PROJECT DESCRIPTION

Overview: A Sequence of Graduated Pedagogical Experiments

Quantitative modeling is the art and science of abstracting the essential features of a situation, building a simplified quantitative representation of that situation, and using that representation to derive useful insights into how to forecast or control its behavior. Building quantitative models has long been routine in science and engineering. It is also becoming common in business, healthcare, and education.

The need to build models and to understand how models can be used (and misused) is no longer restricted to modeling experts. Novice modelers build spreadsheet models every day in thousands of businesses and nonprofit organizations in support of managerial decision-making. Novices also routinely encounter models on the job. (As Sodhi (2003, p.12) points out: “…modeling and abstraction are not obvious to people without OR training. The effect of poor understanding of modeling is a poor understanding of [models embedded in commercial] software.”) Novice modelers typically have little or no training in modeling. Little is known about how they go about their tasks and whether they succeed. No research has been conducted on how best to train novice modelers. This proposal supports an ongoing effort by the co-PIs to train novice modelers and to develop a research-based pedagogy.

The class of model we will study arises from practical problems in business and engineering. These are not homework problems designed to reinforce learning of specific tools, which by their nature are well-formulated problems, with a single preferred approach and a right answer. Rather, we are interested in how novices use modeling to resolve ill-structured problems (Reitman 1965, Simon 1973, Voss and Post 1988) they might encounter on the job, using the basic tools they already know. Two examples follow.

Red Cross: The Red Cross is considering a new policy of paying certain participants for blood donations. Design a model which you could use to advise Red Cross management on the costs and benefits of such a policy.

Boeing: The Boeing Company faces a critical strategic choice in its competition with Airbus Industries for the long-haul flight segment: Should it design and build a super-747 model that can carry 550 passengers at speeds around 350 mph, or a plane that can fly at 95% of the speed of sound but carry only about 350 passengers? Design a model for Boeing executives to use to understand the trade-offs in this problem.

It is generally assumed in the mathematical sciences that modeling is a creative skill or art that is either inborn or can be learned through experience, but that cannot be taught. This assumption may possibly be justified when considering the most elite, expert modelers, yet in many of the arts themselves, including architecture and drawing, creative skills akin to modeling are taught all the time. Furthermore, the question of artistry is less relevant when giving basic training to novices with limited mathematical sophistication who nonetheless are required to build and use models. In our experience, the effective training of novices requires less focus on the sophisticated, creative aspects of modeling and more on simple procedures or heuristics that the novice can use to structure the process.

Powell’s Art of Modeling course will provide the pedagogical vehicle for the research proposed here. Powell has taught this course for 15 years to MBAs at the Tuck School of Business at Dartmouth (Powell, 1995a, 1995b, 1997, 1998). In its design, this course deliberately imitates the art studio, in that the students are presented with modeling problems and coached through
the modeling process rather than being taught through lectures or theory. Willemain has pursued similar themes at Rensselaer with undergraduate management students in a course on Corporate Strategic Planning and Modeling and with graduate engineering students in Modeling Large-Scale Systems.

This proposal describes a three-year research plan that includes both basic and applied components. The primary focus is to carry out basic research on the model formulation processes of novices. No data currently exist on how novice modelers approach the model formulation challenge, so efforts to train novices are of necessity based on personal hunch and experience. We propose to record detailed process information on novice modelers and to create the first research database that others can use for further study of this topic. In this way, the study will advance the fundamental state of knowledge of these phenomena. The study also has an important applied aspect, as we will also test the efficacy of different pedagogical approaches to teaching model formulation. It will generate the first reliable data on how specific teaching strategies alter the model formulation processes of novice modelers. Our results can be expected to have a significant impact on how management science modeling is taught across the curriculum, in both the business and engineering settings.

Now is an especially propitious time to address these questions. AACSB International, the business school accrediting body, recently issued new guidelines that restore the need for “learning experiences in…management science supported decision-making processes.” After a decade of retrenchment, there is a fresh opportunity to restore and reinvigorate management science modeling skills in both undergraduate and graduate business curricula. The president of INFORMS, Tom Cook, has challenged faculty to think hard about new and effective ways to teach management science. We believe our proposal responds to that challenge while also benefitting engineering education.

In brief, we plan a series of graduated pedagogical experiments. Year 1 of the study focuses on individual modelers. We propose to record verbal protocols from novices during the initial stages of formulating models, much as Willemain (1995) did with experts. We will study one group of Tuck School MBAs who are enrolled in the Art of Modeling course both before and after the course. We will also study a control group of MBAs who are not enrolled in the course. This experimental design will allow us to describe the model formulation processes used by these novices and assess the impact the Art of Modeling course has on their processes. As discussed below, we have already conducted a pilot study that has increased our confidence in the feasibility and value of the Year 1 experiment.

