

Prepared Remarks
**Modeling, Simulation and Testing: Collision Course or Happy
Marriage?**

**Philip E. Coyle
March 7, 2006**

**NDIA 22nd Annual National Test and Evaluation Conference
Hyatt Regency Jacksonville**

When I chose as a title for my remarks - *Modeling, Simulation and Testing: Collision Course or Happy Marriage?* - I didn't know I would be following Gen. Larry Welch, always a daunting prospect. Gen. Welch probably knows more about modeling and simulation than any general officer who ever lived. I look forward to every opportunity I have to hear him speak. I've never known anyone who can move from the highest level policy issues down to the individual warfighter with more insight and clarity. Gen. Welch moves with ease from 40,000 feet altitude down to ground truth, and anywhere in between, and he always makes great common sense, as he has again this morning. Mark Twain said "Common sense is not so common," and Gen. Welch proves how true that is. He also made great common sense in a recent DSB study, which I'll come to in a minute.

When I chose my title for this talk, I also didn't know that Admiral Bert Johnston would be giving a talk on Wednesday titled "Open Air Testing vs. Modeling and Simulation: At Last Peaceful Coexistence?"

So before you decide whether you like what I have to say, you'd better wait for Admiral Johnston's talk, as well.

Expectations for M&S - DOD vs. DOE - the differences and the reasons
But to get back to the marriage between modeling, simulation and testing, I've always been puzzled by the fact the M&S is so problematic in the DOD family. In the DOD, M&S kind of nibbles around the edges, which I suppose is alright if that's all the DOD wants, but the DOD and the Services regularly make high sounding pronouncements that M&S is going to be the answer and the greatest thing since sliced bread. While we will hear during this conference examples of excellent M&S work, it is not easy to find examples in the DOD where M&S has really made a difference.

Some of you know that I came to the Pentagon from the Lawrence Livermore National Laboratory where modeling, simulation and testing was - and still is - a very happy marriage. We never did a test without first trying to calculate - model, as you would say - in rather excruciating detail what would happen. It was literally unthinkable that you would spend millions of dollars on a test without making an equivalent effort first in M&S.

For the most part, the DOD does not do that. It's quite common in the DOD to spend millions of dollars on a test without making any significant investment in M&S first. An area of some exception is Live Fire testing, about which I'll say more later.

Nevertheless, I came to the DOD fully prepared support M&S, and did so throughout my tenure at the DOD. But the support I gave to M&S, and the support that Undersecretaries for Defense for AT&L and many other senior leaders have given to M&S, has made very little difference. This is especially surprising since the computational tools that are available today are so much more powerful than they were ten or twenty years ago. Nevertheless, DOD practices haven't changed much. I think there are at least four reasons for this.

First, the focus in DOD defense acquisition is on buying something and moving on, not on understanding for its own sake. Understanding is important, of course, you can't field a system without understanding it, but detailed scientific and technical understanding - both broad and deep - is not the first priority. By contrast, the culture in the development of nuclear weapons has been to achieve first-principles understanding of everything. It's a good thing, because today full-scale underground nuclear testing is no longer permissible. Without those models, the Department of Energy weapons labs would be quite helpless today.

Second, program managers don't want to spend money on M&S if they don't have to. PMs have many demands on their resources, and the incentives are to get the system into production with as little perturbation as possible. In the Department of Energy (DOE) family it is assumed that you will spend money on M&S, lots of money, and it is budgeted every year, year after year, with no expectation that you will ever be finished. Further, in the DOE family there is no expectation that you can trade off M&S for testing; they go together. So there is no PM who is trying to cut testing by spending money on M&S.

Third, the goal for modeling and simulation in the Department of Energy family is to be able to predict with rather astonishing accuracy what will happen. Dr. Keith Bradley will show some examples in his talk Thursday afternoon. While no model is ever perfect, the goal is to come as close to reality as possible, and often the DOE Labs get very close to that goal. This means that modeling and simulation, and the evaluations that come from those models, may produce bad news.

By contrast, in the DOD, the tendency is to expect that test and evaluation will produce bad news and that M&S will produce good news. Thus M&S is often recommended as the better choice, not only because it might supplant testing, but because - generally - the expectation is that M&S will produce better news. Of course, this good news may not have anything to do with reality, especially if the models being used were first developed - as is often the case - to sell the system, not to understand it in technical detail.

In the DOE Labs, a model may produce very bad news, bad news that you have to take very seriously because the model has been doing a pretty good job of predicting reality. That bad news may not actually be bad news, as subsequent tests may reveal, but unless M&S and testing work together, you won't know for sure.

My point is that in the DOE Labs, the intent is that the models will capture the real world, just as tests are trying to capture the real world. Where does a model or simulation get its inputs? Often from test results. How are the test results to be interpreted and understood? Often through modeling and simulation.

