Key Challenges to Model-Based Design:
Distinguishing Model Confidence from Model Validation

by

Genevieve Flanagan

B.S. Mechanical Engineering, Illinois Institute of Technology, 2000
M.S. Mechanical Engineering, Purdue University, 2003

Submitted to the System Design and Management Program
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the

Massachusetts Institute of Technology
June 2012

Signature of Author ______________________________
Genevieve Flanagan
System Design and Management Program

Certified by ______________________________
Olivier L. de Weck
Thesis Supervisor
Associate Professor of Aeronautics and Astronautics and Engineering Systems

Certified by ______________________________
Noelle Eckley Selin
Thesis Supervisor
Assistant Professor of Engineering Systems and Atmospheric Chemistry

Accepted by ______________________________
Patrick Hale
Director
System Design & Management Program

This page intentionally left blank

Key Challenges to Model-Based Design:
Distinguishing Model Confidence from Model Validation

by

Genevieve Flanagan

Submitted to the System Design and Management Program on May 11, 2012
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management
at the
Massachusetts Institute of Technology

Abstract

Model-based design is becoming more prevalent in industry due to increasing complexities in technology while schedules shorten and budgets tighten. Model-based design is a means to substantiate good design under these circumstances. Despite this, organizations often have a lack of confidence in the use of models to make critical decisions. As a consequence they often invest heavily in expensive test activities that may not yield substantially new or better information. On the other hand, models are often used beyond the bounds within which they had been previously calibrated and validated and their predictions in the new regime may be substantially in error and this can add substantial risk to a program. This thesis seeks to identify factors that cause either of these behaviors.

Eight factors emerged as the key variables to misaligned model confidence. These were found by studying three case studies to setup the problem space. This was followed by a review of the literature with emphasis on model validation and assessment processes to identify remaining gaps. These gaps include proper model validation processes, limited research from the perspective of the decision-maker, and lack of understanding of the impact of contextual variables surrounding a decision. The impact these eight factors have on model confidence and credibility was tested using a web-based experiment that included a simple model of a catapult and varying contextual details representing the factors. In total 252 respondents interacted with the model and made a binary decision on a design problem to provide a measure for model confidence.

Results from the testing showed several factors proved to cause an outright change in model confidence. One factor, a representation of model uncertainty, did not result in any differences to model confidence despite support from the literature suggesting otherwise. Findings such as these were used to gain additional insights and recommendations to address the problem of misaligned model confidence.

Recommendations included system-level approaches, improved quality of communication, and use of decision analysis techniques. Applying focus in these areas can help to alleviate pressures from the contextual factors involved in the decision-making process. This will allow models to be used more effectively thereby supporting model-based design efforts.

Thesis Supervisor: Olivier L. de Weck
Title: Associate Professor of Aeronautics and Astronautics and Engineering Systems

Thesis Supervisor: Noelle Eckley Selin
Title: Assistant Professor of Engineering Systems and Atmospheric Chemistry

This page intentionally left blank

Acknowledgments

I would like to thank John Deere for supporting me through the course of this journey, without which none of this could have been possible. To my colleagues at Deere, thank you for your many discussions and support.

In addition, I would like to thank the SDM staff for their continued support and encouragement. Additional appreciation goes to all those in MIT's SDM cohorts for their conversations, encouragement and friendship. We had some good times in Boston and I will miss you as we all begin to disperse.

For the many people that took the 30 minutes to complete my web-based experiment I extend great appreciation. I hope you found it enjoyable. Many of you were concerned if you chose the "right" answer, but I think you will find from this thesis that the results worked out very well and would not have been possible without your support.

I would like to extend special thanks to my advisors, Professors Oli de Weck and Noelle Selin, for their inspiration, wisdom and focus throughout this process. I would not have been able to make this thesis what it is without your guidance.

Finally, I cannot begin to express both my love and appreciation to my husband. We've been through a lot during this journey - living apart while I've been in Boston, countless lost nights and weekends to homework and thesis. Despite that, you continue to be unrelenting in your support and encouragement. Whether it be listening to me ramble about my thesis, proofreading it for the nth time, or just reminding me that I could get through this. Thank you.

This page intentionally left blank

Table of Contents

This page intentionally left blank

1 INTRODUCTION

1.1 MOTIVATION

Model-based design is becoming more prevalent in industry due to increasing complexities in technology while schedules shorten and budgets tighten. Model-based design is a means to substantiate good design under these circumstances as it uses models and simulations of a real system to quickly test many options within a design space. It allows decision-makers choose an optimum solution before the system can be physically tested (Hosagrahara & Smith, 2005). Despite this, however, organizations often have a lack of confidence in using models to make critical decisions. As a consequence they often invest heavily in expensive test activities that may not yield substantially new or better information than what the models already in their possession could have predicted. On the other hand it is also true that models are often used beyond the bounds within which they have been previously calibrated and validated and produce predictions that are substantially in error and this can add substantial risk to a program.

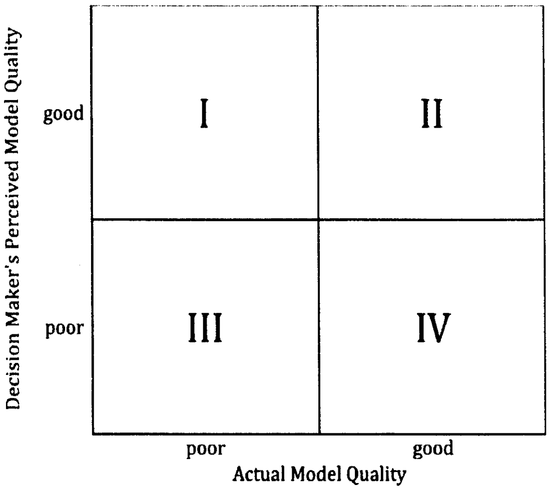


Figure 1: Four-box model of a representing the decision-maker's perception of a model's quality versus the actual quality of the model relative to the problem being addressed.

This phenomenon can be explained using a simple four-box model shown in Figure 1. The x-axis of this figure portrays the actual quality of a model relative to the problem for which it is being used. Two general levels are shown for simplicity of either good or poor; good actual model quality is defined as the model being a sufficient representation of the real system it represents within some a priori modeling error tolerance and therefore an appropriate tool to model that system whereas poor quality indicates that the model may have serious flaws modeling the real system and may produce answers that fall outside the modeling error tolerance band over all or some significant portion of the design or decision space. However, "there can be no proof of the absolute correctness with which a model represents reality" and therefore "confidence is the proper criterion" (Forrester & Senge, 1980). The y-axis of the four-box model represents the confidence, or the perception of a model's quality from the perspective of the decision-maker for the problem. Again, this is rated as either good or poor depending on if the decision-maker trusts the model and uses it to make decisions.

As shown in Figure 1, quadrants II and III indicate appropriate alignment between the perception of a model and its actual quality for the intended purpose. In quadrants II and III, a decision-maker is able to properly distinguish whether a model is appropriate to use for the problem or not. Quadrants I and IV, however, represent where issues can arise in implementing model-based design. Quadrant I shows the case where a decision-maker believes a model to be good for a problem, however, the model is not in fact appropriate to use for the problem and may lead the decision-maker astray. This is in contrast to quadrant IV, where the model would be a good tool to help solve a problem; however, the decision-maker does not agree and continues without input from the model, effectively dismissing its predictions.

For model-based design to be effective in organizations, the optimum is to operate in quadrant II, where mature, high quality models are available and the organization takes advantage of those models. This results in better designs and higher efficiencies (Smith, Prabhu, & Friedman, 2007). Achieving models that are indeed valid and mature and aligning the organization to have confidence in these models may require substantial investment of capital and human effort. But how does an organization know if the models being used are trustworthy to make critical decisions? If an organization were to act on a poor-quality model, as depicted by quadrant I (perceived good, actually poor), the consequences could be severe. For risk-averse organizations, this may shift behavior to quadrant IV (perceived poor quality, actually good quality), where a model may exist and provide appropriate answers, but it may be less of a risk to seek additional sources of input such as results from other models or physical experimentation. For industries that have the option of physical testing of their systems, this can lead to excess resources being used, but with added confidence in the final decision.

1.2 THESIS OBJECTIVES AND APPROACH

The objective of this thesis is to understand the factors that cause perception of model quality to differ from the actual quality of the model. Thus the focus is on quadrants I and IV. Three case studies, drawn from the public domain and industry, will be examined as representative of quadrants I and IV in the four-box model in Figure 1.

Reviews of these case studies will setup the problem space and motivation behind this thesis. This will be followed by diving into better understanding the possible causes of the problems by means of a literature review resulting in the root cause to the problem. From this will emerge a set of factors that describe some of the reasoning as to why model perception drifts from the actual quality of the model.

These factors are then tested in an experimental setting with users from industry to illustrate the effect of these factors on perception of model credibility. The experiment was carried out on the internet, using the model of a simple catapult system that propels a ball through a ballistic trajectory with the horizontal impact distance as the output and the pullback angle, launch angle, type of ball and number of rubber bands powering the catapult as input variables. In this testing, a model's quality was changed from good to poor by asking a subset of the respondents to use the model outside its range of validity. The model and validating data were made available to the 252 test subjects along with varying details surrounding the model that may affect a user's perception of the model either positively or negatively. The subjects were asked to grade the credibility of the model using the framework presented. Success was measured by whether the model was rated appropriately given the known quality of the model.

The results from this experiment reveal not only whether the factors had an impact on the decision-making process, but also suggest methods for how they can be better managed to promote proper alignment between perception and actual quality in more complex decision-making situations in industry and in other settings. This will therefore provide some guidelines organizations can follow to help them adopt an effective model-based design initiative.

1.3 THESIS STRUCTURE

Following this introduction in chapter 1, chapter 2 of this thesis will present the three case study reviews. Each case study will include background of pertinent events followed by a summary of the underlying problems that impacted model usage in those cases.

Chapter 3 will then analyze the problems raised in section 2 further to determine the root cause behind them. It will begin with a statement of the problem and continue with a discussion from the literature that will provide further definition to the problem and also present work that has been done to address these problems thus far. Finally, this section will conclude with a proposal for the root cause of model misuse.

Chapter 4 will discuss the framework that emerged from the research. Eight factors will be presented; for each one, a definition will be provided, a discussion of how that factor impacts the decision-making process via the four-box model (Figure 1), and then examples from the case studies to illustrate the points.

Chapter 5 then discusses the details of the experiment that was conducted to test these factors. It begins by stating the hypothesis, then reviews how the experiment was setup and implemented, and concludes with a discussion of the results relative to each factor.

Finally, Chapter 6 presents the overall conclusions gained from this research including not only recommendations to help address issues with model-based design, but also areas uncovered for future research.