Year 2 of the study investigates how coaching by a modeling expert changes the performance of novices. Coaching is a common mode of learning modeling outside of school; it is also the dominant mode used in the Art of Modeling class. In this set of experiments we will assess the impact coaching has on the novice modeling process and test two variations of coaching.

Year 3 of the study considers modeling by teams rather than individuals. This is the most realistic of the experimental settings, as it most closely mimics how modeling is typically done in practice. Teams of MBAs will work intensively over a short period of time on a modeling task. Teams will work for a client, also represented by MBAs, who will judge their work. Teams will also compete against one another.

These three studies: of novices working solo, of novices working with an expert coach, and of competitive teams of novices working for a live client, will provide the first rigorous and comprehensive research findings on novice modelers and will form the basis for further
descriptive studies and studies that test pedagogies. Furthermore, these experiments will provide a structure within which we can take a more holistic view of the data, with an eye toward drawing useful qualitative generalizations about teaching and learning modeling.

Improving our understanding of the modeling process of novices would be a significant advance in the area of Decision, Risk and Management. After this project is over, it may be possible to design further experiments relating the quality of models to the results of decisions based on the models, but that vastly more ambitious goal is not our aim here. We are willing to assert, for now, that better modeling will lead to better models and thence to better decisions.

**Related Research**

Several recent strands of research have begun to break down the barriers to developing a science (and pedagogy) of modeling. In this section, we place the proposal in the context of several relevant literatures.

The pedagogy of the Art of Modeling course relies heavily on the use of *modeling heuristics*: rules of thumb that can be used to direct the modeling effort (Morris 1967, Powell 1995b, Pidd 1999). One such heuristic, for example, is to draw graphs to make relationships concrete. Schön (1983) cites heuristics as fundamental to the problem solving strategies of experts. However, there is no research that establishes the efficacy of teaching heuristics; perhaps heuristics themselves are simply acquired over time by experts, although Schoenfeld (1985) has shown that explicitly teaching college mathematics students to use problem-solving heuristics is effective in improving their performance in this domain. The course follows the problem-based learning approach to the cognitive apprentice model (Hmelo 1998), in which novices learn by working with an expert on real-world problems, with the expert making his thinking visible.

An important part of the cognitive apprentice approach to learning is to make visible the thinking of experts. Willemain (1994, 1995) has studied how expert modelers behave during the early problem formulation phase of the modeling process. This research established the idea that experts use model structuring as a backbone on which to build, while making frequent side trips to consider issues such as the client’s needs, implementation, data availability, and so on. The protocols from Willemain’s study will provide a reference against which to compare the protocols of the novices.

While the proposed research deals with problems that are more open-ended than traditional homework problems, research on the latter provides some useful insights. Alan Schoenfeld (1985) has studied how college students acquire the ability to solve mathematics problems. His findings point to the importance of developing *control skills*: the ability to observe one’s own problem solving effort from a higher level and to make good decisions, for example, on how much time to devote to a particular approach. His pedagogical results suggest that students need to be explicitly taught effective strategies for problem solving and when to use them. (Note, however, that our interest is not in solving well-formulated problems from a mathematics course, but ill-formulated problems from the practical world.)

The ill-structured problems in the proposed research are more akin to design problems than to homework problems. Atman and colleagues (Atman and Bursic 1998, Atman et al. 1999, Mullins et al. 1999) have conducted research on how engineering students acquire design skills. Engineering design, like modeling, is generally considered to be an art that cannot be taught, rather than a science that can. Atman’s research is useful first in describing the design processes of novices, and then in testing various pedagogies for improving these processes. This line of research is closely related to our proposal in several ways: it investigates how
relative novices (engineering undergraduates) learn the craft skills of their profession, it uses checklists to rate the quality of both process and outcome, and it relies heavily on verbal protocol analysis.

Work on expert-novice differences is obviously relevant to the proposed research. There is an extensive literature on this topic in fields ranging from serving a volleyball to piloting an aircraft to reading X-ray films to wine tasting. There is a trickle of papers using functional magnetic resonance imaging (fMRI) to study expert-novice differences (Cincotta and Seger 2000, Lee et al 2001); ultimately, such methods would be a fascinating way to pursue our topic. For the present, however, we have found no literature on the specific topic of differences between expert and novice modelers. The proposed research would fill this gap. Apparently the most similar research has been in studying physics and chemistry problems, since these at least involve equations. For example, Slotta et al (1995) discovered that novices interpreted electromagnetism problems in physical terms, whereas experts used “distinctly nonmaterialistic representations.” Similarly, Chi et al. (1981) found that novices represented problems in a more literal way and experts in a more fundamental way. Crismond (2001) found that novices working mechanical investigate-and-redesign tasks were much more focused on the details of the problem at hand, whereas experts were more likely to link to basic scientific concepts and more probing and strategic in their thinking. Heyworth (1999) found contrasts in “conceptual understanding, use of a qualitative procedure, and the type of strategy used.” Savelsbergh et al. (2002) noted that experts have sufficient knowledge to “take full advantage of problem features at a glance” and can generate multiple representations of a problem. Clement (1998) documented that both novices and experts relied on analogies in problem solving in mechanics; though experts were more effective at using this non-deductive reasoning strategy, their methods were successfully taught to novices. These results in the field of science education are consistent with the general summary of expert-novice differences provided by Glaser and Chi (1988) and Glaser (1990). They reported that experts: perceive larger patterns, work more quickly and accurately, have superior memory, see problems in more fundamental terms, spend more time understanding a problem, and more consciously monitor their own progress. We will look for these contrasts in our protocols.

