course, they will always find statistically significant results. Then you ask them to interpret it and, by God, they can; they are very creative. And of course, they are interpreting noise. Then, you reveal the truth; you unmask the data. And it is a humbling experience and as a good warning about some of the problems one can get into in data snooping.

We also will simulate various kinds of other bad practices which they can then see themselves getting into trouble on. So the point of the simulations is not just to do the usual properties of estimators but to consider other things, for example, the role of chance. I just gave one example with data snooping but I can simulate crime with a simple Poisson process which has spikes and, by God, they interpret it. And then we do an intervention and, of course, the spike drops and they are prepared to assume that, by God, we have got an effect. And of course, that is just the progressions of the mean and you can show them that in a very kind of explicit way.

We also use it for what-if applications. I'm sure you are all familiar with this. You have a particular set of results; you set up a model; you can simulate the future. That turns out to be real instructive because when you simulate, even if the simulation is not valid - these are after all didactic exercises – it's humbling. And the idea is that they had—that they were sure where correct turned out to be not so good when you simulate what would happen if you intervene. So that is the second tool we are using.

The third tool is dealing with statistics meta-issues. This has already come up in a lot of conversation so I will be even more brief. A lot of the important work that gets done in program evaluation is sort of above the level of the nuts-and-bolts of actual data analysis. And things like, for example, keeping track of the kinds of transformations you do so you can backtrack—and those are available to other folks; keeping things simple and so on; being able to resist the temptations of a client who wants results in a particular way.

These are important topics that very rarely get taught, at least in the American scene, at least in program evaluation. And we do this by role playing. So we will have a circumstance where we will have one of the students being interviewed to undertake an evaluation. Maybe I will be the bad guy; I will be the client, or some of the other students. And we actually go through—we script it to some degree; we actually go through the kinds of discussions which would unfold in that circumstance to help students become practiced at maintaining at least some level of integrity.

Now, of course, it is really difficult to do that as we all know. There are inevitable compromises; a student says to me, "Oh, if I do not do it I will not get it published." Well, it is easy for me; I have tenure. If you do not have tenure or if you were in a firm where your income depends upon getting this contract, it is hard for me to tell you categorically what to do. But at least what we can do in the role playing is bring these explicit tensions to the surface and then have the students talk about them.

I find this works remarkably well with students from different backgrounds, from different countries, from different experience. We have in our class right now several individuals from the Japanese police agencies and they are working perfectly well with people from Taiwan, with people from Central Europe who come with very other different experiences. So the role playing seems to work quite well - I think - and I would recommend it to you as something you might want to try for getting at these more thorny issues that you cannot kind of present theorem-proof-style.

So that is really all I have to say. I'm anxious to hear your remarks. If you have questions, I can certainly elaborate on this, but we have had a lot to chew on and we will turn to the questions.

Rebecca Maynard: Thanks. I want to thank our presenters. I think we have had a lot of provocative statements up here. Let's open it up to questions. Yes?

[Unidentified speaker]: Hi. Yes, yes. I do not think we did background biographies for the panel. You [Richard Berk] also edit a journal called *Evaluation Review*, and many of the articles now are from non-U.S. sources?

Richard Berk: Well, it is interesting. I appreciate the advertisement. Send us your papers. We do program evaluation. We are empirically oriented so we want to see data—the results. And we really like to see results from around the world. And so about a quarter of our publications now are from outside the United States and I would like to get it to over a half.

[Unidentified speaker]: And could you describe what those—the submissions are like—the successful ones and the unsuccessful?

Richard Berk: Well, as I said, we tend to be much more empirical than some other journals. We do not really want to seek topologies or proscriptive things: Here is what you are supposed to do. We want to see where the rubber hits the road: Show me data, show me results but in the context of a particular national or international local scene that you are working in, so the successful publications are generally empirically based. We also are pretty flexible about the level of technical sophistication. We want the tools done with the simplest means possible. You do not have to trot out some fancy structural equations if they are not necessary. We do not care;

in fact, we do not like it—if you need those kinds of tools, by all means. What this often means is that, often, from the developing countries where the computer expertise or the skill building has not yet reached the level, let’s say, of Great Britain and the United States, we get very, very good papers which are nevertheless not fancy by American statistical standards, and we love them.