Thus, in the DOE Labs, M&S and testing are partners that go together with more or less equal likelihood of providing realistic insights. To get to the bottom of the questions M&S and testing both raise, it is necessary to iterate, back and forth, between the models and live tests.

This iterative process is essential for a healthy M&S program. Accordingly, in the DOE Labs if a model produced bad news, you wouldn't just change the model so it produced happier results, and then stop. This is something I've seen done in the Department of Defense. If a model produces bad news, you shouldn't just change it; you need to understand why! Maybe the problem is real. And to find out takes an iterative process that essentially

never ends. This type of sustained, iterative process between M&S and testing is not very common in the DOD.

Fourth, it is more difficult in the DOD to identify a manageable set of M&S programs and goals. You can't model everything - any more than you can test everything - and DOD systems are often so complex that the number of areas where one could conduct M&S is mind numbing.

In short, in the DOE nuclear weapons family there is really only one overall mission supported by a relatively small number of programs, whereas in the DOD family there are many missions and hundreds of programs.

Recommendations - Four Steps for Change

So to establish a happy marriage between M&S and testing in the DOD it's going to take several new things:

First, DOD leadership will need to make larger and sustaining investments in M&S that are not tied to earlier full-rate production dates or reduced testing. DOD investments are needed in M&S for its own sake. This doesn't mean that M&S won't have to meet programmatic goals. Of course it must. But so long as it is a zero sum game, where a dollar spent on M&S is expected to save a dollar on testing, M&S and testing will have no more than a shotgun marriage relationship, at best.

Second, M&S will have to be supported by an iterative process from continuing and ongoing testing. Some of these tests will be conducted primarily to support M&S, and thereby understanding, not necessarily to support a programmatic milestone. And it also means that the DOD will sometimes get bad news from M&S, and when it does it cannot react by canceling the M&S, or cutting its budget, or cutting its reach.

Third, to make these investments manageable, the DOD will need to focus on mission areas, not programs. Here is where the DSB report that Gen. Welch co-chaired with Bob Herman is important. That report - the DSB Summer Study on Transformation - points out that the DOD must discipline its resources to mission purposes. Under the current approach, DOD always has many more programs all fighting for dollars than it can afford - the bow wave effect in "a sea of force provider interests". The DSB Summer Study recommends - among many other things - allocating resources to mission purposes, constraining plans to intended resources, and measuring progress

against plan objectives. The DSB study also recommends - for the reasons Gen. Welch explained a few minutes ago - limiting "the first stage of spiral development to designs providing: A useful increment of added military capability where there is no more than moderate risk in achieving cost, schedule and performance goals."

Measuring progress against plan objectives, and measuring risk requires modeling, simulation, test and evaluation, all working together.

In this regard, Live Fire testing is an area where M&S practices have been laudable. I don't know how DOT&E does it these days, but Jim O'Bryon used to insist that a Live Fire test not be conducted until it had been modeled first. I thought this was a very important policy and supported him in it throughout. He didn't always get what he wanted. Sometimes the M&S work was poor or half-hearted, sometimes the Services just wanted to do the test and get it over with, but Jim tried to instill in Live Fire testing a partnership with M&S, a relatively happy marriage.

In this sense you could think of Live Fire testing as a kind of "mission" area. Metal hitting metal is something that has to be expected with most battlefield systems, and so you could think of Live Fire testing as something common and overarching individual programs. Often the details in Live Fire tests might be quite common between programs that were otherwise quite different. It is in this sense that one can think of Live Fire testing as a "mission" area not unlike the context in which the word "mission" is used in the DSB report, although they are referring to mission areas in an operational warfighting context. And in this sense one can imagine organizing and managing M&S efforts along mission lines, as well.

My only point here is that by focusing on mission areas - not necessarily individual procurements - just as the DSB report recommends, M&S also could better manage the dazzling array of tasks it might undertake.

The emphasis in the DSB study on increased oversight and measuring risk, brings me to a fourth needed step. To adequately measure risks, the DOD will need to improve its understanding - broad and deep - of the mission area systems it intends to field. This new emphasis on understanding - so as to truly understand the risks involved - will be good for M&S also, and will produce a climate where understanding - not just procurement - is valued for its own sake.

The Defense Acquisition Performance Assessment Report

Having mentioned how helpful I believe Gen. Welch's DSB report on Transformation can be - not only to our overall military establishment, but by extension to how we think about M&S in the DOD, I need to comment on the Defense Acquisition Performance Assessment Report, which came out last month. The DAPA report, so called, complete with glossy cardboard cover and a set of cardboard "3-D" prism glasses such as seen in science fiction movies in the 1950s, is neither effective nor suitable. What the "3-D" glasses are for is a mystery, the DAPA links to their web sites don't work, and the little prisms at the bottom of each page don't shed any light on the real problem.