2 CASE STUDY REVIEW

The four-box model presented in Figure 1 represents two domains of the problem space: one of actual model quality and one of perceived model quality. For model-based design to be effective in organizations, there are two key activities. The first is to get the actual quality of the model to be good; the second is to then get the perception of that model to match, represented by quadrant II in the four-box model. In this section to follow, three case studies will be presented that are examples of misalignment between perception and actual model quality to demonstrate the problem space. First, a case will be presented from industry that shows the resulting behavior when model-based design is not internalized within an organization, as the models are good, but the confidence to accept their results is lacking. Following this example, a highly publicized case study will follow from the Eyjafjallajökull 2010 volcanic eruption that closed much of European airspace. In this case, there again emerged perception issues of the models. This case also presents the concern of potential risks of making decisions in the absence of a model. The final case is from the space shuttle Columbia accident in 2003. This example demonstrates the potential hazards that result from quadrant I (perceived good, actually poor) model behavior.

2.1 EGR MEASUREMENT VENTURI AND THE LACK OF MODEL-BASED DESIGN

The first case to review is one from industry based on experience from the author. It is cases such as these that present frustration in organizations trying to integrate model-based design within their processes. Models may be well done and executed; yet design decisions are made based not on the correct model results, but it is decided instead to go forward with physical testing resulting in unnecessary prototype costs and extended development time in order to confirm what the model has already predicted. As the details of this case are not in the public domain, this section will begin with some background and will further discuss the elements of the model and decision-making that ensued.

2.1.1 BACKGROUND

This case comes from a company that manufactures and sells heavy-duty diesel engines for use in industrial off-highway applications and is based on a real design problem that occurred in the 2008 timeframe. Industrial off-highway diesel engines are regulated by the Environmental Protection Agency (EPA) with respect to harmful substances they can emit during operation (emissions) – of particular concern being Nitrogen Oxides (NOx) and particulate matter (PM). The levels of these emissions are measured in a laboratory where an engine is running a series of tests and the engine speed and load are either steady state or highly transient. The EPA also prescribes the duty cycle and test profile the engine has to be subjected to during testing.

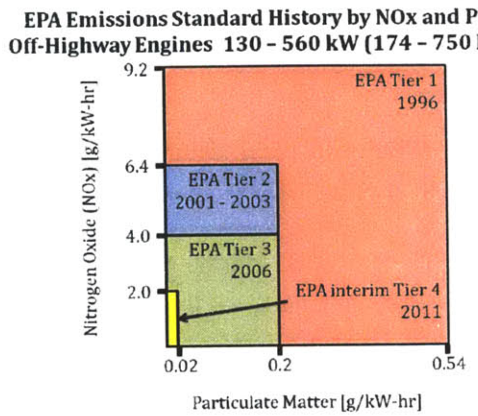


Figure 2: Evolution of EPA Emissions Standards for heavy-duty off-highway diesel engines. The x-axis shows the PM standard level and the y-axis shows the NOx standard level. To be in compliance, the engines must have composite emissions within the boundaries of their respective tier level [adapted from (EPA, 2011)].

Beginning in 2011, diesel engines with power outputs greater than 37 kW (50 HP) were required to meet a new emissions standard called interim Tier 4 (EPA, 2011). This emission standard required a 50% reduction in NOx and 90% reduction in PM as compared to the prior Tier 3 emission standard (Figure 2). In addition to reduced emissions, the transient test requirement came into effect in addition to the existing steady state test protocol. The transient test consisted of 1,239 engine speed and torque conditions run at 1-second intervals, thereby resulting in a 20-minute test procedure (right chart in Figure 3). The transient test shown below is plotted over the steady state points from the left chart but has included a point at each speed and torque condition from the test procedure with lines connecting these points to show the movement in speed and torque throughout the test.

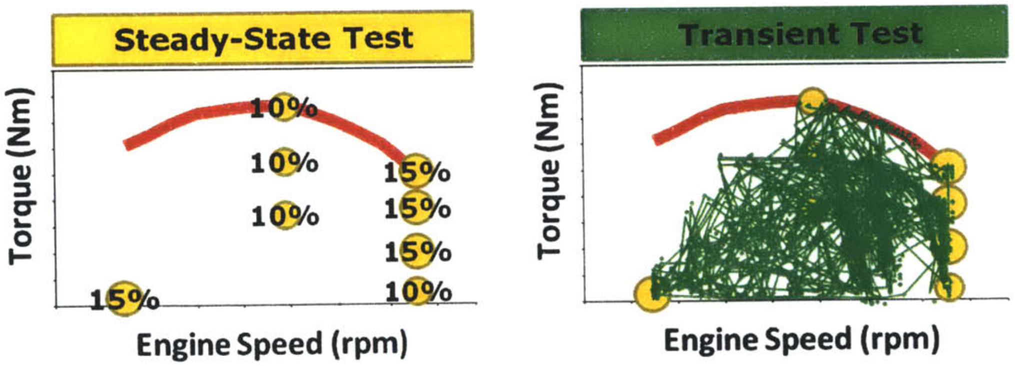


Figure 3: EPA prescribed test protocol for engine certification of emissions. The left chart shows the steady state test procedure where emissions from eight points (yellow dots) with constant speed and load are combined into a single composite emissions level. On the right, the transient test procedure shows the steady state points as a reference, but is actually the green dots connected by a line to represent the order they must run [adapted from (EPA, 2011)]

The new tier 4 emissions standard applies to new engines and does not require retrofitting of the existing vehicle fleet. Each of the new requirements drove complexities on the engines. New technology was needed to reduce the emissions from the engine (Figure 2), but also had to be capable of controlling those emission levels through transient operation (Figure 3). The dominant design architecture to control NOx emissions used in similar industries was a system that recirculated exhaust gas to the engine's intake (EGR). The amount of EGR flowing through the engine is inversely proportional to the resulting engine's NOx output. Precisely controlling this EGR flow during transient operation is challenging. The dominant solution for measuring and controlling the flow of exhaust gas was to use a measurement venturi with a delta pressure sensor across the venturi (Figure 4). This highly responsive delta pressure reading was then used, along with other measurements, to calculate the desired flow of EGR being returned to the engine intake manifold in both steady state and transient operation.

As mentioned, this technology was the dominant design in similar markets – of most interest was the on-highway heavy-duty diesel engine market. These engines were similarly sized and operated and thereby comparably regulated by the EPA. However, the on-highway market generally precedes the off-highway regulations by two to four years. Therefore, because the EGR system with measurement venturi was the dominant design on products in the on-highway market that were already in production, very little upfront analysis was done to validate this technology or the design parameters prior to its implementation within the off-highway engine system. The design revolved primarily around packaging constraints from other hardware and ultimately fitting the engine into various off-highway vehicles. Therefore, there was also a lack of requirements generated for this sub-system that could adequately guide its design at the component level.

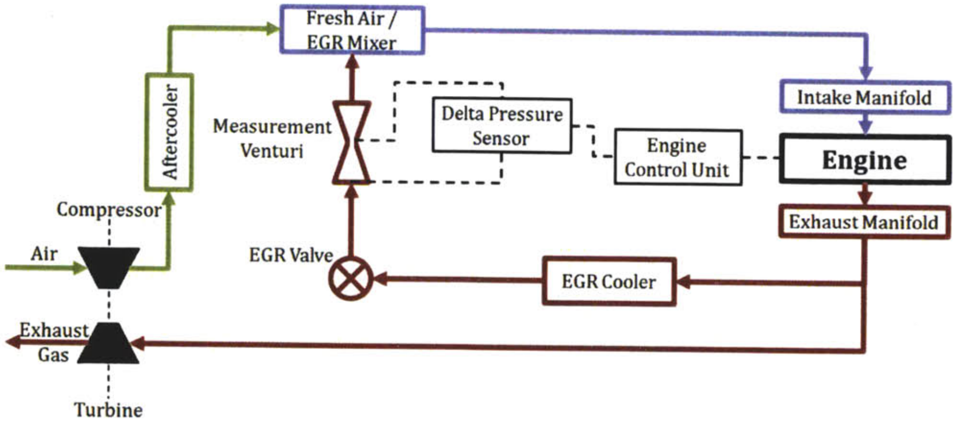


Figure 4: Air System Diagram for heavy-duty diesel engine. Air enters the system at the compressor and enters the engine by way of an aftercooler, fresh air / EGR mixer, and intake manifold. The air leaving the engine will either exit the system through the turbine or will recirculate back through the engine, re-entering at the fresh air / EGR mixer [adapted from Baert, Beckman & Veen 1999]

With this system, there are two models used in the design process that must be clarified. The first is a model that is embedded in the engine control unit (ECU). This is the model that transforms the real-time measurements from the delta pressure sensor along with other sensors to calculate real-time exhaust gas flow being returned to the engine and is then issued as an opening/closing control input signal to the EGR valve. This model is critical to real-time engine operation.

During the engine's product development process, this embedded model must be calibrated to actual performance. This can be done either by running a physical engine in a test cell or using a model to predict the engine's behavior. This off-line engine cycle simulation is the second model of interest. It runs on a computer-based model of the engine and estimates the crank-angle resolved parameters for the engine. This model is able to run many more scenarios and in a much shorter period of time as compared to running the engine in a test cell. It is this engine cycle simulation that will serve as the primary focus for this case study.

The engine cycle simulation used in this case was a simulation that was used regularly by the company. In fact, it had been used to make other design decisions on similar engines in the past. It had been validated various times against physical test data. The model was robust to changes as it was a physics-based model (as opposed to empirical), and the engineers running the model had a long history of using it and were highly qualified. Its primary fault was the lack of quantification of uncertainty bounds on the model outputs. The measures of uncertainty in the model were not aligned with outputs related to program requirements; what the program needed to understand was the impact of the EGR flow measurement on NOx levels, where the simulation provided uncertainty only in the flow of EGR itself. Although EGR flow is a leading indicator of NOx, the correlation is not fully understood and therefore adds uncertainty. The question was how does predicted EGR mass flow uncertainty propagate through the model to bound NOx emissions uncertainty.

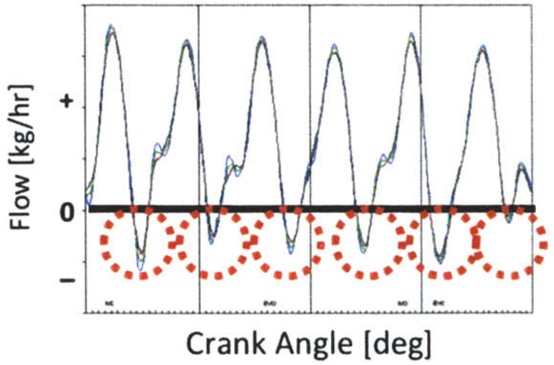


Figure 5: Mass flow of air by engine crank angle. A positive flow indicates the EGR is moving in the intended path whereas negative flow indicates the EGR is flowing backwards.