Rebecca Maynard: Peter?

Peter Reuter: Dr. Alberto Martini was amusingly pessimistic about demand for evaluation in Italy and I think we also have a sense that politics is perhaps more factionalized and chaotic. I was curious that Christiane would say something about the demand for evaluation in Germany.

Christiane Spiel: I think, unfortunately, we also have not [demand for] an evidence-based policy, neither in Germany nor in Austria. This year we had the first conference of the German speaking countries. That means Switzerland, the German part of Switzerland, Germany, and Austria. And we compared what is going on in the different political fields. And it is very interesting that in Switzerland—Switzerland is completely different to Germany and to Austria.

In Switzerland, they have in the basic law an article that says all decisions of the Parliament have to be proven for their effectiveness. And as a consequence, not only the politicians but also the general public is much more aware of evaluation. And if you compare in Germany, there are 80 million inhabitants and in Switzerland there are 7 million. But the numbers of the—if you look at the members of the two evaluation professional societyies, I think, they are the same. And I think that means that the Swiss are much more aware of this problem.

And also this year, in advance of this conference, I sent out a letter to all of the federal ministries in Germany and in Austria and I asked them how many evaluations are conducted, what they are doing with the results of their evaluation, and how evaluations are organized, and so on. And from a lot of the ministries there was no answer and if there was an answer it was clear evaluation is not really well-established. And in some of the ministries, the answer was, in some cases, “We are working on that.” Really, “We are working on that and we plan to do something.” And I’m now conducting additional studies; starting Austria where we are conducting interviews with the persons who were responsible within the ministries. And also to find out how they define evaluation and how evaluation politics is really organized—very difficult.

Rebecca Maynard: Yes?

Laurel McFarland: I’m interested in the fact that in the original statement here, it talks about multinational settings. But I, actually, in the first questions in the second session have not heard people talk a lot about preparing students for jobs in multinational settings—policy analysts who will be working in multilaterals; policy analysts who will be working for NGOs who are policy researchers and evaluators who will be working at multinational settings. I would like to hear some comments.

Richard Berk: Well, I'm the tool guy here so for me the answer is pretty straightforward. If you believe that there are some fundamental tools that are required, regardless of setting, admitting that it may have to be tweaked substantially to work—the students that I worked with, many of them wind up in multinational settings and seemed to—like the World Bank—and do just fine as long as we are talking about the tools.

Rebecca Maynard: Next question?

[Unidentified speaker]: I was actually going to ask Robert to expand on the comment he made about MDRC because it is not just MDRC that benefits— it is the number of American firms.

Robert Walker: Talking about the last question, though, the multiple settings? Because the European Union is a classic situation where policy is negotiated across nation states and where evaluations are, in theory, meant to be conducted also across multiple states. And we have a series of programs in place which are motored by a device called the open method of coordination, which is basically a beauty contest of countries where we sign up to a common goal. We leave the ways of reaching that goal to the individual members but we evaluate - and I use that word with great caution - the success of those sets of policies and, subsequently, exchange in a process of peer review the mechanisms for good practice.

Now, that is wonderful on paper. But the practice of doing it is incredibly difficult, not only at the political level but also at the technical level. Because in a sense, what you are trying to do is evaluate packages of policies and at the moment there is a subterranean process going on which Alberto and I have engaged in, which is in a sense beginning to try and raise up quality methodological issues and to try and plant that into a process which is spending many, many millions of euros trying to engage in a process of common social policy development based upon evidence but where the evidence is incredibly weak.

The MDRC, the mathematical-type stuff—in '97, we, to my knowledge, have conducted about three random assignments—experiments of which I picked up one of them and I'm deeply embarrassed every time it gets mentioned to me. Subsequently, since '97, I sat on a government working party. We got up to about 220 prototypes or pilots of policies that would be nationally implemented.