We conducted a pilot test of our proposed Year 1 experiment during the winter of 2003, when the Art of Modeling course was offered. We collected nine verbal protocols from student volunteers. Volunteers worked for 30 minutes each on two problems before the course began, and again for 30 minutes each on two problems after the course was completed. The problems included the Red Cross and Boeing problems cited earlier, as well as the Alumni Giving problem used by Willemain (1995) with experts. We are currently coding transcripts of these protocols using the same procedure used by Willemain (1995). This pilot study established the feasibility of our approach and suggested several tentative conclusions. We have shown that we can attract volunteers from among MBA students, that such volunteers can think aloud with little intervention for 30 minutes on an ill-formulated problem, that our research associate can manage the process effectively (thereby preserving anonymity for the students from the professor, as required by the Human Subjects Committee), and that (with some minor changes) the problem statements we have used are adequate. The verbal protocols and sketches generated by these volunteers already form a suggestive data set. Two tentative conclusions have emerged about novice modeling. One tentative conclusion is that, unlike experts, novices devote most of their time to problem structuring and little to assessing client needs, data availability, or implementation of results. Another tentative conclusion is that novices tend, under some circumstances, to attempt to develop an answer (or recommendation) to the problem rather than to build a flexible model with which to explore alternatives. Both conclusions are
consistent with Glaser and Chi’s (1988) statement that novices are less self-critical and more superficial in their work.

**Year 1: Experiments with Individuals**

We begin this series of experiments at the most elementary level of analysis: with a single novice modeler working alone. We wish to understand the process novices use to formulate and build models, and we wish to better understand the factors that lead to high quality models. In particular, we wish to test whether courses designed to improve modeling skills actually do so in a positive and measurable manner.

In line with previous NSF-funded research on expert modelers by Willemain (1994, 1995), we will gather data on the modeling processes of novices by recording think-aloud protocols as they work through the initial stages of problem formulation. These protocols will be coded for content and then analyzed for patterns. Willemain (1995), for example, identified a typical organizing framework that expert modelers use in their initial modeling efforts, in which model structure provides the backbone around which the modelers take periodic excursions into other domains. We will be interested to see whether novice modelers also use a backbone structure or, as Alan Schoenfeld (1985) found with college math students, that weak control skills lead novices to focus too heavily on one approach or one aspect of the problem formulation process. (Atman et al. (1999) found something similar when comparing freshman and senior engineers’ approaches to design problems.)

In order to meet the goals of this first year of the study we will need to gather data on the performance of novices both before and after they take Powell’s Art of Modeling course at Tuck. We will also test a control group of second-year MBA students at Tuck who do not take the modeling course. In the first year, students will work two problems (spending 30 minutes on each) in the first phase of the study, which will be timed to coincide with the start of the course. At that time, they will also complete a Myers-Briggs questionnaire, the results of which can be compared to unpublished Myers-Briggs data obtained by Willemain from expert modelers, and a personal questionnaire. Some 10 weeks later, after the course has ended, the students will be tested again, in the same format. They will again work two problems, one a repeat and one new. This format will allow us to measure the improvement, if any, in their efforts on one familiar and one unfamiliar problem. The control group will work the same problems. Exhibit 1 illustrates the experimental design.

Here is a discussion of the critical elements of the experiments.

**Subjects**
The experimental subjects will be second-year MBA students, divided between those who take the Art of Modeling course and those who do not.

**Treatments**
The experimental treatment will be whether the student takes Powell’s Art of Modeling course. The control group will perform the same model formulation tasks as the study group, at approximately the same time in their program, but without taking the modeling course.

**Assignment to treatments**
No randomization in the assignment of subjects to treatments is possible, since students must be free to choose which courses they take. We will collect background data on both study and control subjects in order to compare individuals across groups. We have designed the study to allow us to cancel out selection bias arising from the inability to randomly assign subjects to treatments (see Exhibit 1).

