Minutes of the Advisory Committee
for the Model Analysis Grant
Prepared by Dean Zollman
June 11, 2001

The Advisory Committee for the Model Analysis grant met in Columbus, Ohio, on May 29 and 30, 2001. Members of the Advisory Committee who were in attendance were Andy diSessa, Pat Heller, Jose Mestre, Joe Redish, and Elaine Seymour. Project staff members who attended the meeting were Alicia Allbaugh, Andy Bennett, and Dean Zollman from Kansas State University, and Lei Bao and Gyoungho Lee from Ohio State University. Rasil Warnakulasooriya from Ohio State attended part of the sessions.

We had just started our introduction to the project when the members of the Advisory Committee interrupted to ask about our meaning of the word context. We responded primarily that context meant the scenario in which a physical situation or a problem was set. Pat Heller responded that the word context is used in a broader sense in the proposal and does not necessarily mean just the scenario of the situation. She and Elaine, particularly, suggested that we make our meaning of the word context much clearer.

The discussion of context continued and several additional points were made. Andy diSessa pointed out that using scenarios as a primary way of looking at context made our task very difficult. We will have to be able to show that any new context is isomorphic to others to which we are comparing it. Assuring this type of isomorphism will be a difficult task.

Pat raised a question about whether the students must be able to experience the scenario for it to become part of our efforts. She pointed out that students can not experience things at the microscopic level. But, if we limit our idea of context to scenarios that students can experience, we eliminate anything related to the microscopic world.

Some questions that we should consider and which are related to context include: Does a student think of a certain principle within the context that we have presented? Can a student apply a principle in a particular context?

Joe pointed out that with regard to these questions we need to look at both what is presented to the student and how the students respond. That is, we may see something within a certain context and feel that we have presented it in that context, but the students may see it somewhat differently. Elaine agreed that this was a very important point and that students could recognize things as being quite different from what we think we have presented. In the end, we need to separate the triggers (cues) from the students' responses. In this regard Elaine recommended that we read the papers of Alfred Schutz who wrote about similar ideas in the 1930s.

Andy pointed out that the interpretation of a context is related to the students' prior experiences. We need to understand that students "memories" will affect how they view the context of a situation and that those "memories" are not always correct. People remember observations that they never could have observed because they are not consistent with the physical world. However, the "observations" support the students' view of the physical world. They can remember things that never happened or remember them with different results because they can support alternative frameworks. Several members of the panel agreed that they had observed this phenomenon in their own students and that there is some research that also supports this idea.

Andy also pointed out that we need to look at what is left out of any scenario. Students may be seeing an "agency" which can be cued by many different things, some of which are not intended by us. For example, a student may not have a concept of mass at all. Thus, we need to be sure that we understand how the students interpret what we say. For example, we need to understand what the students understand when we talk about the size of an object. Do they focus on mass, weight, or volume? Each of these concepts could trigger a different response from the student. Thus, in our interviewing process we need to be sure to structure the interviews in such a way that we learn about this type of information.

Pat pointed out that context should include different modes of representation of physical principles. For example, a set of words, an equation, a graph, or a picture of the physical situation may each trigger a different response in a student. We must be certain that we understand how each of these representations are being received by the student.

The topic of discussion then changed to dissemination and how one gets faculty to respond even if we are successful in showing the importance of context. Elaine, in particular, commented that faculty can see the information as it is presented in our data but still may not respond by modifying their teaching style. Several members of the Advisory Committee suggested that the level of dissemination that we presented in the proposal was greater than we could realistically expect to accomplish during the course of the grant. They suggested that we look at the dissemination component as, at best, a pilot study to see how faculty respond to these types of issues and results. Elaine, however, suggested that our results could really "shatter" some of the ways that faculty think about how students approach physics problems. She further suggested that we prepare the data in a form that will help faculty understand this difference. Others expressed some skepticism about whether faculty would respond even to conclusive data.

Overall folks agreed that we need to look at the choice between developing tools ourselves and collaborating with other practitioners in physics teaching. By collaborating with the practitioners we build in more dissemination automatically. However, we may be limited to the work that is being done by others. Thus, we would have certain advantages if we developed the tools for our investigation ourselves. However, then the dissemination might a little bit more limited. The Advisory Committee did not give us a clear direction on this point. However, during further discussions it seemed as if developing our own materials and limiting the amount of dissemination was the preferred approach.

The group then moved to the Ohio State Faculty Club for dinner and further discussions on these issues. The issue of context and its meaning continued to be an important part of the discussion through dinner. No new ground was covered that hasn't been incorporated into the previous paragraphs but the emphasis on making a clear definition was quite well stated.