This engine cycle simulation was of particular importance during the development program of an engine for the interim tier 4 emissions standard using the design architecture in Figure 4. At the start of the program, initial design and analysis activities lead to the first set of prototype engines that could be tested in a test cell to verify the design. During these initial analysis activities, the engine cycle simulation was showing an anomaly with regard to the amount of air exiting the fresh air / EGR mixer. The first prototypes were tested for the phenomenon and high-speed data collection confirmed that fresh air in the mixer was flowing backwards into the EGR measurement venturi during favorable conditions through the engine's rotation (Figure 5). This impacts the delta pressure reading across the venturi, as it is no longer representative of forward flow through the venturi. Thus the model predicted bi-directional flow through the measurement venturi which is undesirable.

The ECU embedded model used for engine operation uses the delta pressure reading across the venturi to predict EGR flow, however because of the reverse flow phenomenon, predicted by the physics-based model, the embedded model would not function properly. This resulted in a lack of controllability for the EGR flow and therefore for NOx emissions control. The effect is portrayed in Figure 6 showing the physics-based relationship using Bernoulli's principle used to predict flow. An indicator of Reynolds number and resulting discharge coefficient are plotted for a series of data points collected from a physical engine test. For Bernoulli's principles to be valid, the fluid must be assumed as incompressible which is the region circled with a dashed line in Figure 6. However, as the data moves left on the x-axis, the relationship becomes invalid and can no longer be used. This area is shown by the circle with the solid line and is the region where back flow is occurring in the venturi. The embedded model operating in this region, therefore, cannot use the physics-based calculations and instead relies on an empirical regression relationship to determine EGR flow from the inputs. Although an approximate linear relationship can be established from Figure 6, the empirical model loses robustness as compared to physics-based models. For instance, if later design changes were to be made to the air system for improved performance or reliability, the empirical relationship would have to be recalibrated, requiring significant effort. Imagine, also, over the life cycle of the engine as component features in the air system change with time, this model would progressively drift away from optimal.

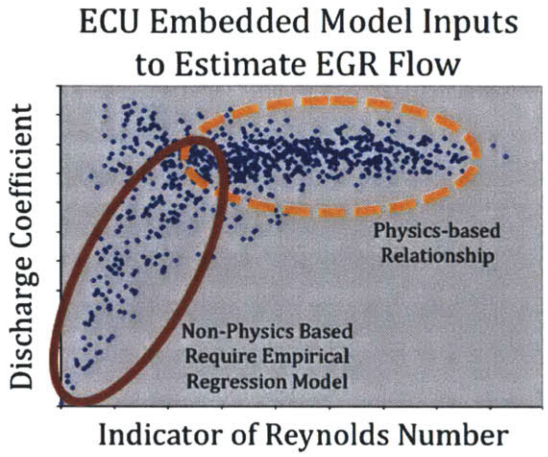


Figure 6: Effect of backflow on critical inputs to ECU embedded model. Bernoulli's principle assumes an incompressible fluid, which is represented by the region circled with the dashed line. As air begins to flow backwards through the venturi, the Reynolds number indicator shows the assumptions for Bernoulli's principle is no longer valid and an empirical regression relationship must be used.

Once this was discovered, it became a design problem: how to redesign the EGR system to eliminate this backflow effect, thereby making the ECU embedded model more robust. Because this issue was detected relatively late during engine development, crucial real estate surrounding the engine was not available to provide a lot of flexibility in potential redesign options. However, as the engine cycle simulation had found this problem initially, the model was changed to investigate different geometries and layouts that might improve the situation within the constraints of the design. Of note, of the designs that were ultimately tested, the drive cycle simulation indicated that simply adding length between the mixer and measurement venturi appeared to give the best results. Although the backflow phenomenon was not eliminated with this change, it was rare that air made it to the venturi to affect its measurement. The decision became whether to move forward with the design change as recommended by the model, or run further physical testing at the risk of making a significant design change even later in the program.

At this stage in the program, there was only one additional build remaining before production started. The decision makers had the following options, as depicted by the decision tree in Figure 7: they could proceed based on the physics-based model results alone. However, there was a chance that the model prediction could be incorrect, and in this case, the final prototype build would have a sub-standard design and there would be no opportunity to conduct another build before the tier 4 regulations would take effect. This would correspond to the situation in quadrant I behavior from the four-box model. The model would be believed to be of good quality but the predictions would ultimately turn out to be in error. However, if the decision makers chose to delay the decision until physical test results were available that would either support or refute/correct the earlier predictions made by the physics-based model, they would definitely lose several months in the schedule but could still make getting the new design on a portion of the final prototype build in order to gain experience before production.

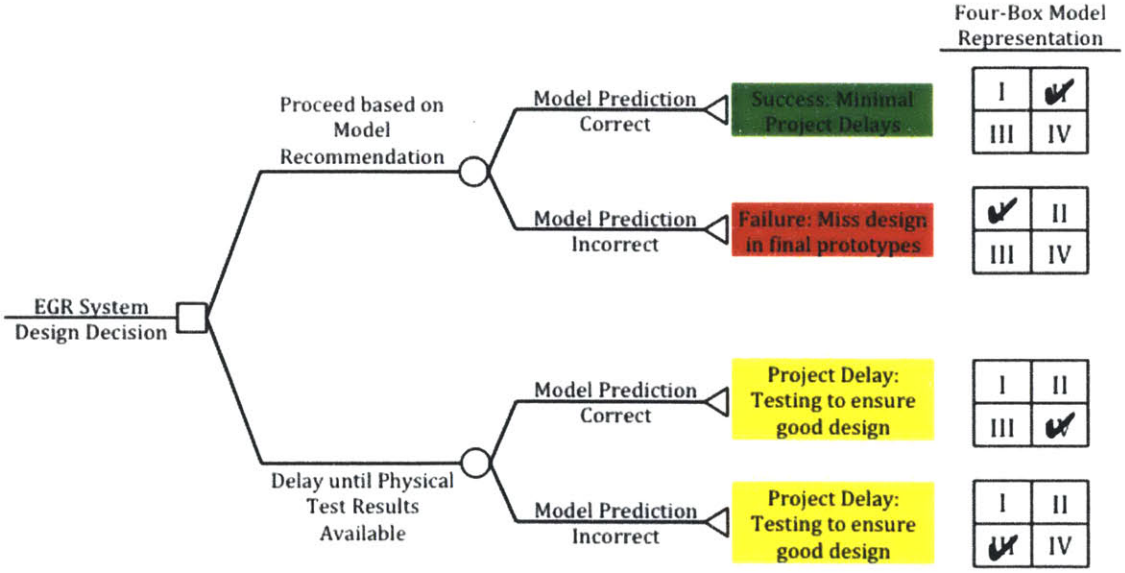

Four-Box Model Representation
Proceed based on Model Recommendation
Model Prediction Correct
Success: Minimal Project Delays
I ✓
III IV
Model Prediction Incorrect
Failure: Miss design in final prototypes
✓ II
III IV
EGR System Design Decision
Model Prediction Correct
Project Delay: Testing to ensure good design
I II
III ✓
Delay until Physical Test Results Available
Model Prediction Incorrect
Project Delay: Testing to ensure good design
I II
✓ IV

Figure 7: EGR System Design Decision Tree. There were two options for the decision, whether to proceed or to wait and pursue additional testing. In both cases, there was the possibility of the model being correct or not resulting in different outcomes shown. Each outcome is shown relative to the quadrant in the four-box model it represents.

For some industries, the prospect of physical testing is not an option, or is at most as uncertain as the model. In the diesel engine industry, however, testing is not technically difficult to do but it is costly and time-consuming. The test cells with dynamometers create fairly representative conditions to what the engine might experience in the field. However, with increasing prototype costs, fuel costs, increased instrumentation along with a larger number of tests required to qualify the growing complexity in the engine system, it is becoming more difficult to do extensive testing within the schedule and budget constraints of an engine development program. Despite the drive for increased model-based design, there still is a bias that decision-makers have in this industry towards physical testing.

In the end, the decision maker chose to physically test the design options, pursuing the third decision path shown in Figure 7. The physical test results matched what the physics-based model had already predicted. On the one hand the physical testing confirmed what had already been predicted by the model and this confirmation can be viewed as a positive in having reduced perceived risks to the program, on the other hand the physical testing did not generate substantially new information and can be viewed as a waste of resources and project schedule by introducing redundancy between model-based predictions and physical testing. Interestingly, another engine program followed this first case with a later regulation date. As it was similar hardware to the first, the design team, armed with these experiences, did early design analysis using the engine cycle simulation to determine the optimal design configuration to limit backflow. The design was accepted without testing and was found to be successful in preventing backflow once prototype engines were built. Thus, in the later engine program the situation moved from quadrant IV to quadrant II.

2.1.2 SUMMARY

The engine cycle simulation used in this case was shown with respect to the four-box model in Figure 8. The problem in this case was that of perception of model quality. The model is deemed as good; it has not been changed as a result of this incident, it predicted the right answer, and was specifically designed for these types of air system problems. However, due to pressures that affected how people perceived the model and the potential consequences of model error, its results were questioned and the program chose to run physical tests to confirm the proposed design decision, accepting a guaranteed program delay and extra costs as a result. This case demonstrates the impacts of quadrant IV where significant expenditures and schedule risk resulted from misaligned model perception and actual model quality.

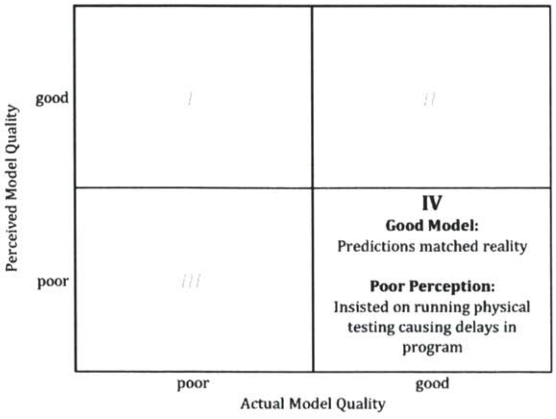


Figure 8: Four-box model of EGR Measurement case

For model-based design to be effective in this organization, there needed to be an assimilation of lessons learned on how to move to the upper right quadrant – how to know and have confidence that a model is good and act on its results; not doing so cost the program significant prototype and testing costs plus several months in the development schedule.

2.2 EYJAFJALLAJÖKULL VOLCANO ERUPTION

The next case study is, as before, an example of quadrant IV in the four-box model illustrating another example of a model providing an adequate representation of its underlying physical system, but still coming under heavy scrutiny. Besides showing another example in quadrant IV, however, this case also introduces the potential risks of quadrant I behavior where model results are used in a decision where the model is not in fact appropriate. This case is based on the Eyjafjallajökull volcanic eruption in Iceland in 2010 that closed much of European airspace when the ash cloud it emitted had spread across the continent. This case has some common features compared to the previous one, primarily in how model perception can be affected by significant exogenous pressures.