**Exercises**
Each participant will work three modeling exercises, labeled A, B and C, with A done twice. (See Exhibit 1.) Each exercise will take 30 minutes, during which time the participants will think aloud as they formulate models to address the issues in the exercises. Final choice of the exercises awaits the results of our pilot study, but Exercise A is likely to be the Alumni Giving exercise previously used with expert modelers (Willemain 1995) and engineering graduate students at Rensselaer; six expert transcripts of this exercise are available for comparison with the students' efforts. Exercises B and C are likely to be the Boeing and Red Cross exercises shown above. The exercises chosen will have certain key characteristics: they will call for a quantitative model but be open to more than one technical approach, they will not require specialized knowledge beyond the novices' level, and they will permit reasonable progress to be made in the space of 30 minutes.

Process measures of quality
Process will be documented primarily through protocol analysis (Ericsson and Simon 1984, Van Someren et al. 1994, Chi 1997). Process measures must be salient and reliable. One way to proceed is to use a checklist with points awarded for explicitly considering specific items or issues, such as:

- Context: client objectives, client competence
- Structure: number of critical variables, number of critical relationships, extent of mathematical development of relationships
- Realization: data availability, data quality, computation time
- Assessment: comparison to alternative approaches, number of moments of reflection
- Implementation: implementation plan (e.g., schedule of steps).

We have previously developed a prototype checklist and field tested it using student raters at Tuck. The checklist will be converted to a numerical score between 0 and 100. An important part of the first year research will be to refine this checklist.

We will also obtain subjective assessments of process quality from panels of expert modelers. The assessment form will have the same categories as the checklist as well as an “overall” category and will be graded on a 7-point Likert scale. Raters will be invited to participate based on two qualities: their record of accomplishment in operations research modeling and their interest in the teaching of modeling. We will do this fully realizing that Reitman (1965) pointed out that getting agreement on the solution of ill-structured problems is difficult by definition, and Voss and Post (1988) noted the possibility of getting agreement at the cost of narrowing the problem frame. Nevertheless, some form of feedback (e.g., grading) is valuable in the classroom, and the feedback can be valid if it is “…based upon the extent to which a solution can be rationalized…” (Voss and Post, p. 282). We hope to exploit the twice-annual gathering of experts at INFORMS meetings to help accomplish the assessments.

Outcome measures of quality
The 30 minute period that the students will have to work the problem precludes their developing a complete model for us to assess. Instead, we will focus on the quality of their initial approach to the problem, as described above. This is significant, since a problem well formulated is a problem half solved. Outcomes (in the form of complete models) will be assessed in Years 2 and 3.

Sources of experimental variability
- Students’ initial skill in modeling. The students may vary significantly in skill level even within their specific cohort.
- Effectiveness of modeling course. Since any course is an intensely human endeavor, its effectiveness will vary, depending, e.g., on the “chemistry” between the professor and the student and among the students themselves.
- Effectiveness of other courses. Students in both the control and study groups will be growing in skill and experience as they take other courses. That process can be quite
variable, and its effects can splash back into the modeling course, for good or ill. Fortunately, given the Tuck curriculum, students will not be likely to take closely-related courses in the term of the study.

**Threats to internal validity**

The classical Campbell and Stanley (1963) list of threats to internal validity (see also Willemain 1979) can be grouped into several categories: “other influences,” “biases,” “measurement,” and “randomness.”

Other influences include “history” and “maturation.” The proposed nonequivalent control group design (NECGD) is known to be well-insulated from common influences (history) and from maturation (trend). Its potential weakness comes in protecting against differential history or maturation, but differential effects are usually less likely than shared effects.

Simple biases should not be a problem because the NECGD computes differences in gains (see Exhibit 1). “Selection-maturation interaction” is a potential threat to internal validity, since the lack of randomization leaves open the possibility that students who enroll in a modeling course are on a steeper learning curve than those who do not (perhaps because of their interests and skills, perhaps because they take related courses simultaneously, although there are no other courses at Tuck specifically designed to enhance modeling skills). “Experimental mortality” could be a problem if possibly less-motivated students in the control group were more likely to drop out of the study before completing the second pair of exercises at the end of the course.

Measurement problems include “instrumentation” issues (problems in gathering data from the subjects) and “testing” issues (reactions by the subjects to the measurement process). Instrumentation problems will be minimized by standardizing the process of coding both the transcripts of the modeling sessions and the evaluations of the work products. Standardization will follow from clear concept definitions supplemented by examples, adequate coder training, and achieving consensus among multiple coders. Our own prior coding experience (Willemain 1995) and that of others (Sen and Vinze 1997, Mullins et al. 1999), as well as our early efforts at coding the pilot study, suggest that we should achieve at least 80% reliability in coding think aloud protocols. Testing problems will be mitigated by prior acquaintance of subjects and experimenters. The balanced nature of the NECGD will minimize the impact of measurement issues.

Randomness includes “instability” and “regression artifacts.” Regression artifacts will not be a threat because we will not be assigning participants to groups based on their prior modeling skills. Instability becomes less of a problem as sample size increases. Unfortunately, we will be limited to relatively small numbers of subjects (though not small relative to similar studies that have been published), so careful statistical analysis will be a must.