The morning session on May 30 began with Alicia Allbaugh providing a review of the literature related to context. Pat Heller added to her list a paper that she and co-workers had done on proportional reasoning. They used identical proportional reasoning problems but changed the objects that were discussed in the problems. Several members of the committee suggested that we do a literature search on "modes of representation" because much of the work related to context is also very closely related to how a particular problem or situation is represented. Jose mentioned a study by Judah Schwartz on intensive versus extensive variables that he thought also could fit into this general category. Andy diSessa stated that Jerry Golden had done a definitive review several years ago related to the context in which mathematics problems are presented to students. Pat noted that her group at Minnesota had looked at their context-rich problems carefully and discovered 21 characteristics that make these problems difficult. I do not have in my notes whether this work has been published. Jose mentioned that a student working with him, Tom Koch, is looking at how students cue on different issues. He is finding that the students see a situation quite differently when they are discussing two balls on a ramp simultaneously rather than looking at the behavior of one ball on the ramp then removing it and placing the other ball on the ramp.

These studies are apparently being done in relation to how items with different moments of inertia will roll down an incline.

Modality is also a word that we may want to look at with regard to our efforts. Andy diSessa mentioned a study by Susan Golden-Meadows and her co-workers in which they look at items such as how students are gesturing as they're discussing a particular topic. Their conclusions indicated that changes in the way students gesture about a topic is an indication that they are beginning to have a conceptual change. He also mentioned in this regard the work of McClosky on dynamic versus static images and how that affects students' views of problems. For calculus Ricardo Nemirovsky at TERC has done some work on representations in studying calculus. Joe pointed out that the context of a problem can affect what intellectual resources are activated when a student thinks about the problem. Steinberg and Sabella have looked at this with respect to Newton's First Law and discovered that the context of the problem can affect students if students remember about the fundamental principles.

The conversation then changed to a question about whether students ideas concerning a certain situation or physical principle can in fact be classified into a few models. Some of the work that Andy diSessa is doing indicates that there is a very large range in the types of thinking and models which students use. The data of his present study seem to be supporting this very well. Joe mentioned a recent paper in the *International Journal of Science Education* on models of the sun and moon. The conclusion was that there were about as many models as there were students. He wondered if we were looking at a "disease caused by doctors." That is, we ask a student question, and they respond in a certain way but only do so because we have forced them into thinking about it in a certain way. It could be that students do not have any particular models of the physical world at all. Instead, they just make something up at the spur of the moment in order to satisfy our need to get an answer from them. Andy followed this up with the question, "Do we learn something when we lump a large number of student models together?"

The final topic in this discussion was that of the time scale for conceptual change. Jose and Pat pointed out that if we look at how conceptual change occurs, we know that students must confront their preconceptions and then adopt the accepted physicists' conceptions. In the transition period the students are likely to have more confusion than they will have either by using just their preconceptions or just their "post conceptions." Putting this idea in terms of Bao's vocabulary: a student in a mixed state might actually seem to be acting and reasoning in a worse way than students who hold non-physical models of the world. The period of time to become an expert (or have a poor state along the expert axis) is quite long. While we are looking at students in introductory courses, we are not likely to see the complete transition to the expert state and, therefore, may well see students in a more confused state than they were when they came into their physics class.

The next event was Lei Bao presenting a view of his approach to model analysis. The first question that was discussed by the Advisory Committee was whether students appeared to have a consistent ontological model in their reasoning. Andy stated that everyday reasoning seems to be quite ontological for most people, while Joe stated that the reasoning of students about physics does not seem to have a consistent ontological model.

A discussion of what we call "the null model" was next. Bao stated that the null model, in effect, included all of the stuff that we didn't understand. Joe questioned whether most of the students end up in that particular state. It seems from our data so far that they do not. Bao pointed out that in our pilot test the null model had many more students in it when the analysis was applied to P. World students than it did for more advanced students. Andy pointed out that we are using this analysis to look at the dimensionality of the students' state of understanding but not at the dimensionality of the context of the problem or situation. As we are using this analysis at present, each context must have a different model space because the model space represents the student mental models.

Joe raised two issues with regard to what Bao called concentration analysis. The first one is what is the distribution of model that the students are using? Secondly, do the questions themselves trigger certain models? For example, the assumptions of the Force Concept Inventory are, first, that there are common models among the students' thinking about forces and that the students will see their models in the collection of responses that is presented to them. It is possible that a question will trigger a particular model in a student's mind, but then he or she will not see that model in the responses. As a result the students responses may seem to be less consistent than they would be if the students recognized their model in the responses. The situation can become even more complex because the physicists could see the model in a particular response even though the students do not recognize it in that way.