And of those, I think four or five were random assignment programs. Most of them were—no, not most of them. Probably about 40 percent—possibly, 50 percent of those were quasi-experimental. And it was, in a sense—in '97 it was nowhere near the capacity within British social science to deal with that demand. And part of the consequence of that is that a significant proportion of those pieces of evaluation were very, very badly done.

Rebecca Maynard: Michael?

Michael O'Hare: So almost all of the technology that you guys are describing in your programs is classical inference statistics. I would like you to re-work the course relevance of statistical inference in terms of two approaches – inferential and Bayesian. One is a very big problem, we have no data. And really, you do not want to find another paper, right? There may be no more data in the other case; and yet, you have this approach and this tends to drive a wedge between the analyst who is trying to understand and the administrators who are trying to manage.

The second one, I feel that one of the elements that may be different is the way [indiscernible] of the world, so you could imagine designing policy such that it was to constantly generate data but with that policy operating in a small and passive way rather than having a way for a big study that takes a year to do and by then somebody else is on this. To what extent will these ideas succeed?

Rebecca Maynard: Do you want—?

Richard Berk: Yeah, I will answer it. Ninety percent of what we do is not statistical inference because it is very rare to have the requisite data to do that. So we are mostly involved in designing really good research designs, getting good measures of what happened at impact. And then at the very end, yeah, you might worry about some inference. So I do not want to leave the impression that, primarily, we would use inference.

As for Bayesian stuff, I love Bayesian stuff when it works and we use it when it works. As you know probably very, very well, there is a lot of controversy about whether that is the most effective way to go. One thing I like about it enormously is it does allow the revisions that you described. I wish it were more effective at building in the costs of mistakes. So you get the posterior but no one has built in the costs of being wrong. You can do that but it is much harder.

So I think the bottom line is Bayesian stuff is great. It is not perfect; there are things that we can do to teach it better and to use it better. And I think it is one of the many tools that we want to convey but it is only one, and statistical inference is only one small piece as well. It certainly is not excluded.

Rebecca Maynard: In my experience in developing countries, in particular, there is a lot of use of causal inference in the roll-up of new policies and practices. So in fact, what is happening in many places is that governments are making decisions to roll out a program in a staged way where they're randomly selecting the order of introduction of the intervention which allows them to get causally valid data on the effectiveness of a program, to get information that would allow them to tweak it as they go and make modifications. So that is particularly quite popular in developing countries where there are a lot of villages and resources are strained.

So I had a question way in the back that I was waiting on and then I'll come back.

Alberto Martini: I have an idea for an exercise for your students in the program evaluation class. It is parallel to what Richard Berk said about generating random data and let students analyze that. Take the European Union methodological documents and let your student comment on them, then you reveal to them that it is lot of bull#@%$. They may take it seriously; it is in English.

Christiane Spiel: I would like also to say that I fully agree with you, Alberto. I think we have on the one hand too many poor evaluations but not enough good evaluations. And we have also have some fields with many evaluations; for example, now, in the German speaking countries—but not only in them—they have so many evaluations in the university sector, too many. That means we have so much to do to working on that paper that we have not enough time for research and teaching. But in other fields there are not enough evaluations; and in any case, they are too many and results are not really used in a good way. That is the problem.

[Unidentified speaker]: This is a workshop. In the workshop, it hinges quite on the extent that there is a policy analysis being done throughout the world because by training more people involved in policy analysis as waspointed out here, or for Americans, there are no policy analyses out there to look at in many national settings. Well, we have heard some examples—pessimistic or how limited their—but I cannot believe that is true that this is a representative sample of the world because if you would back up this huge demand for our students around the world to be taking these courses.

So, stepping back from their own countries a little bit, looking at the world and just saying what types of policy analysis is being used at various parts of the world? And why are we seeing this demand in the surge of students if it is of very limited use?