In fairness, the DAPA report points out that over the years there have been 128 separate studies of DOD acquisition performance, this being the 129th - and does quote others to identify certain problems. On its front page, the DAPA report quotes two recent Acquisition Policy reports by the Congress. The first from the Senate says, " Problems occur because the Department of Defense's weapons programs do not capture early on the requisite knowledge that is needed to efficiently and effectively manage program risks."

The second quote from the House says, "...that the current Defense Acquisition Framework is not appropriately developing realistic and achievable requirements within integrated architectures for major weapons systems based on current technology, forecasted schedules, and available funding..."

The Congress clearly understands the problem. So does the GAO who reported way back in 1971 that "untested and undetermined technology risks" and "poorly defined requirements" were two major factors why DOD acquisitions so often failed. Of course the GAO has pointed this out countless times since 1971, also.

However, the DAPA report does NOT own up to the principal problems already well identified by the Congress or the GAO.

While the DAPA report makes 1,069 recommendations in 42 issue areas, most prominently, the need for improved oversight, it never addresses the

problems with risk that the Congress, the GAO, and Gen. Welch's DSB report address so clearly.

How can this be, you ask? In this issue area - the need for improved oversight - the DAPA report made more recommendations than in any other issue area. Yes, but instead of recommending mechanisms to strengthen oversight and the understanding of risk - such as earlier and more rigorous testing, and concomitant efforts in M&S, - the DAPA report recommends constraining oversight, and constraining the scope of OT&E, operational test and evaluation, which is hardly the problem when it comes to understanding risk.

If in the DOD, M&S actually identified risks, as I am recommending, the DAPA report probably would have recommended constraining M&S as well. Thus the DAPA report, if implemented, would be bad for the future of M&S and the larger role M&S could play.

In addition to being silent on addressing risk and technological maturity, the DAPA report does little to address the incentives needed to change behavior. You all know that incentives are what drive any process, and a lack of incentives is part of the reason why modeling and simulation is so different in the Department of Defense than in the Department of Energy.

Also, the DAPA report does not address the many problems in contracting that come from an inability of the government to get needed information.

This latter point often comes up in an M&S context where contractor models - originally developed to sell a system to Congress and to senior DOD and Service leadership - are treated as proprietary under many contracts even though those models are then used, sometimes misused, in the development of a system.

So unless you want to give the 3-D glasses to your kids, I wouldn't waste your time on the DAPA report.

M&S and Testing - Sharing the Same Goals

Jack Krings has spoken candidly about these things at many conferences. And Tom Christie has spoken and written about them as well. In the February, 2006, issue of Naval Institute Proceedings Magazine, Tom Christie has published a wonderful article titled, *What Has 35 Years of*

Acquisition Reform Accomplished? If you haven't seen his paper, I recommend it. As Tom explains so well, in system after system - time after time - the front end of the acquisition process sows the seeds of future problems. To quote Tom, "...by the time technical and cost issues come to the fore in spades, few, if any, of those involved in the process can bring themselves to admit they were wrong, to cut their losses before inevitable further cost growth and schedule slips, or to demonstrate much needed discipline by making an example of program officials and their contractors who have sold the department and the taxpayer a bill of goods."

Tom points out that what is needed is discipline, exactly what Gen. Welch's DSB report on transformation also recommends. And modeling and simulation can be part of that discipline in the DOD, just as it is in the DOE.

Like all successful marriages, for modeling, simulation and testing to be a happy marriage they have to share the same goals - namely to understand how a system really works, its strengths and weaknesses, and whether the system will be effective and suitable for the warfighter. The purpose of M&S and testing together is to produce insight and understanding. You all know from your own work that when you can measure a result in a test, and then calculate - model - that same result from first principles, that's when you feel you really understand it. You may still have many questions, but you know you are better off than you would be from either testing or M&S alone.

Conclusion

The future of M&S in the Department of Defense will be determined by the goals that are set for M&S. The relationship between modeling, simulation, testing and evaluation will be determined by those goals also. With the computational tools that are available today, there is no longer an excuse for the DOD to expect so little from M&S.

If M&S is going to continue to only nibble around the edges, if M&S is going to be viewed as a source for good news and never for bad news, and if M&S is thought of as a way to cut or avoid testing, then peaceful coexistence may be the best we can hope for. For M&S to not just nibble around the edges, transforming leadership is needed as Gen. Welch calls for.

But if M&S adopts the goals of true understanding, and providing evaluations that both require and complement testing, then a happy marriage

will result. To make a difference to the warfighter, both M&S and testing have to deal with the real world. Testing already does that. It's time that M&S did too.

Thank you.