**Sample sizes**

Consider a quantitative response variable, such as a numerical score of modeling process quality. Assuming normal distributions for individual students, the net gain in a NECGD has a normal distribution with mean equal to the size of the experimental effect, $\mu$, and standard deviation proportional to the individual variability in net gain, $\sigma$, and inversely related to the sample sizes in the study and control groups, i.e., $\text{net gain} \sim N(\mu, \sigma \sqrt{\frac{1}{Ns} + \frac{1}{Nc}})$. Given that our informal pilot study achieved $Ns = 7$ and $Nc = 2$, we expect that more organized recruitment would yield at least $Ns = 10$ and $Nc = 4$. These sample sizes would allow detection of a 1.5 $\sigma$ effect with 80% power.

**Threats to external validity**
With reference to Campbell and Stanley’s typology, the two major classes of threats might be called “artificiality” and “replicability”. Artificiality is an issue to the extent that novice modelers in these experiments behave differently than they would either in a classroom or on the job. Differences from classroom behavior could be greater or lesser effort or greater or lesser openness in expressing ideas. We do not expect these differences to be large, since in many ways the proposed experiments will seem quite similar to academic exercises and will be executed in familiar academic surroundings. Differences from work behavior may be greater, since the work environment has many aspects that cannot be duplicated in the proposed experiments. Relevant aspects include the presence of a client and supraclient (Little 1994) distractions and interruptions, interactions with colleagues, even the requirement for a more professional demeanor. We see no easy way to bridge the gap between school-based experiments and on-the-job behavior, but we would argue that the experiments bridge the gap better than conventional classroom exercises or homework problems.

**Year 2: Master/Apprentice Experiments**

In the second year of the project, we will move on to a more realistic type of experiment. While we expect to learn a lot from our experiments with solo modelers, in professional practice it is unusual for a new analyst to work entirely alone. Instead, he or she is almost always part of a team comprised of other analysts and a project leader. We will defer our experiment with full teams to year 3. In year 2, we will perform experiments using a master/apprentice (or coach/student) format. It is common in practice for a new analyst, the apprentice, to work under the supervision of a team leader, the master. The master provides guidance to her apprentice, who then does the technical work.

We envision an instructional format in which a student (apprentice) formulating a model gets some form of mid-course guidance from his instructor (master). This pedagogical format is clearly more costly than having students work on their own, since the instructor has to be directly involved, however briefly. However, we expect this format to be not only more realistic (which is valuable in its own right) but also more effective in developing the skills of the student. The instructor can steer students away from dead ends, reinforce good ideas, and ask questions to help a student having trouble getting started. This format is used routinely in the “studio” approach to teaching professional craft (Powell 1998, Pollack 1976, Schön 1987).

We want to document the effect of providing expert consultation, and to compare consultations provided at two different phases in the problem solving process. Our control condition is to provide no consultation: the student works for 60 minutes, takes a 10 minute break, and then works for another 60 minutes. One experimental condition provides consultation early in the process, allowing 15 minutes with the master after 15 minutes of initial work. The other experimental condition delays the consultation until right after the 10 minute break. Exhibit 2 illustrates the experimental design. We expect that the earlier consultation will have its greatest effect on problem framing, and the later consultation will have its greatest effect on model realization.

Here is a discussion of the critical elements of the experiments.

**Subjects**
The roles of master will be played by Powell and Willemain. The apprentice rolls will be played by second year MBA students.

**Treatments**
The experimental treatment will be level of expert consultation: none, earlier, or later.

**Assignment to treatments**
Students will be randomly assigned to one of the three treatment levels. This randomization will protect against biases due to systematic differences in the skill levels of apprentices in the three treatments. This true experimental design is known as a posttest-only control group design (POCGD).

The sequence in which the treatments are applied will also be randomized. This randomization will protect against learning effects on the part of the masters.

Exercises
There will be at least four exercises, administered to each subject twice over the course of an academic term. Having multiple exercises will allow for a greater range of behaviors. The exercises must meet certain criteria, i.e., they must:

- be doable in two hours
- conclude with a tangible work product that can be assessed for quality (e.g., an equation or set of equations, a computer program, pseudo-code, Excel spreadsheet)
- be self-contained, not requiring resources (data or computation) beyond those provided during the experimental session
- allow for multiple possible technical approaches

Process measures of quality
The same measures of process quality used in Year 1 will be used again.