Doug Besharov: I do not share quite the pessimism but I'm also not as knowledgeable as my colleagues here about what goes on in Europe. It is interesting that there is a whole sector that has been neglected, and that is the private sector. I cannot keep my students through the Ph.D. because they get bought out from under me at the Master's level to go to pharmaceutical firms, public opinion firms, MDRC, and others; it is an enormous demand. Literally, we cannot keep the students long enough because they get paid more than assistant professors when they just have a Master's. I do not know if that is true in Europe. I would be curious to hear if the private sector in Europe also has that sort of demand. But in the United States it is enormous.

Robert Walker: I'm not sure the private sector has quite that level of demand, but there is a lot policy analysis goes on. It is all around the quality of that and which is in the sense why the organizations are being set up to boost the quality. There is in a sense a role which the experience that APPAM—I mean, in terms of giving momentum to those improvements.

Equally, I think there are areas of policy-relevant work where APPAM members can learn from things which go on elsewhere. My view is that some of the formative evaluations of the experimental work that is always—my collaborative [sounds like] experimental work—experimental economics—is better developed in other parts of the world than it is in the U.S. So I think it is a two-way street.

The reason for the demand—well, it is a rather trite sort of observation but economic development is not inconsequential. What you are finding is that the policy language and the economic social development language are coming together as countries' economies generate. And those countries' economies call and create the sorts of needs which traditional welfare states have been created to meet.

And in that context, anybody that has got evaluation on their CV is now on the calling list of vast parts of the developing world—and it is—one is constantly being invited and they now are open for expertise. They have got—and beginning to formulate the questions which are bread and butter to your business. But they have yet got the experience upon which to build in a homegrown setting. And so they are happy to do a cultural leap to take your techniques and see how they might map in into sub-Saharan Africa for example.

Rebecca Maynard: Okay, we have time for one or two more questions.

[Unidentified speaker]: Yeah. It is sort of a question and a comment, but I have been listening to a number of these things I guess I was struck by the fact that on one hand you are putting it up as a set of tools which precisely because they are derived from mathematics, statistics and so forth are somewhat universal. But as you get them into play with policy analysis you need sort of a set of supporting institutions. You need governments generating data or you have to deal with the fact that you do not have that. And to some extent, it is just like the U.S. or in the U.K. you are relatively favored in that resource. I mean you can actually teach these things and expect to do it, but that I think in other contexts we have got to figure out what works.

But they could confront the problems with what data—very poor administrative data with the usual defects. And maybe that is really a subject for another session, but it is almost what else can you think of ways of putting value to the classroom, too, so that we are not assuming that students are going to have access to the whole infrastructure, I guess, that I take for granted, for example, like policy analysis here.

Rebecca Maynard: I take your point. I would say that it is—in my experience this point I would even be, what countries have what kinds of data. Some quite poor countries sometimes have certain kinds of data that are really very good and surprisingly good. But your point is well taken.

Christiane Spiel: We have, I think, the problem that people in the public administrations are not trained in public policy. And for that reason, we do not have this data because the governments do not commission such evaluation to produce such data and this is, I think, the real big problem in many European countries. And we have to work on that that they were much more trained. And for that reason, we publish guidelines for people who commission evaluation; guidelines to make them aware what they have to look on, what are good evaluations.

Richard Berk: Alright. I just got a little reprieve from our chair here, and they are teaching a method for it in evaluation for its policy students. I mean, ideally, I think prepared them to be able to go in and to be in such institutions as the Department of Justice; then you have the Department of Labor, Health and Human Services. In different context there has been a reevaluation of the work that has been prepared by economists, by criminologists, by psychologists, by epidemiologists, by political scientists. Each of those disciplines has some tradition of the statistics of evaluation research; they give different names to the same things, so on and so forth to bring up students. My sense is that many students, if they are trained, pick another way up with a lot of difficulty. I have a lot of difficulty with other ways. I can prepare them in the context of policy tools to be able to do evaluation and understand other evaluations.