2.2.1 BACKGROUND

In April of 2010, the Eyjafjallajökull volcano in Iceland erupted. Although volcanic eruptions in Iceland are not rare, due to unfavorable atmospheric conditions, the eruption caused northern European airspace to close, affecting the major European hub airports (Bolić & Sivčev, 2011). Policy makers in Europe faced a decision that would either risk lives and equipment by continuing to fly in the ash where history had shown this to be detrimental to planes in flight (Aviation Week and Space Technology, 1990; Guffanti, Casadevall, & Budding, 2010), or to close airspace with the consequences being billions of dollars of lost revenue by the airline companies in cancellations and rerouting logistics besides the personal strife felt by the many passengers who were stranded for the weeks affected by the volcano (Ragona, Hansstein, & Mazzocchi, 2011; Ulfarsson & Unger, 2011). Due to European application of the precautionary principle (Alemanno, 2011b), guidance to policy makers was actually quite clear:

> "The recommended procedure in the case of volcanic ash is exactly the same as with low-level wind shear, regardless of ash concentration – AVOID AVOID AVOID." (ICAO, 2007)

There were two models in this case that helped to determine the areas to avoid. The first is an atmospheric model, NAME, that uses input about the eruption and meteorological data to forecast the movement of the cloud of ash from the volcano. Any region where the model said there was a concentration of ash greater than 0.2 milligrams per cubic meter (mg/m^3), the region was determined to be a no-fly zone (ICAO, 2007). This model was developed in 1986 as a result of the Chernobyl tragedy (Alemanno, 2011b) and had "evolved into an all-purpose dispersion model" (ICAO, 2007, pp. I-3-16). Although ash cloud propagation does not match the original purpose of this model, thereby bringing it into question, it had been validated against other models used by other Volcanic Ash Advisory Centers (VAAC), satellite readings, as well as physical instrumented test flights to show that it was fairly successful (Brooker, 2010) in predicting ash concentrations emitted from a point source. The model had been used regularly and was used by qualified personnel.

Despite this validation, there were still many uncertainties in this model. First, there were large uncertainties in the model's inputs. Information required about volcanic eruptions relies largely on observations which are not exact (Stohl et al., 2011). Meteorological forecasts are another primary input to the model that are notoriously uncertain. All of these inputs are translated to the model outputs by a set of complex calculations and simulations, where predictions become less reliable and less certain than the uncertainties on the input data.

The second source of uncertainty in the atmospheric model is validation against the real system. Large-scale distributed physical measurements of the ash cloud are not feasible. Satellite imagery is used to estimate it, but as imagery is not optimized for ash cloud observation, there are numerous uncertainties in that measurement as well (ICAO, 2007). As such, validating the model to the real system is difficult within close precision.

In the days following the eruption, as pressures mounted to reopen airspace, airline companies began running test flights through the regions impacted by the ash cloud. Upon their return, airlines reported no damage to the aircraft or the engines. This began to raise many questions as to the validity of the atmospheric models (Ulfarsson & Unger, 2011) and the damage threshold that was assumed to be the correct one for purposes of defining the no-fly zone.

The other model that plays a large role in this case is that of the level of ash concentration that aircraft engines can fly through without experiencing damage. This model, at the time of the eruption, was conceptual as opposed to data-based. In theory, the model would look something like what is shown in Figure 9, but as such, does not exist, or at least not in the public domain (Brooker, 2010). The absence of this model has been recognized for some time (Brooker, 2010). Although some testing has been done to understand

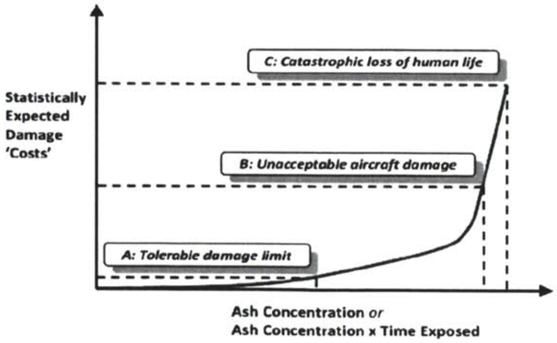


Figure 9: Ideal model of effect of ash concentration on aircraft engines. The desire is to know at what level of ash concentration (x-axis) the aircraft engine reaches different severity levels of damage against its operation (y-axis). (Brooker, 2010)

the effect of ash and dust on jet engines (Ulfarsson & Unger, 2011), the data does not support the critical specification as to the max tolerable limit (shown as "A" in Figure 9).

Despite this, five days following the initial eruption, after much coordination between European officials, airlines, and engine manufacturers, the guidance for airspace closure was modified to a tiered approach (Johnson & Jeunemaitre, 2011). Still based on the atmospheric models, an ash concentration up to 2 mg/m^3 was deemed safe. High ash concentration levels remained a no-fly zone; however, a new intermediary zone was added where it remained within the individual countries' discretion to allow flight operations. This new tiered approach allowed much of the European air space to reopen, thereby allowing some return of normalcy, however accepting some residual risk in the intermediate regime. This approach was based heavily on the conceptual model of aircraft engine's resilience to ash.

Although loss of life was avoided in this case, the financial impact to airlines and passengers alike was significant. In essence the tradeoff to be made from a financial perspective was between short-term loss of revenue due to suspended flight operations versus longer term costs due to increased engine maintenance and repair. This put a lot of pressure as to the validity of the atmospheric models being used to forecast the movement of the ash cloud and understand what portion of air space was to be closed (Alemanno, 2011b; Stohl, et al., 2011).

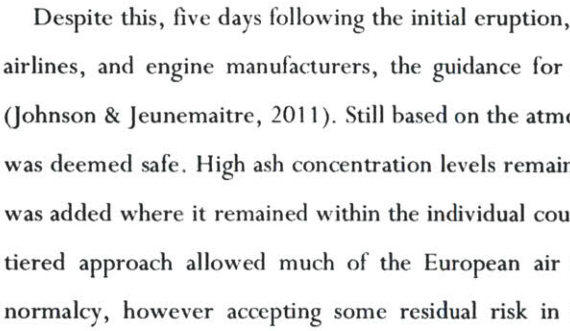


Figure 10: Comparing airline traffic in the week of the eruption and the week preceding. The number of flights following the eruption dropped to as little as 20% of normal. (Bolić & Sivčev, 2011)

The disruption due to the days of airspace closure is portrayed well in Figure 10, where the number of flights one week prior to the eruption is compared to the week following the eruption. In the end, estimates showed that US$1.7 billion was lost by airlines in revenue with 10 million passengers affected (Ragona, et al., 2011). On the other hand no flight accidents occurred that were directly attributable to volcanic ash.

2.2.2 SUMMARY

The two models used in this case present good examples of two different quadrants behavior and risks from the four-box model (Figure 11). Quadrant IV was examined in the previous case where cost and schedule were impacted negatively as a result of not sufficiently believing in the model without additional testing. Quadrant IV behavior was also reflected in the application of the atmospheric model in the days following the initial volcanic eruption. The model has a long use history and validation background, but came under heavy scrutiny as airlines began investing its validity using physical test flights which undermined

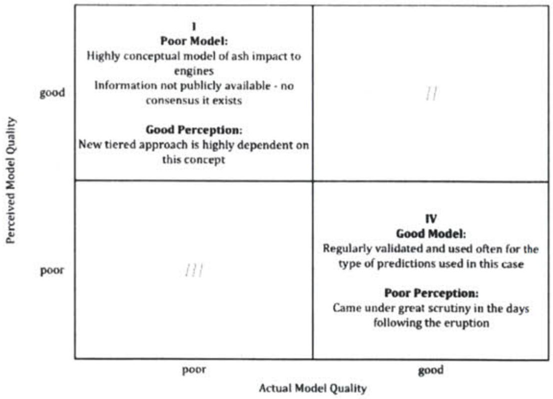

I
Poor Model:
Highly conceptual model of ash impact to engines
Information not publicly available - no consensus it exists
Good Perception:
New tiered approach is highly dependent on this concept
II
III
IV
Good Model:
Regularly validated and used often for the type of predictions used in this case
Poor Perception:
Came under great scrutiny in the days following the eruption
good
poor
Perceived Model Quality
poor
good
Actual Model Quality

Figure 11: Four-box model of Eyjafjallajökull volcano eruption case. Atmospheric model in quadrant IV, engine damage model in quadrant I.

some of the model predictions. This became a problem of perception, where confidence in the model began to wane and decision makers needed to act quickly.

The action taken, however, was based on the second model in this case, supposing the relative resilience of engines to ash clouds. The decision makers' concept of this model was enough to create new legislation on it, however, as it is still primarily conceptual in nature, very little is still understood about it. This is a problem of model validation: how do decision makers know the model is good? The potential consequences, if this model is in fact wrong, are much greater than that of quadrant IV behavior. Although there is no evidence to this effect it becomes a concern for future events.

In this case, model-based decision-making was effective in the beginning although fraught with serious consequences. The perception issues that emerged with the atmospheric model had the potential to limit the effectiveness that models can have in this kind of scenario. The validation issues that come into question on the engine damage models raise serious questions as to the capability of the new legislation, based heavily on these models, to act appropriately in the event of a future incident.

The conceptual model of the effect of ash on engine deterioration is the second model used in this case. For the purposes of classification within the 4-box model, it was rated as a poor quality model; since it is uncertain that these models exist and if they do, what the credibility of their own validation or input data looks like. However, it was the introduction of this conceptual model that turned around the crisis and brought European airspace back to normal. Therefore, despite their uncertain validation, the aircraft engine damage models were perceived well enough by policy makers to redefine policy guidance. In part this may have been influenced by the short-term financial pressures to quickly return to full flight operations.

2.3 SPACE SHUTTLE COLUMBIA TRAGEDY

The first two cases demonstrated the consequences of quadrant IV (perceived poor model quality, actually good model quality) behavior. In these cases, there was cost and schedule issues, but the consequences were not as traumatic as quadrant I (perceived good model quality, actually poor model quality) behavior could be. The Eyjafjallajökull volcano hints at the potential risks of using only conceptual models for major policy decisions but it remains to be seen if the engine models will be validated in time to prevent possible catastrophe in the future. The third and final case from the space shuttle Columbia accident will illustrate the real consequences of quadrant I behavior and will put more details behind the problem of validation. It is consequences such as these that make it a challenge to implement model-based design in support of good decision-making.

2.3.1 BACKGROUND

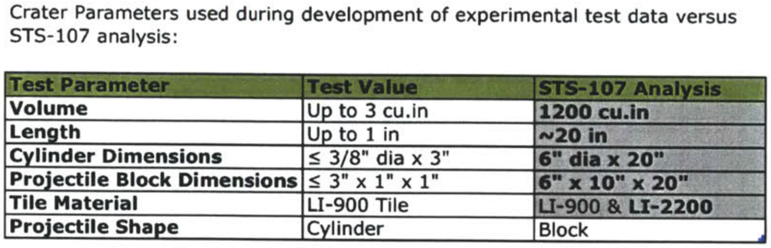
On February 1, 2003, the Columbia space shuttle reentered the Earth's atmosphere as it was returning from orbit upon completion of the STS-107 mission. During reentry, a hole in its wing caused by foam debris impact during its launch 15 days earlier caused the heat of reentry to breach the wing and destroy the structure. The space shuttle broke up during reentry, leaving only a trail of debris across the western half of the United States (CAIB, 2003).