Outcome measures of quality
Unlike in the Year 1 experiments, this time the students will have enough time to actually produce a model. A major part of our research will be developing instruments to rate the quality of the models produced by students in each of the three treatment groups. We expect to have to tailor the ratings to the specific modeling exercise, but some general principles are clear. Our approach starts with the fundamental realization that a model is an artifact created to do a job for a client. This means that the traditional standards for evaluating a model in, say, physics, do not apply. For instance, “beauty” carries very little weight, complexity is more a vice than a virtue, and the ability to compute results before a decision deadline is crucial. Willemain’s (1994, p. 215) survey of expert modelers provided useful guidance for developing an assessment instrument: “When listing important qualities of an effective model, the experts mentioned, in decreasing order of frequency, qualities having to do with 1) validity, 2) usability, 3) value to client, 4) feasibility, and 5) aptness for client’s problem.” These general attributes must be interpreted and weighted in the context of the exercise at hand. For instance, one expert interpreted “value to client” as meaning “provides insight”. In this regard, a “bad” model - in the sense of a model that cannot be validated or is actually invalidated by evidence - may still be very valuable (Hodges 1991, Bankes 1993). Our plan is to develop exercise-specific rating sheets to provide a numerical score for model quality. These sheets would be filled out by the PIs and also by panels of expert modelers recruited to judge model quality. Since this kind of rating necessarily has a large subjective component, we will ask for ranges rather than single-number scores and then take proper account of this ambiguity when comparing the performance of students in the three treatment groups.

Sources of experimental variability
- Students’ initial skill in modeling.
- Consultant skill in providing guidance. Masters may be more or less alert to problems in the apprentice’s approach. They may provide more or less constructive suggestions to problems that they do recognize. They may be more or less emotionally supportive in the way they make their suggestions.

Threats to internal validity
A POCGD is well-protected against all threats to internal validity (Campbell and Stanley 1963, p. 8). This means that any differences observed among the three groups (no consultation, early consultation, midway consultation) will reliably be attributed to the experimental intervention.
Sample sizes
We expect to have about the same sample sizes in Year 2 as in Year 1. A preliminary power analysis (assuming normally distributed scores, 5% significance, 80% power, and effect sizes of 0, 1 and 2 Sigmas), suggested the need for about 6 subjects per group. These sample sizes seem achievable for the second year of the study. More detailed planning is impossible because this experiment has never been run. One of the benefits of the experiment will be to get first estimates of effect sizes and noise variances.

Threats to external validity
Artificiality is a potential problem in that subjects in the control group might learn of and resent the extra attention given to subjects in the two experimental groups. However, we do not consider this a high risk since the treatment is not remedial.

Year 3: Competitive Game Experiments
In the third and final year of the project, we will move on to a yet more realistic type of experiment. In today’s professional practice, a novice modeler is most likely to be assigned a role within a team consisting of several technical analysts and a team leader. Some newly-minted MBAs may get an early assignment as a team leader rather than analyst.

The third year experiments will document experience with a form of pedagogy very different from the traditional group report: role-playing games within the context of competing teams. Willemain has used this format successfully in classes for management undergraduates and engineering graduate students.

Exhibit 3 shows the structure of the games. Each game is organized around a technical competition in response to a request for proposal (RFP) from a client. The RFP will call for a model to solve a problem. An example which Willemain has used in prototype form with both undergraduate and graduate classes at Rensselaer is to develop algorithms and software to do timetabling, i.e., to assign university classes to rooms and time periods. The client team manages the proceedings and decides the result. Two competing teams bid for the work; each team has one leader and three analysts.

The competitive game has a number of pedagogical advantages over the conventional group report:

- It emphasizes the role of the model’s client. The competing teams must not only accommodate to the existence and objectives of a client but probe for the relevant characteristics of their particular client, such as level of background knowledge. Modeling is very different when done for a client than when done for a class (Little 1991, Quade 1988).

- It emphasizes aspects of modeling beyond the technical. In particular, the response to the RFP will require estimates of time and cost to do the job, and it must be both understandable and persuasive to the client. Most novice modelers have little experience with thinking about these aspects of modeling: rarely do they appreciate that the modeling process itself needs to be managed (Gass 1987).

- It emphasizes the contrast of competing technical approaches to the same problem. Client team members must formally evaluate the competing approaches presented by the two teams of bidders. Technical analysts must propose, assess, and at least partially develop competing approaches. Team leaders must consider the tactical advantages of each approach and choose one.

- It emphasizes the problems of effective modeling in a team setting. While a team can be much more knowledgeable, creative and productive than an individual modeler, it can also be inefficient or become deadlocked. Important skills are needed to keep a team productive, including effective problem decomposition, good leadership, and good followership.
• It provides stronger motivation than the typical report format. Willemain’s experience at Rensselaer is that students quickly get very involved in competitive games.

Here is a discussion of the critical elements of the Year 3 experiments.

**Subjects**
Second-year Tuck MBA students will play the roles of clients, team leaders, or analysts. In the traditional report format, the MBA students will work in conventional groups.

**Treatments**
The control treatment will be developing a model in the form of a traditional group project. The experimental treatments will be playing a competitive game.

**Assignment to treatments**
Subjects will be randomly assigned to either the study or control group. Within the games, students will be randomly assigned to client or team leader roles. Within the traditional groups, students will be randomly assigned to the role of group leader.