A discussion about what are appropriate contexts for other concepts to be considered in Newton's Third Law listed a few that we had not included. For example, students may consider the duration of the collision or the hardness or softness of the object in the collision as being important. Our experiment would not have detected this kind of thinking. Thus, we need to be careful how we construct the interviews so that we extract from them all of the different contents which students may consider important. We cannot limit ourselves to those that are traditionally thought of as important by a typical physicist.

Andy Bennett now took over the presentation and had the Advisory Committee work with his online homework system. Andy showed a graph of student participation and noted that it follows an exponential decay curve rather well. The number of participants seems to be exponentially related to the number of trials in which the students participate. Joe suggested that we look at the scores of students on the trial previous to the one they have done. For example, it would be interesting to know how the students who completed a second try did on their first try; how the students completed a third try did on their second try and so forth.

He also noted that in some ways the representations used here are similar to some work by Judah Schwartz in which he looked at how students solved mathematics problems with a broken calculator. Schwartz apparently would provide students with a problem and with a calculator that had lost one or more of its functions. Students then needed to solve the problem with this broken calculator. Thus, they needed to use some of the mathematical logic in order to accomplish all mathematical functions. Joe suggested that the responses that are requested when students are asked to plot lines are similar. They could, for example, obtain a fraction for one of the numbers that they are entering into the computer. However, with the graphs they must use whole numbers. As a result they need to take a reasoning step to figure out what to do. (I suspect that this might provide a useful way to do some think aloud research. We could present this situation to the students and have them think aloud about how they are going to move from an answer that they can't enter into the computer to one that they can enter. The reasoning might be interesting.)

Andy diSessa raised the question of whether the system should provide worked out problems as feedback. In this way the students could see how a problem was worked and then try to use that reasoning to work further problems. In the present system if they get the problem wrong because their reasoning is incorrect they are likely to continue to get it wrong unless they seek help elsewhere. Andy also suggested that we might look at a variety of input mechanisms to see if any of them result in students performing better on these problems than others. Joe mentioned a study by John Seely Brown in which he looked for bugs that students displayed when they were doing subtraction problems. Brown's results were that, while students made a large number of errors, these errors could be explained by only a small number of bugs. He wondered if we could do something similar here and see if errors related to graphical representations in a trig course were related to only a small number of bugs that the students make. (Editorial comment: This suggestion seems to be somewhat the inverse of some of the suggestions about students' mental models in physics. When we discussed the mental models in physics, members of the Advisory Committee questioned whether in fact there were a small number of models that students were using. Here there seems to be some evidence that the students are having a small number of underlying problems that display itself as a large number of errors. Because subtraction and applying models are quite different mental processes, these two statements are not necessarily inconsistent.)

Andy diSessa suggested that another line of research would be simple observation and classification of what the students are doing as they complete some of these exercises. Graduate student observers could watch how the students do problems and simply classify the behaviors. We would then obtain some ideas about their reasoning as we watch these activities take place.

The discussion then changed to how web-based homework might help a physics or mathematics class. Jose mentioned that in their calculus-based physics class, the average time spent on out-of-class activities was two hours per week before they introduced web-based homework. When the University of Massachusetts started using web-based homework in their calculus-based physics classes the amount of time increased to four hours per week --- still admittedly a rather small number compared to faculty expectations. Pat said that they had done a similar study with their calculus-based physics class that also has context rich problems. The average amount of time for their students to work out of class is seven hours per week, and they do not have web-based homework. Jose mentioned that the Physics Department at the University of Illinois has been using web-based homework in their physics classes and collecting a very large amount of data. Perhaps we could find something to use in that collection for our data mining activities.

Joe suggested identifying facets of reasoning in the sine graphs as students work through a trigonometry course. We could then look at the same thing and see if we see the same types of reasoning when the students conclude their physics course and need to use graphs of the sine function.

Andy suggested that we look at the computer-assisted instruction literature for research on various types of feedback and various types of representations that have been used in other computer-assisted situations.

The next session was a presentation of the posters which described research that had been done in the first few months of the project. Because the poster environment lends itself to small group discussions, I did not collect notes during that time. I hope that each of the individuals presenting posters will be able to provide some comments on their discussions to include here.