Following this tragedy, the Columbia Accident Investigation Board (CAIB) was commissioned to investigate the details behind the accident and provide guidance to the National Aeronautics and Space Administration (NASA) as to the cause and preventive actions that could be taken to promote safe shuttle missions in the future. The CAIB produced a report seven months later that provided a comprehensive review of the history of the space shuttle program, the events leading up to Columbia's demise, and review of the organization and culture at NASA culminating in a series of recommendations. This report is the primary source of information for this case study review (CAIB, 2003).

Following the launch of the Columbia shuttle on January 16, 2003, the Intercenter Photo Working Group reviewed tapes of the launch and noticed debris hitting the space shuttle 81.7 seconds after launch. It was later determined the debris was a piece of insulating foam from the external tank that struck the leading edge of the left wing of the shuttle penetrating the structure. A series of requests were made by this working group to obtain photos of the shuttle in orbit to inspect the potential damage, but none of the requests were granted. Therefore, the Debris Assessment Team turned to models to understand potential location, type and size of damage stemming from the impact.

Many models have been used in post-analysis of the Columbia accident; however, the models of interest for this report are those used while Columbia was still in orbit – primarily the Crater model used to calculate penetration due to impact of debris on the thermal protection tiles on the shuttle. There was an additional model referred to as a Crater-like algorithm that was designed to do the same analysis with ice impacts on reinforced carbon-carbon (RCC) panels which line the leading edge of the shuttle's wing. The Crater model was developed during the Apollo program and updated for the shuttle program through 1985. The Crater-like algorithm was developed in 1984 when testing was done using ice impacts on the RCC panels (CAIB, 2003).

The primary issue with regard to Crater and the Crater-like algorithms is the difference between the empirical data used to calibrate the model as compared to the use case in the case of the Columbia shuttle. Two tables are shown in Figure 12 that show the difference between the values used to develop the Crater algorithm and its parameters limits next to those values being used to test the Columbia scenario. Emphasis has been added to show where the Columbia scenario was outside the validated region of the models.



Crater Parameters used during development of experimental test data versus STS-107 analysis:

Test Parameter	Test Value	STS-107 Analysis
Volume	Up to 3 cu.in	1200 cu.in
Length	Up to 1 in	~20 in
Cylinder Dimensions	≤ 3/8" dia x 3"	6" dia x 20"
Projectile Block Dimensions	≤ 3" x 1" x 1"	6" x 10" x 20"
Tile Material	LI-900 Tile	LI-900 & LI-2200
Projectile Shape	Cylinder	Block

Figure 12: Comparing Crater model to Columbia STS-107 Analysis. Several parameters of the model are shown with their tested values for typical operation on the left and the values used during the analysis during the STS-107 mission on the right side [adapted from (CAIB, 2003)].

The two models were clearly used well outside their original purpose. In fact, the Crater-like algorithm for the RCC panels was designed to determine necessary thickness of RCC to withstand ice impact, not to determine penetration depth (CAIB, 2003). The teams performing the analysis recognized this, but due to the lack of photographic evidence requested and absence of other certified models that were suitable for this level of analysis, it was the only scenario that could be exercised to understand the potential damage.

Despite the models being used well outside their intended region, the Crater model predicted full penetration through the thermal protection tiles due to foam impact. The Crater-like algorithm predicted that RCC penetration would occur with debris impact angles greater than 15 degrees where further analysis showed the potential for a 21-degree impact to the RCC panels causing a breach. Engineering judgment was then applied to these results to correct for the known errors in their initial usage. Although this was the first time this team was performing the analysis, it was generally known that the Crater algorithm was a "conservative" judge of tile penetration. Since it assumed constant material properties of the tile, and in reality there is increasing density of the material deeper in the structure that may hold up to impact better. These two reasons caused the Debris Assessment Team to discount the results from the Crater model. Regarding the Crater-like model for the RCC panels, a "qualitative extrapolation" was done to determine that an impact angle of 21 degrees would not cause penetration of the panel. To put it simply, the indications from the models were that a full RCC panel penetration had likely occurred, but due to the high modeling uncertainty outside the validated range, these engineering predictions were not believed and management eventually took the position that a full breach had likely not occurred and that reentry should be attempted.

Given the region the analysis was conducted relative to the calibrated inputs to the model, uncertainty became a large question. It was not understood how good the model was at predicting so far outside its validated region, but any implicit or explicit model assumptions or known uncertainties were not conveyed to the management team. And "management focused on the answer – that analysis proved there was no safety-of-flight issues – rather than concerns about the large uncertainties that may have undermined the analysis that provided that answer" (CAIB, 2003).

As management weighed the decision of what to do about the foam impact, there were several factors at play. First and foremost was the fate of the space shuttle program as a whole. After heavy budget cuts, the program had strict schedule milestones related to the International Space Station (ISS) that, if missed, could result in further budget cuts or program termination. Many of the internal communications while Columbia was still in orbit were focused more on schedule delays as a result of return-to-flight maintenance issue from damaged tiles rather than the possibility of loss of the shuttle during reentry (CAIB, 2003).

Besides the schedule pressure impacting management, tile damage due to foam shedding during launch was not a new issue. Nearly every shuttle launch experienced this as confirmed by either imagery during launch or by divots found in the tiles upon the shuttle's return, and about 10% of missions experienced shedding of the foam around the left bipod ramp of the external tank which was the source of debris on STS-107. After successful completion of a shuttle mission, concerns found from that mission are noted as In-Flight Anomalies, in some cases, the next shuttle launch cannot occur until concerns are addressed before the Flight Readiness Review. There were many instances in space shuttle history where foam shedding during launch had been made a concern, preventing flight of the next mission until resolved (Figures 6.1-6 and 6.1-7 in the CAIB report review these in detail). However, the many changes made to reduce exposure were enough to continue with the next mission, but not enough to eliminate the issue altogether.

The space shuttle Atlantis flew mission STS-112 that was the latest launch to experience significant foam loss prior to STS-107, and it was from the same location and preceded by only 3.5 months. The damage from the foam loss was significant, but this was the first time the incident was deemed an "action" as opposed to an In-Flight Anomaly. This then allowed the following shuttles to launch without steps being taken to solve the foam shedding issue. The CAIB report states "this decision ... is among the most directly linked to the STS-107 accident. Had the foam loss during STS-112 been classified as a more serious threat, managers might have responded differently when they heard about the foam strike on STS-107" (CAIB, 2003). As stated by Woods (2005), they were using "past success as a reason for confidence" (p. 9).

First, it was not believed that foam could ever penetrate an RCC panel entirely; to the point that the CAIB had to "prove or disprove the impression" which "prompted the investigation to develop computer models for foam impacts and undertake an impact-testing program" (p. 78) in the post-accident analysis. When STS-107 was still in flight, this same sentiment was noted by the CAIB: "Analysts on the Debris Assessment Team were in the unenviable position of wanting images to more accurately assess damage while simultaneously needing to prove to Program managers, as a result of their assessment, that there was a need for images in the first place" (p. 157). The management team could not or did not want to believe that full RCC panel penetration could occur as a result of foam loss, and therefore because of their inherent bias tended to reduce their perception of quality of analysis despite the fact that the engineering community stated otherwise.

2.3.2 SUMMARY

The events surrounding the Columbia space shuttle's final flight can be used to show how issues with model validation can impact the quality of information available to make decisions, and how factors affecting perception of quality can further impact the decision making process.

It is interesting to plot this case study on the four-box model presented earlier and shown again here (Figure 13). In this case study example, there are potentially two ways to plot it depending on the point in time of this case. Quadrant I indicates the situation while the shuttle was in orbit: a poor model that was perceived well. There are numerous examples of how the actual quality of the models used in this example are poor (using the language from the NASA-STD-7009): Input pedigree, Uncertainty, People Qualifications, M&S Management, Use History. However, as a result of overriding factors related to people's perception of quality, such as the pressure from schedule deadlines, the potential consequences of the decision, or the lack of consistent communication across the organization, the model was deemed good enough to pursue reentry. This is particularly interesting when taken in context with other models being available at that time that could better address these validation concerns with the model, but were not certified by NASA for use.

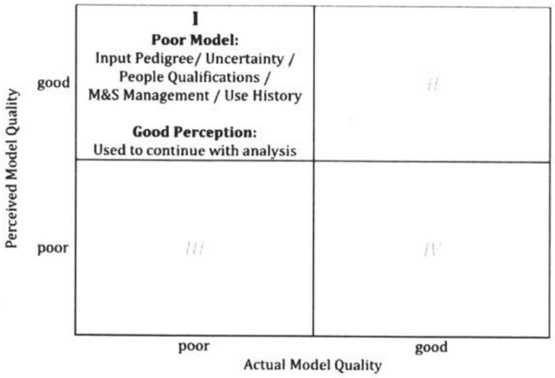


Figure 13: Space Shuttle Columbia Accident plotted on four-box model

Using the benefit of hindsight, as actual model quality is rarely if ever known ahead of time, this case could also be classified in Quadrant IV where the quality of the model is good, i.e. it correctly predicted full RCC panel penetration, but it was perceived poorly. Despite the various issues with the models used, the resulting prediction it gave was in fact representative of reality where full penetration was achieved. Despite this, the management team's perception of the model was to discount its findings due to inevitable model uncertainty, not based on the model's quality itself.

3 ANALYSIS OF THE PROBLEM

3.1 PROBLEM STATEMENT

For model-based design to be effective, it is necessary for models to be validated and appropriate for their use and for the decision makers to trust the results. The previous case studies demonstrated the consequences if either of these two criteria are not met. In the first and second cases, decision makers sought other sources of information from which to make a decision usually costing time and money. In the last case study, there is potential for dire consequences upon making the wrong decision from misleading model results.

In the case study review, where there is the benefit of hindsight, it was clear where there was a problem of perception in the first cases where models were actually providing appropriate information but decision makers required more confidence before acting or did not like the outputs of the models because it contradicted their pre-conceived notions and biases. In the final case, there was a clear problem of validation where a model was used that misled decision makers, resulting in tragedy. Given these consequences, for model-based design to work successfully in organizations (i.e. resulting in most or all cases that reside in quadrants II and III), decision makers need to know whether to believe the model results as the consequences otherwise are too great.