**Exercises**
There will be two exercises for the teams to work. One is likely to be the timetabling exercise mentioned above, since this has been used three times successfully, if informally, at Rensselaer. One advantage of this exercise is that all students will be well-versed in the higher education system, so a lot of problem context will be immediately obvious. The other exercise might be one with which few students will be personally familiar, such as the Boeing problem cited earlier.

Having multiple exercises will allow for a greater range of behaviors. The exercises must meet certain criteria, e.g., they must:
- be doable in one week of calendar time and perhaps 8 hours of work time
- require more than one person to be completed
- allow for multiple possible technical approaches.

**Process measures of quality**
The modeling processes used by each team will be evaluated in part using the same criteria employed in the Year 2 experiments. Because of the extra complexity of group interactions (e.g., simultaneous side conversations), we will not attempt to obtain transcripts from recordings of the group meetings. However, we will make audio (and perhaps video) recordings for internal use, and we will observe and assess the group meetings using the same checklist used in the Year 1 experiments. Furthermore, additional dimensions will be monitored having to do with the effectiveness of the various actors in their roles. For instance, analysts should propose and develop alternative technical approaches; leaders should develop frameworks for evaluating the alternatives. We will also obtain evaluations by the students of their peers on how well their peers performed their roles (e.g., decisiveness of team leaders, responsiveness of technical analysts, and clarity of clients’ statement of intentions). Regarding the value of video recordings, Garber and Goldin-Meadow (2002) pointed out the value of simultaneously analyzing words and gestures during a problem solving session.

**Outcome measures of quality**
The models proposed by each team will be evaluated using the same criteria employed in the Year 2 experiments. In addition, the client teams will provide additional data in the form of their selections of the winning bids. There will also be an additional dimension to evaluation of the game itself: we will devise a survey to assess the students’ perceptions of the educational value of the competitive game compared to the traditional group project.

**Sources of experimental variability**
- Students’ initial skill in modeling.
- Students’ skill in leadership.
Students’ ability to be productive in a spatially-distributed team setting.

Threats to interval validity
The use of a POCGD will protect against most threats to internal validity.

Sample sizes
The longer duration, greater number of work hours, and larger number of students associated with competitive games means that this experiment makes more operational demands. We will require enough students to run at least one traditional group and one game. Each game needs 8 students (2 members of a client team and two teams each consisting of one leader and 2 analysts). To hold team size constant, each traditional group needs 3 members. Therefore, we need at least 11 students. Each team in the game might require five sessions of perhaps 90 minutes each, for a total of about 8 hours over one week: an initial meeting, two or three working meetings, and a final session at which the two teams present their responses to the RFP and a winner is chosen by the client team. We plan to run the experiment a second time, using the same team assignments to exploit familiarity from the initial run but focusing on a different problem.

Threats to external validity

• Artificiality: Competitive games are less common and more exciting than traditional classroom formats; they sometimes lead to odd behavior. Willemain has seen some undergraduate teams’ efforts devolve into elaborate shows of “PowerPoint fluff” with little technical content (This could be realistic behavior, but not one we want to encourage; graduate students have not reacted the same way.). Basically, however, competitive games are more like what many modelers really do for a living than are typical academic assignments.

• Replicability: If the games prove to be as useful as we expect, they should be within the means of most graduate programs that teach modeling.

Results from Previous NSF Research
Powell’s research has been internally funded by the Tuck School, as is common in business schools. Willemain has received three NSF awards during the previous ten years. Two have been from DRMS and lead directly to the present proposal. The first award studied the first hour in the life of a model. It exposed expert modelers to multiple modeling exercises, then recorded, transcribed, coded and analyzed their think-aloud protocols and sketches. It also solicited experts’ opinions about good modeling practice and descriptions of their own work.

• NSF award number, amount and period of support: Directorate of Social, Behavioral and Economic Sciences, Program in Decision, Risk and Management Sciences; Grant #9012094; $94,820; August 15, 1990 to October 31, 1992.

• Title of the project: Exploring the Process of Model Formulation

• Summary of the results: We obtained think-aloud protocols from 12 expert modelers. A core group of 4 experts each worked 4 different model formulation exercises, spending one hour on each exercise. A supplementary group of 8 modelers added another 2 think-aloud protocols for each exercise, for a total of 24 protocols. We coded and analyzed the protocols and made them available for other researchers to use. Some of the findings from analysis of the protocols were: experts vary their attention among multiple topics while nominally engaged in model formulation; the core activity is an alternation between doing a little modeling and doing a little assessment of the model; there is a lot of variety in experts’ approaches to an individual problem; across problems, there is quantitative evidence of individual modeling style; sketches and other visualizations play an important role in experts’ approaches to model formulation. Before obtaining the protocols, we also obtained from
each expert their opinions on fundamental modeling questions (e.g., what constitutes a “good” model), a description of their own practice, stories useful for novices, and Myers-Briggs assessments.