One conversation that I had during the poster session and during the lunch was with Pat Heller about the work that she and her students have done on the nature of forces. They seem to have a fairly large amount of information about how students think about forces and how students apply Newton's Second Law to a variety of situations. This work includes such conclusions as many students assign acceleration to one force even though there may be several forces applied to an object. If a situation has an unknown force, that force is almost always equal to mass x acceleration even though that might not be appropriate for the problem. Pat and I discussed the possibility of us using the University of Minnesota data as a foundation for a study of Newton's Second Law and the context that we are investigating. This looks like a good possibility to build on work that has already been completed.

During the morning session Elaine asked the Advisory Committee the question, "Is this project doing something which is useful and will contribute to our understanding of student understanding?" When she asked the question, she phrased it in somewhat more diplomatic terms than I have here, but the general idea is to address the usefulness of the project in terms of the broader context of research into student understanding and student mental models in physics. For the most part the members of the committee did not address the question directly but instead focused on what we could do to make sure that our project had the maximum usefulness for the educational research in physics and math teaching communities.

Andy diSessa opened the discussion by stating that careful analysis of the changes in physical features and how they affect student reasoning is at the core of our program and a particularly interesting feature is how the reasoning and mental models develop over time. He said that he would like to hear "some kind of story" about how model analysis relates to the students' reasoning processes. Engaging students in questions about the physical features and about the physical concepts could go a long way toward us being able to answer that type of question. However, we need to keep in mind that the process data is constrained by the questions that we ask. He pointed out that we should keep in mind that we are getting answers to questions, not necessarily unveiling concepts. He suggested that we look at theories of conceptual change and see how they interact with model analysis.

Joe took a somewhat different point of view. He pointed out that our project grew out of the impact of the Force Concept Inventory on the physics teaching community. Our present techniques will not help us understand what individual students are understanding but instead allow us to understand the state of reasoning of a group of students. There seemed to be some agreement among the members of the Advisory Committee that our approaches were looking at classes as they were studying physics rather than at individuals. That led to a discussion of the instructional implications of our results and how we should try to affect faculty and their teaching. One issue that was mentioned repeatedly was that we can inform faculty that the students' wrong answers are as important as the right answers. In fact, our tools and results can help faculty focus on the details of the wrong answers and understand better how to obtain useful data from the wrong answers and the different types of wrong answers. We can in the process of our research develop some tools that we can later "sell" to faculty as innovations in ways of analyzing responses to classes. These tools should be well informed by research. However, there was some concern expressed that we were a long way from effective, easy to use assessment tools and the accompanying instructional techniques. The advice seemed to be to focus on the research aspect during the coming three years. (This item was also discussed later. See below.)

Andy once again emphasized the need for us to do some case studies and to be searching for other concepts that might be valuable and that we might be overlooking. We need to also look carefully at the assumptions that we are making about how people learn and how these assumptions are similar to or different from other research in this area. Overall, he felt that the case studies would help us make better connections with other ongoing research in science education.

Pat expressed concern that what we had done to date seems to be too close to the Force Concept Inventory. She suggested that we need to be careful about that issue and not let the FCI guide us into too narrow of a study. Joe suggested that we could our development process to show how the FCI could be done correctly. The Advisory Committee seemed to be in general agreement that this issue was an important one for us to consider and that we needed to be careful to exploit the impact that the FCI has had on the physics teaching community but at the same time be well aware of its strong limitations as a research tool for the type of work that we are trying to do. As a particular example Pat pointed out that the FCI does not look at many aspects of Newton's Third Law. In particular, there are no questions related to static situations which involve the Third Law or long range forces involving the Third Law.

The discussion then turned to the overall goals of the project. The committee agreed that the number of goals listed in the proposal were greater than we could realistically expect to be done within the amount of time that we have available. A discussion of creating a user friendly form of model analysis that could be used in the classroom and well tested was raised. The general feeling of the Advisory Committee was that it certainly should be a long range goal of the researchers involved in this effort. However, this effort is probably beyond the scope of the present project. By the time we accomplish the research necessary to be able to create a user-friendly form of the material, we will probably have finished the time that we have available.

Another issue raised by Jose was the linkage between the physics component and the mathematics component of the project. At present, the physics and math components seem to be working somewhat independently of each other. This approach is necessary as we are getting the tools and techniques in place. However, the committee is concerned about whether we will be able to measure the transfer. In general the committee seemed to agree that doing the independent components well was more valuable at this time than short-changing that component in order to get more information about the transfer from math to physics.