Visualizing the problem space will help to distinguish where these problems originate. Figure 14 shows a generic decision support system (DSS) whereby a model and data become a DSS that interacts with a user or set of users, generally the decision maker (Bonczek, et al., 1980). In this system, the user will "[search] for information about the current and desired state of affairs, [invent] possible courses of action, and [explore] the impact of each possible course of action" (Brennan & Elam, 1986, p. 49) with the help of the model and associated data. It is the interfaces in this system where the problems originate. In the interface between model and data within the DSS, there is the problem of validation whereas the interface between the user and model

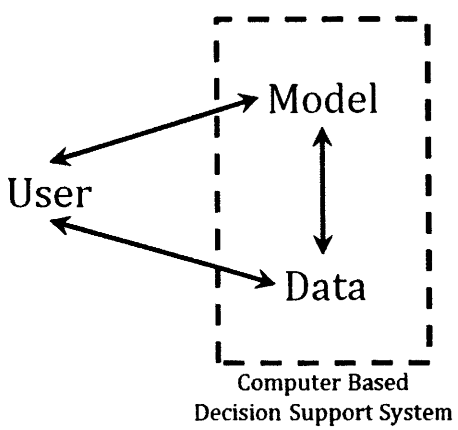


Figure 14: Interfaces in a Decision Support System. [Adapted from(Bonczek, Holsapple, & Whinston, 1980)]

introduces the problem of perception. Before a solution to these interface problems can be provided, the potential underlying causes must first be uncovered. The following sections will seek to do so by first ensuring a definitive understanding of the aspects related to model-based design. This includes identifying what defines a model and which models are relevant to this thesis. In addition, attention will be given to defining model validation and how it is different from other similar terms often used such as verification, assessment, and confidence.

Following these definitions, the problem of validation will be further investigated by first understanding some of the challenges to validating models. This will be followed by some prescriptive techniques available in the literature to help address the problem of validation. The problem of perception will also be analyzed further by first understanding factors that can influence the perception process specifically as it relates to model quality, which will illustrate how perceived quality may differ from actual quality.

3.2 DEFINITIONS RELATED TO MODELS AND SIMULATIONS

Confusion in the problem space can originate by what is meant by model, simulation, verification, or validation (Brugnach, Tagg, Keil, & de Lange, 2007; Oreskes, Shrader-Frechette, & Belitz, 1994). This section will review some of the definitions from the literature to both define the scope of this thesis in the case of models and simulations, and to present the varied field that is verification and validation.

The concept of a model can be quite varied. The exhibit "Making Models" at the Museum of Science in Boston illustrates the vast definitions of models. They can be physical, conceptual, mathematical, computer simulations (Boston Museum of Science, 2001) and more. The exhibit has fundamental concepts it intends to teach its clientele about models: that a model is not the real system, but can be used to learn more about it, that models can be effective communication tools, and that the usefulness of a model can be determined by comparing its predictions to actual observations (AAAS, 1993).

3.2.1 DEFINITION AND SCOPE OF MODELS AND SIMULATIONS

Although the number and types of models are diverse, the definition of a model is concise and can be used to describe any model. Highland (1973) summarized many definitions into simply "a model may be defined as a replication of a real world entity" (p. 11) and describing further "no model is a complete and true representation when we attempt to model a real world entity ... at best it is a simplified version of the real world" (p. 12).

Some try to further define models beyond this broad definition. In many cases, this becomes a model classification to be discussed later in this section. However, Refsgaard and Henriksen (2004) recognize the importance in distinguishing between the concept of a model encapsulating the governing theories, the code that implements the model in a software package, and finally the "site-specific" model which is generally what most consider to be the final product. Although this definition is more confined to a class of models that are computerized and mathematical, it is helpful to recognize these distinctions when we define processes that add to a model's credibility.

A simulation is defined by Forrester (1961) as the "process of conducting experiments on a model instead of attempting the experiments with the real system" (p. 18). Simulations allow for better understanding of the problem being addressed and the full design space. They can run "what-if" scenarios to explore system responses in cases where it is prohibitive or impossible to do on the real system (Banks, 1998).

By the definitions described here, the model would be considered the operand or instrument and the simulation the operation or process. However, throughout the literature as well as this thesis, the term "model" is often used in place of "simulation." In many cases, rather than provide a distinction, the terms are lumped into a single abbreviation: M&S.

This thesis cannot purport to be applicable to any model or simulation. This difficulty of defining scope was realized by Bertch, Zang, and Steele (2008) during development of their model validation standard. They noted that there are numerous types of models and simulations in use at NASA and validation efforts were in place that were specific to individual types of models. However, they were tasked with providing a broad enough approach to encapsulate all models and simulations used at NASA. Their work on developing a NASA-wide standard for validation and credibility of models and simulations, which resulted in NASA-STD-7009, was a direct outgrowth from the CAIB final report.

Gass and Thompson (1980) scoped the problem well: "The types of models considered as the primary basis for developing these guidelines have the following general characteristics: (1) they are models that are developed to assist the policy analyst or decision-maker in selecting or evaluating various policies regarding governmental issues and programs ... (2) they are mathematical models of a complex system and have been computerized; and (3) they are large scale models." The applicability of this thesis follows similar guidelines as above.

It is most important to remember that models and simulations are simplifications of reality. They are tools used to support decisions required for the real systems they represent. However, as they are simply a representation of real systems, they are all, in effect, wrong (Box, 1979). Whether or not models are in fact wrong or not depends on whether the predictions and recommendations flowing from these predictions would lead a rational decision maker to take a course of action that was – in hindsight – judged to be right or wrong. This of course is fraught with difficulty since even with hindsight different stakeholders – presented with the same facts and outcomes – may evaluate the decision that was actually taken quite differently. The key is to determine which models are useful by means of processes to enhance and understand their credibility.

3.3 MODEL VALIDATION AND RELATED TERMINOLOGY

A common expression in any realm that works with models is that "all models are wrong, but some are useful" (Box, 1979). The processes of verification and validation describe the activities that move a model from the first part of Box's statement, to the latter part. What makes a model useful, however, is a fuzzy proposition making the verification and validation processes difficult to define and execute with a consistent result in the end. Yet these processes are the means by which credibility is attributed to a model and therefore a critical component to impacting the perception of a model's quality. Therefore, this section will discuss in detail what is actually meant by verification, validation and related terminology, as there are some important distinctions as well as some debate on the subject.

It is helpful to see details of the modeling process to better understand how the processes of verification and validation come into play. Sargent (2001) presents a descriptive paradigm of the modeling process (Figure 15) that has enough detail to show the intricacies that distinguish the verification and validation processes. In this paradigm, he shows two realms: the real world and the simulation world. The real world shows the actual system and its results in physical reality while the simulation world shows the various processes and stages required to convert a real system into a model or simulation. These include abstracting a system in the real world into system theories that serve as the interface between the two realms. Using these theories, a modeler generates a conceptual model of the system, which is then converted into a specification illustrating how the conceptual model is to be implemented. At which point, the model is implemented into computer code where it can then be used and repeatedly exercised to generate data that reflect back to the real system.

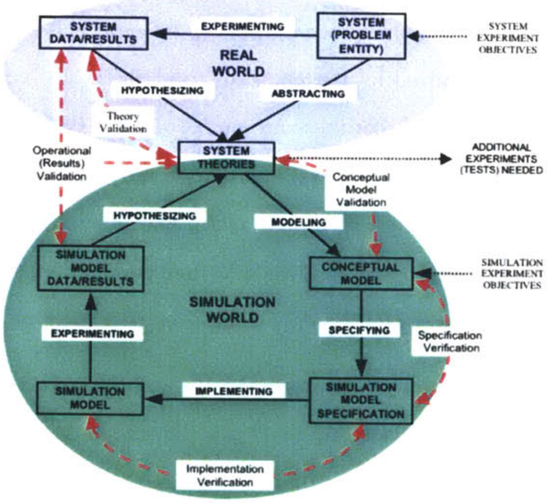

SYSTEM DATA/RESULTS
EXPERIMENTING
SYSTEM (PROBLEM ENTITY)
SYSTEM EXPERIMENT OBJECTIVES
REAL WORLD
HYPOTHESIZING
ABSTRACTING
Theory Validation
Operational (Results) Validation
SYSTEM THEORIES
ADDITIONAL EXPERIMENTS (TESTS) NEEDED
Conceptual Model Validation
HYPOTHESIZING
MODELING
SIMULATION MODEL DATA/RESULTS
CONCEPTUAL MODEL
SIMULATION EXPERIMENT OBJECTIVES
SIMULATION WORLD
EXPERIMENTING
SPECIFYING
Specification Verification
SIMULATION MODEL
IMPLEMENTING
SIMULATION MODEL SPECIFICATION
Implementation Verification

Figure 15: Sargent's detailed paradigm of the modeling process including the simulation world and how it interfaces to the real world. (Robert G. Sargent, 2001)

Verification is the process to understand "did I build the thing right?" (INCOSE, 2011; Pace, 2004). It is generally accepted that to verify a model, is to confirm that the model, as implemented generally in software code, correctly instantiates the originating conceptual model (Balci, 2004; Banks, 1998; Gass & Thompson, 1980; Refsgaard & Henriksen, 2004; Robert G. Sargent, 2001; SCS, 1979). In this more detailed paradigm in Figure 15, verification is done in two steps. First, by verifying the specification for a model to the conceptual model, then by verifying the model to its specification.

It is common in publications and particularly in industry that verification and validation are used interchangeably (Naylor, Finger, McKenney, Schrank, & Holt, 1967; Refsgaard & Henriksen, 2004). In what is a minor change in wording from above, validation is the process to understand "did I build the right thing?" (INCOSE, 2011; Pace, 2004) which is a significantly different proposition. The difference is illustrated in Figure 15 where the model is being checked against reality in three different manners: the governing system theories to the system behavior, the conceptual model to the system theories, and finally, the simulation results to the real system behavior. In this point of reference, much of the literature describes validation as a behavioral assessment of the model to the real system (Balci, 2003; Lewandowski, 1981; Romero, 2007). Banks (1998) goes on to say "whether the conceptual model can be substituted for the real system for the purposes of experimentation."

Although the paradigm from Sargent helps to define the verification and validation processes, there is still some debate over what these processes mean. Oreskes et al. (1994) disagree with the common definition of verification used in the modeling community. In their opinion, verification "is to say that its truth has been demonstrated, which implies its reliability as a basis for decision-making." Further in their discussion, no open system can ever demonstrate truth in this manner as the cause and effect to a variable is not strictly dependent on other known variables. Refsgaard and Henriksen (2004) discuss the issues in defining verification as presented by Oreskes and others and tried to address the issue by separating the model's computerized code from the model itself. In so doing, they have created a closed system between the requirements created by the concept to those implemented in the code. With this distinction, there is no assurance given by the verification process that the model represents reality, only that it correctly represents the abstraction of reality embodied in the conceptual model.

Oreskes et al (1994) also warn against the common use of validation by stating that substituting a model for the real system may mislead what people believe is reality. They define validation as a process to ensure the absence of flaws in the model. Many of the processes defined by others to complete validation – those of comparing model outputs to reality – are further criticized by Oreskes, stating that "congruence between a numerical and an analytical solution entails nothing about the correspondence of either one to material reality."