• Brief description of research products not described elsewhere: n/a

The second award investigated the role of visualization in modeling. One stream of work executed and analyzed experiments in visual display of modeling heuristics. Another stream rigorously developed the concept of model context and developed guidelines and methods for implementing visual representations of model context in the future incarnation of the world wide web, i.e., the Semantic Web.

• NSF award number, amount and period of support: Directorate of Social, Behavioral and Economic Sciences. Program in Decision, Risk and Management Sciences; Grant #9730465; $264,690; June 1, 1998 to May 31, 2002 (William A. Wallace, Co-PI)

• Title of the project: “Visualization and the Process of Modeling: A Conceptual Approach”

• Summary of the results: The research centered around three dissertations. L. Waisel’s PhD dissertation involved human-subjects experiments in which decision makers were presented with varying degrees of visual feedback about modeling heuristics. The goal was to relate task performance to the degree of visual feedback support. The task was a problem in groundwater modeling using Darcy’s Law. Waisel also conducted a quantitative study of expert’s use of sketches during model formulation. S. Ganaway’s MS thesis repeated some aspects of Waisel’s heuristics visualization experiments under different conditions. A. Crapo’s PhD thesis involved interviews with modeling experts about the role of visualization in modeling. It also formalized the idea of the context of a model and implemented context information in a form suitable for the Semantic Web.

• Publications resulting from the NSF award: Crapo et al. 2000; Crapo et al. 2002; Waisel et al. in process; Willemain et al. in press.

• Brief description of research products not described elsewhere: n/a

Summary
Novice modelers are building more and more models in business settings because of the widespread availability of appropriate software tools. However, no data currently exist on how novice modelers approach the model formulation challenge, so efforts to train novices are of necessity based on personal hunch and experience. The primary focus of this proposal is to carry out basic research on the model formulation processes of novices. The class of models we will study (1) arise from practical problems in business and engineering (2) are ill-formulated in the sense that the problem itself, the approach, and the nature of the solution are not known; and (3) can be resolved using basic methods (not the advanced methods of expert modelers.) We propose to record and analyze detailed process information on novice modelers and to
develop and use instruments for assessing the quality of students’ models and modeling processes.

The study also has an important applied aspect, as we will also test the efficacy of several different pedagogical approaches to teaching model formulation. Year 1 of the study will focus on individual modelers, using verbal protocols from novices during the initial stages of formulating models. Year 2 will investigate how coaching by a modeling expert changes the performance of novices. Year 3 will consider modeling by competitive teams rather than individuals. All studies will be quasi-experiments or true experiments. The results can be expected to have a significant impact on how modeling is taught across the curriculum, in both the business and engineering settings.

Exhibit 1: Individual Modeler Experiment, Year 1

<table>
<thead>
<tr>
<th>Control Group</th>
<th>Exercise A</th>
<th>Exercise B</th>
<th>Exercise C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Group</td>
<td>Exercise A</td>
<td>Exercise B</td>
<td>Exercise C</td>
</tr>
</tbody>
</table>

- Exercises A, B and C take 30 minutes each. There will be a short break between the two exercises worked on the same day.
- Each will be scored in two ways: by a checklist of process events, and by ratings by multiple experts of process quality.
- We will obtain two estimates of the differential gain in performance due to the course by computing:

  \[ [A(\text{study,after})-A(\text{study,before})] - [A(\text{control,after})-A(\text{control,before})] \]
  and
  \[ [C(\text{study})-C(\text{control})] - [B(\text{study})-B(\text{control})] \]

Exhibit 2: Master/Apprentice Experiment, Year 2

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Segment Time (minutes)</th>
<th>Segment Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>60</td>
<td>Solo work</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Break</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>Solo work</td>
</tr>
<tr>
<td>Early Consultation</td>
<td>15</td>
<td>Solo work</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>Consultation</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>Solo work</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Break</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>Solo work</td>
</tr>
</tbody>
</table>
Exhibit 3: Competitive Modeling Game, Year 3

**Study Group**

**Client Team**

Played by: Two students  
Responsibilities: Respond to questions about RFP; manage the presentation meeting; determine criteria for winning bid; assess competing teams’ presentations and written bids; determine a winner; explain and justify the decision. (Post-game: evaluate the performance of the two team leaders.)

**Team Alpha**

Role: Team leader  
Played by: One student  
Responsibilities: Determine team strategy in responding to RFP; assign work to team members; direct the development of the response to the RFP; lead the presentation of the team’s bid. (Post-game: Evaluate the performance of the client team; evaluate the performance of the technical analysts of Team Alpha.)

Role: Technical analyst  
Played by: Two students  
Responsibilities: Help the team leader select a response strategy; execute the technical work; provide inputs to the leader’s presentation. (Post-game: Evaluate the performance of the leader of Team Alpha.)

**Team Bravo**

(Competes with Team Alpha)

**Control Group**

Traditional team of 3 members