Pat raised a question about exactly what kind of tool we were proposing to create. She felt that sometimes our proposal implied that it would be a somewhat standard pre-test, post-test type of assessment tool. In this case the students would answer the same questions in the pre-test and the post-test and our analysis would be to use that information to see what, if anything, the students had learned. In her view this type of assessment tool is quite different from a diagnostic tool that is used throughout the instruction to assess how students are learning. For this type of instrument one needs a set of questions that changes as the instruction progresses so that one can see how the students state of understanding is changing. For example, in Newton's Third Law we could imagine having questions that look at how students use techniques such as free-body diagrams, drawing vectors and mathematics and how that fits into their understanding. At the beginning of instruction it would not make much sense to ask them questions about free-body diagrams because they would have no understanding of what that term meant. Others agreed and stated that there is not much research on the progress of student learning during a semester of instruction. This type of research could be quite valuable to the community but would require something different from a pre-test, post-test research design. In this context the idea of doing a case study with a few students who were followed throughout the semester was raised again. Several members of the committee were in favor of this type of approach in an attempt to try to understand how students' state of understanding changes during the semester. (Editorial Note: Along these lines we may want to look back at a study which was done by one of Hans Niederrer's students. He actually followed very carefully *one* student through two semesters of quantum mechanics and tried to follow his state of understanding. I believe that that article was published in the *International Journal of Science Education* about two years ago.)

Returning to the goals of the project some questions were raised about the number of different concepts or topics that we might address. Pat suggested that we limit ourselves to one concept per institution particularly if we are trying to trace students' states of understanding. Developing a sufficiently large number of questions to do this type of tracing throughout a semester or year will be about as much effort as this level of project can undertake.

The conversation then returned to our impact on faculty and their actual teaching behaviors. Elaine mentioned that faculty will respond to well-grounded data on student learning. She felt that we needed to be sure that whatever our research concluded, we would be able to communicate to faculty in a very clear way and provide them with some diagnostic tools to be useful. However, she emphasized that the important part is to have data that are very believable and show the careful nature of the research in order to influence faculty. A couple of other members of the committee expressed some skepticism with how rapidly faculty respond to this kind of data. However, her point was well taken.

Elaine also addressed the point of the separation of the mathematics and physics components of this project. The feeling seemed to be that it is not a problem to keep them separate at this time because the transfer part may be clearly difficult to do. It is better to have well-founded data in both aspects and maybe look toward developing measures of transfer at a later time than to move too quickly and thus have not as well-founded of a study.

To change the topic we then discussed specific concepts which this project might find fruitful in its research. As might be expected the first one that was raised was the concept of Newtonian forces. Pat, in particular, questioned whether students ever really do get the idea of forces. Her view is that students create magical forces in order to be able to apply their understanding of Newton's Laws. She pointed out that the term "the net force" seems very confusing to students and that common language gets in the way. For example, students will talk about the force of an engine and mean something quite different from what physicists mean by forces. (Editorial Comment: The letter to the editor in the latest issue of the *American Journal of Physics* and written by Dan Styer is somewhat relevant to this point.)

Another topic that is closely related and was mentioned several times is the general area of circular motion, centripetal forces, and centrifugal forces. Both Andy diSessa and Jose mentioned that students will give different answers depending on the context in which an object is moving in a circle. For example, different explanations are given for the reason, in terms of forces, that a ball on a string moves around in a circle and a ball in a curved tube moves around in a circle. Pat suggested that forces on objects moving in circles would be a useful one for our investigations but we should put them into a larger context such as Newton's Second Law. Again, she mentioned the work that the Minnesota group had done and how it could be relevant to our efforts.

Andy diSessa suggested that we might want to look at how students understand buoyancy. In part, he suggested this to us because he had tried to understand the student understanding himself and had found it extremely difficult. He felt that every student had a different "story" and that none of them made any sense to him. In particular, he had investigated students' explanations of what happens with a beaker of water that is sitting on a balance and has a piece of lead lowered into the water. Jose mentioned similar experiences with a question in which he had a ball barely floating in some water and then asked students what would happen if he poured oil so that it was on top of the water. Students' explanations of what would happen here seem to be all over the place.

Pat concluded by summarizing some general criteria that she thinks we must consider as we are choosing the concepts. The two most important criteria should be

1. What concepts have the largest research based on which we can build?
2. Which ones are most frequently encountered in the introductory physics course?

The committee generally agreed that these were critical criteria and suggested that we make our choices accordingly.

At this time the committee was getting ready to catch airplanes. During the lunch and post-lunch sessions they had not directly answered the question about whether we were doing something useful. However, because most of the afternoon discussion, which was addressing that question, focused on how we could make our efforts useful to the largest number of communities, we will assume that what we are doing is useful and we need to focus on how to make it the most useful research to the largest number of different people.