Under this premise, how can one determine if the model truly corresponds to the system it intends to represent? Oreskes suggests confirmation of the model whereby "if a model fails to reproduce observed data, then we know that the model is faulty in some way, but the reverse is never the case." Forrester and Senge (1980) further elaborate this concept as building confidence, which "accumulates gradually as the model passes more tests and as new points of correspondence between the model and empirical reality are identified." Therefore, successful validation processes result not in a model that mimics reality, but rather one that builds confidence that the model is an appropriate tool to use with the real system.

Confidence is not an inherent attribute of the model, but rather an attribute of the user or decision maker (Gass & Joel, 1981). Therefore, it is necessary to distinguish between credibility and confidence. Confidence is a measure of whether the model should be believed. Gass and Joel define confidence as "the user's total attitude toward the model and of the willingness to employ its results in making decisions" and "is expressed by the influence the model's outputs had in the decision" (p. 341). Credibility is developing the potential to build confidence in the model (Robert G. Sargent, 2005). Under these definitions then, the validation process builds model credibility. This in turn may improve the confidence of the decision maker in the model and increase the probability that the model will indeed be used to influence real decisions. However, as will be shown later in this thesis there are factors such as time pressure to make a decision or consequences from an erroneous decisions that can influence a user's confidence in a model, even though the attributes and credibility of the underlying model may be invariant to these exogenous factors.

The final terms that are used often in the literature are accreditation and certification describing a formal process that results in a recorded approval that a certain model is indeed fit for use. Generally, accreditation refers to the result that the model meets a set of acceptability criteria (Balci, 2004; Pace, 2004; Robert G. Sargent, 2001). However, Sargent discusses some arguments in this definition where accreditation may be taken to mean the procedure to give a third party the ability to certify models. Here, then, certification becomes the official documentation that a model meets its criteria. This distinction becomes important as the literature describes the importance of using a third party to validate models in order to enhance their credibility (Robert G. Sargent, 2001). In general, however, there is terminology in use to describe the documentation of a model meeting its specifications.

Having a solid understanding of the terminology related to the modeling process is key, and a summary of definitions is shown in Table 1. This will help to introduce the idea of confidence, a key element required to model perception.

Terminology:	Definition:
Verification	The model meets its specifications
Validation	The model represents the real system
Credibility	Potential to build confidence in the model
Accreditation	Model meets given acceptability criteria
Certification	Documentation of completed model verification, validation, and accreditation process

Table 1: Summary of terminology related to model-based design

3.4 MODEL AND SIMULATION CREDIBILITY ASSESSMENT

In Sargent's illustration of the modeling process (Figure 15), the validation processes are shown relative to their integration within the process but with no guidance as to how to complete those validation processes and how to do them in such a way as to build the needed credibility in the model. This section will describe first what the assessment or validation process is and why it is important. This will be followed by discussion on the importance of model classification in assessment processes finishing with some example assessment frameworks from the literature.

Forrester and Senge (1980) noted that the decision-makers using the models are within the system boundary. Therefore, "validation includes the communication process in which the model builder ... must communicate the bases for confidence in a model to a target audience. Unless the modeler's confidence in a model can be transferred, the potential of a model to enhance understanding and lead to more effective policies will not be realized."

Brugnach et al (2007) noted the resulting issues that arise as a result of this lack of communication including "policy makers do not understand models," "lack of certainty or validation of the models," "lack of integration of policy makers and modellers," and a "lack of stakeholder involvement in the whole modelling process." Yet despite these issues being known, modelers believe they are not being listened to, while the decision makers "do not hear much they want to listen to" (Clark & Majone, 1985; Lindblom & Cohen, 1979; Weiss & Bucuvalas, 1977).

The assessment process has been established to provide "a practice intended to enhance societal understanding of the broad implications of science and technology and, thereby, to improve decision-making" (Sclove, 2010). It is meant to be a vehicle with which modelers can input pertinent information regarding the model and decision makers will have the necessary information with which to make decisions.

Therefore, it is important to remember in the following section that a key aspect to assessment processes is communication. In the following sections, several frameworks will be presented that will describe technical aspects of model validation, but if the resulting credibility of the model does not get communicated well to the decision maker, then the process loses value or breaks down entirely.

3.4.1 VALIDATION DEPENDENCE ON MODEL CLASSIFICATION

Classifying models is important from various aspects. First, it can help to organize a field that is incredibly diverse. Second, frameworks to verify and validate models and simulations can be defined differently depending on how the model is classified (Lewandowski, 1981; Oren, 1981). The potential options in model classification are vast (Highland, 1973), and therefore a list of examples shown below is not complete.

However, it is meant to demonstrate the breadth of scope used to classify models to help understand how validation techniques might vary among classifications.

Oren (1977) introduces a concept that is more a taxonomy of model classifications as compared to a taxonomy of models themselves. However, it illustrates the potentially numerous ways that models can be classified. Each of these methods may require a different framework for validation of the models within it as it values different characteristics of the models differently.

- Goal of experimentation or the model's purpose
- Type of application area
- Type of system
- Nature of the model
- Nature of the relationships
- Time set of the model
- Ratio of simulated to real time
- How the state of the model is updated
- Device used to do the experimentation
- Way of accessing the computer in computerized simulation
- Simulation executive (time structure of the simulating software)

3.4.1.1 Classification based on System Behavior

Forrester utilized a model classification in order to identify the current state of modeling and illustrate gaps in the field. In so doing, he identified a segment of models that most closely represented organizational behavior, but was not captured by existing models; hence became the origin of system dynamics modeling (Forrester, 1961).

Forrester used a hierarchy of fundamental attributes of a model's behavior. He was interested in using models to describe how systems respond and therefore emphasis was placed on classifying the outputs of the models as opposed to its inputs or its attributes. The structure of his resulting taxonomy is shown in Figure 16.

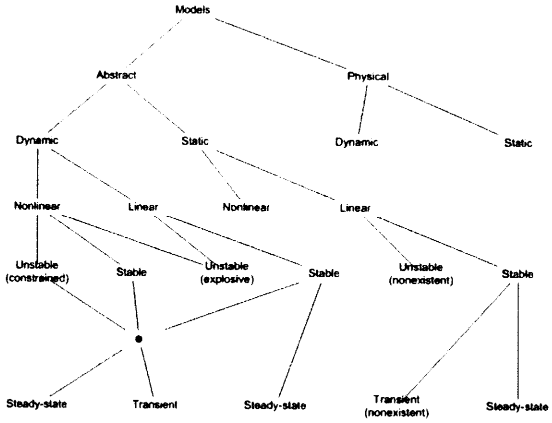


Figure 16: Forrester's Model Taxonomy. A hierarchy of attributes can be used to describe model behavior (Forrester, 1961)

3.4.1.2 Classification Based on Model Architecture

Highland (1973) begins his discussion of model taxonomy with a review of existing classifications at the time of his publication. From this review, which included Forrester above, he proposed a new taxonomy that would provide a set of identifying features for improved communication of the model. The result is a system of classifications that is based on a hierarchical architecture of the model, beginning at the highest level of abstraction and decomposing down to the variables of the model.

He first identifies the broad classification to start similar to how Forrester had done. He then diverges to classifying the purpose or function of the model. This is followed by a classification by relationships of the modules within the model to gain an understanding of complexity. Finally at the lowest level of abstraction, the system variables are classified by their attributes.

3.4.1.3 Model Classification for Validation

In Lewandowski's (1981) discussion on issues in model validation, he pointed out the importance of specifying a model in regards to attributes that are most relevant to validation of that model. He identified three attributes each with two contrasting levels that could best describe a model in ways that could best prescribe the method for validation required.

- Model Background
    - Natural
    - Behavioral
- Logical Type of the Model
    - Causal
    - Descriptive
- Interpretative Type
    - Probabilistic
    - Deterministic

3.4.1.4 Validation dependence on model classification

The validation processes available can depend on the way a model is classified. Each of the above examples of model classification presents its own challenges to validation. Kleindorfer and Ganeshan (1993) discuss two primary segmentations of models (with a third as a mix of the first two), each with their own method of validation. Justificationism describes the belief that a model has a firm grounding to experimentation or sound theories. Therefore, validation of such models is tying the experimentation to the models. In contrast, antijustificationism is the belief that judgment is required, that empirical validation, as is done with the first theory, is not possible with models. In this case, validation "consists of persuading someone that one's model falls into a well-accepted way of seeing a problem." Another perspective is to validate by forcing "the system that it describes to act like the model." As seen from their work, validation techniques can vary significantly based on how the model is viewed.

Much of the work above was inspired and developed from some of the original work by Naylor et al (1967). They identified three ways of viewing models: Rationalism, Empiricism, and Positive Economics. In the first case, rationalism is a model comprised of logical deductions. Therefore, validation in this case is dependent upon aligning the model with first principle theories. In the case of empiricism, the models are data based as opposed to theory based, and therefore validation is a matter of aligning the data from the model to the real system. The final case, positive economics, describes the model's ability to predict future behavior of the system. This focuses validation on matching output behaviors as opposed to theories and data as in the first two cases.

Voas (1998) introduces another perspective to approaching validation; in this case, the interest is in certification. They decompose the problem into three segments: Product, Process, and Personnel. With each segment, the way in which certification is done is different. Although their paper focuses more on certification, the ties to validation have been made by Balci (2004) who presents quality indicators for each of the segments. These indicators are used then to provide an assessment for the model, its processes, and its developers and users. More on Balci's framework will be discussed in the next section.

Another presentation of validation techniques is based on the purpose of the model. This technique is particularly interesting as general validation theory suggests that validation is done as to the purpose of the model. Therefore, it is important to understand how the validation may change as the purpose does. Lewandowski (1981) provides four potential purposes, each with varying validation challenges.

The first purpose is to gain understanding of the real system on which a model is based. In this case, the primary challenge is "the relationship between the structure of the process and the structure of the model." The importance of understanding the impact of model assumptions is stressed as a primary goal to be sure the model is an appropriate representation for understanding.

The next purpose may be for prediction or forecasting. Lewandowski describes this purpose being "the most frequent situations, and probably the most difficult ... from the point of view of validation approach." It is described that even if a model can be validated well against its reference data, it is difficult to assume that the model will behave well outside the validated region as there may enter in new parameters that were not a factor before. This was clearly the case in the Columbia space shuttle case study.

Similar to prediction, models can be used for scenario analysis. For this case, a model is used to view future system behavior, but based on predetermined scenarios. According to Lewandowski, "the methodology for validation of scenario models does, as yet, not exist."

The final purpose presented by Lewandowski is using models for optimization. As discussed, there are several variations of optimization, but in all cases, an objective function is developed to describe the system, and the validation of that function is most essential. Proposals are made to use data, if available, or possibly subject matter experts who have a better understanding of the system behavior.

The Department of Defense, in its Verification, Validation, and Accreditation guidelines (DoD, 2006) provides one final purpose on which validation may be dependent. In many of their cases, training is an important purpose of a model. In this case, validation is based on activity being trained and the accuracy required for that activity.