Rebecca Maynard: That is a very good point. We did not focus on the training to be a good consumer of evaluation. So that is an excellent point. Anybody want to comment?

Robert Walker: Well, social policy in the UK has no disciplinary base. It pulls ideas from any discipline that happens to be around and that is what we actually teach. It is dangerous but that is how it—so in a sense, the issue does not emerge from the social policy-type people. It is much more problematic than the disciplines looking out and we, therefore, lose a great deal of insight because there is a lack of exchange between those people engaging the policy world and the disciplines per se.

[Unidentified speaker]: My question is to the three case studies from the European countries. Where are these sort of graduates going? What are the sort of destinations of the graduates of the policy institutions? Can you answer this question with three different aspects? One—what are the particularly different career routes compared to regular economics or politics and other departments at your institutions? Two—what are the sort of idealistic models of careers for the institution that you are aiming to change? And third—in reality where are the actual majority of the graduates being employed?

Rebecca Maynard: Alberto, you want to start on—?

Alberto Martini: Where are the graduates going? Mostly in consulting because the government in Italy does not absorb other than people with law degrees. So it is a big barrier we have not been able to break. So it is mostly consulting for the public sector but not within the public sector. We have not broken that barrier.

Rebecca Maynard: Robert?

Robert Walker: In terms of the courses that I have talked about, I do not know the answer. In terms of my own course, which is comparative social policy, most of them are in NGOs and government departments in their country of origin. And they returned to pursue careers there. So in a way, it is basically a career training that they are getting.

Rebecca Maynard: So the foreign students that you are training in—?

Robert Walker: In terms of comparative social policy, they applied social research. It was very explicitly designed to fill the gap that was created by 1997.

Christiane Spiel: I have to say we have many, many students in psychology in Vienna; that means 5,000. It is unbelievable for the United States I think. And we can have a specialization in evaluation. And I think most of them are similar than in Italy. On the one hand, go to consulting or public administrations.

Rebecca Maynard: Okay, Doug.

Doug Besharov: Could I ask—speaking now in my role within APPAM. I'm interested and worried about where we go as an organization. Can you think a little bit and the other members of the panel as well to give us some directions to consider?

Alberto Martini: Well, I think we should have an European APPAM warning about the duplication and I think we should make an effort in that direction because the establishment under which evaluation is done in Europe is beyond redemption. So I'm sorry for the understatement.

Robert Walker: I think I agree with the spirit and the humor of what Alberto says. I disagree profoundly with the content of what he says. I think I share much of the APPAM tradition but I do not—I do move across. I have just written a paper on the application of entropy to policy evaluation. The basic point is that the ontological and epistemological positions of most evaluators in Europe; and therefore, the content of the *European Evaluation Journal* is profoundly different from the content of *Journal of Policy Analysis and Management*. And indeed, if you pick up the American evaluation journal, you will see the content of that is very different from the content of JPAM and so they are different worlds. What I would argue and what I would agree with Alberto is that I think the perspective which JPAM represents is under-represented in European policy evaluation.

Christiane Spiel: As I mentioned at the beginning, I think the European Evaluation Society is a very far way from what you are doing and also what we are doing. And so I fully agree with you. We need much more in the way like APPAM is doing it. I am the first time here and when I listened to the presentation it was only on very few conference sessions. I have the impression everything is very much based on quantitative research and it is very much dominated by this variable-centered approach. That means it is all general linear models on a very advanced way.

And I think this is also a difference to the European tradition. We are much more also likely to apply person- centered approach that we are looking on types of persons. We are also much more like to apply qualitative approaches. That means using more a big variety and I think this is also not so bad. But having in mind it is important also if you start with qualitative data analysis, ending up with quantitative conclusion. That means to applying content analysis and whatever. That means coming away from only subjective perspective toward maybe into subjective—close to an objective one.

Rebecca Maynard: Thank you. I think we are out of time. Let's thank our panelists.