This section has illustrated the vast field of model validation. Defining a fixed process is difficult as the model type or purpose for which it is used is different (Bertch, et al., 2008; Kleindorfer & Geneshan, 1993; Lewandowski, 1981; Naylor, et al., 1967). Add to this the complexity surrounding the definition of model validation, and a large, varied field emerges in attempting to understand how to validate models in order to boost their usefulness in the decision making process.

3.4.2 EXISTING FRAMEWORKS

A number of different ways to validate models have been presented in the above sections. This section will describe some specific assessment frameworks from the literature. In most of these cases, some of the techniques described above are simply a subset of the validation process which typically can include more activities. For example, Figure 17 shows the Problem Solving Process published in the Department of Defense (DoD) in the Recommended Practices Guide (RPG) (DoD, 2006). This process represents the system of activities relating to using models and simulations, including a section that addresses the Verification and Validation processes. Other cases include more aspects of the problem solving process than just focusing on the model development. The frameworks presented in this section will begin with fundamental model validation steps followed by examples that increase the detail against which models are judged in order to demonstrate the potential breadth of simple model validation. However, as mentioned above, a critical step to the assessment process includes communicating with the decision makers; two frameworks will be presented that introduce attributes of the communication process within their model assessment process. To conclude, three additional frameworks will be discussed that further attempt to build upon the communication process by introducing a numerical rating of models that results from the assessment process.

3.4.2.1 Fundamental Model Validation Steps

Naylor et al (1967) present a "multi-stage" process to validation. It consists of three steps:

1. Formulate hypotheses describing system behavior
2. Using statistics, validate these hypotheses to the system (recognizing the challenge in some cases to have data available with which to do this)
3. Test the ability of the model to predict system behavior

This framework demonstrates the fundamental model validation process. It provides a generic roadmap, but does not give much guidance as to the details required for each step.

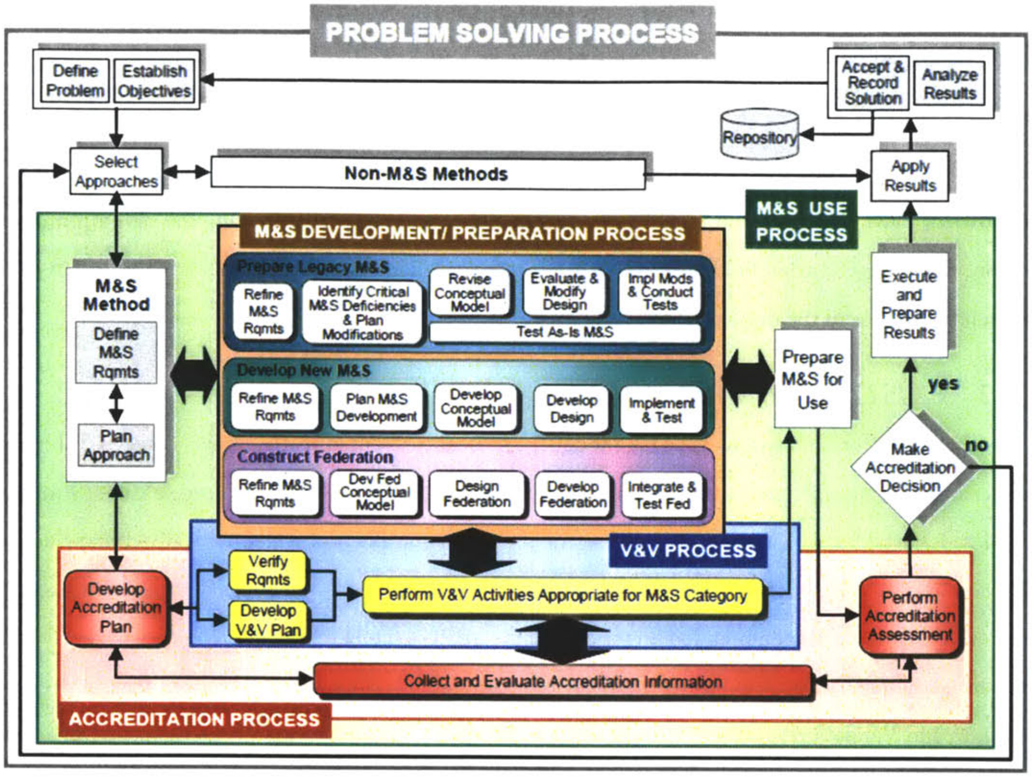


Figure 17: DoD Problem Solving Process including Verification and Validation steps (DoD, 2006). This process illustrates the series of activities around problem solving of which verification and validation is a part. It is a large part of the accreditation process but also overlaps with model and simulation development and use processes.

Oren (1981) presents a much more detailed process of model validation. He identifies a matrix (Figure 18) of required assessment linking the aspects of the model to be validated and the criteria to which it should be validated against. The checkmarks indicate that an aspect of the model should be judged against the criteria along the top. For example, the data from the real system should be judged against the goal of the study as well as the norms of experimentation technique likely used to collect data from the system.

For each of the aspects of the model, a series of deeper criteria are provided for each of the intersections in the matrix. These criteria are meant to be at levels that are individually assessable.

Although Oren's framework is much more descriptive than that shown by Naylor et al, it remains at a relatively abstract level related to model validation.

Balci (2004) presented an extensive framework for model quality assessment. The framework was developed around Voas' (1998) segmentation of the modeling environment which involves discerning between the model product, the modeling process, and the project.

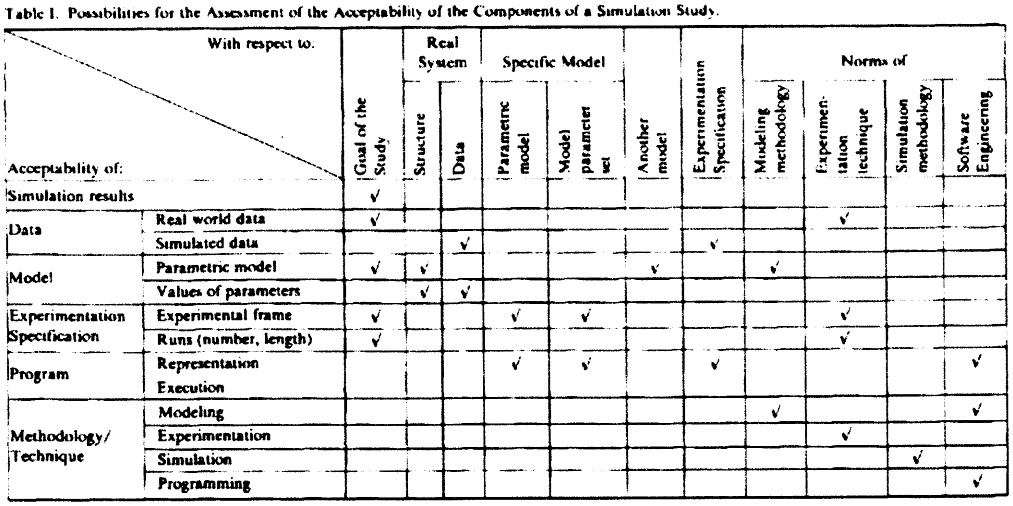


Table I. Possibilities for the Assessment of the Acceptability of the Components of a Simulation Study.

Acceptability of:	With respect to:		Real System		Specific Model				Norms of			
		Goal of the Study	Structure	Data	Parametric model	Model parameter set	Another model	Experimentation Specification	Modeling methodology	Experimentation technique	Simulation methodology	Software Engineering
Simulation results		√										
Data	Real world data	√								√		
	Simulated data			√				√				
Model	Parametric model	√	√				√		√			
	Values of parameters		√	√								
Experimentation Specification	Experimental frame	√			√	√				√		
	Runs (number, length)	√								√		
Program	Representation Execution				√	√		√				√
Methodology/ Technique	Modeling								√			√
	Experimentation									√		
	Simulation										√	
	Programming											√

Figure 18: Oren's Model Validation Framework (Oren, 1981). The rows of the table show the attributes of the problem where the columns across represent measures by which the attributes can be tested.

The product quality framework presented by Balci recommends creating a hierarchy of quality indicators related to the product. Similar to Oren, the indicators are decomposed to a level that is assessable. He presents an example including the following high-level indicators:

- Dependability
- Functionality
- Performance
- Supportability
- Usability

Balci next proposes indicators to evaluate the modeling process:

- Acceptability
- Maintainability
- Quality of the methodology used to execute the process
- Quality of the way the methodology is applied
- Rapidity
- Reliability
- Robustness
- Supportability
- Understandability
- Visibility

Finally, the project indicators are proposed where the maturity of an organization is graded. For this, Balci borrows from CMMI (2012).

Balci, like Oren, provides a very thorough framework for model assessment with respect to verification, validation, and quality assessment particularly as Balci has done for the three aspects of the modeling environment as defined by Voas. What these frameworks lack is inclusion of the communication process amongst the various roles in the modeling process without which makes the assessment process incomplete.

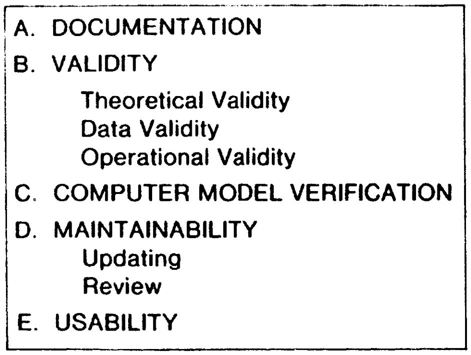
3.4.2.2 Assessment Frameworks with Communication

The importance of the communication process is recognized and accounted for in other frameworks; two examples are presented in this section to discuss the means used to improve communication. Sargent (2001) introduces a recommended framework for model validation that begins to include elements critical to communication between the modelers and decision makers.

1. First agreement between the modelers, sponsors, and users on the validation approach and minimum techniques to be used
2. Specify the level of accuracy required for the problem
3. Test assumptions and theories of the model
4. Continually validate the conceptual model throughout development
5. Explore the model's behavior throughout development
6. Compare model behavior to system behavior in at least the last revision
7. Develop documentation of the validation process
8. Propose a schedule to periodically review the model's validity

Sargent provides much less detail compared to the earlier frameworks as to some of the specific validation procedures to conduct. Although early agreements regarding model assessment may have been achieved, later documentation and scheduling are critical to communication between the modelers and decision makers. Extensive model validation may be of little use if the results of the model validation, including a clear recording of the model's limitations, are not subsequently communicated orally and in writing or in electronic form to subsequent users and consumers of the information produced by the models.

A framework presented by Gass and Thompson (1980) (Figure 19) is very similar to Sargent's above. There is emphasis placed on documentation, maintainability, and usability. However, more guidance is given to the documentation, validation and verification aspects.



A. DOCUMENTATION

B. VALIDITY

Theoretical Validity
Data Validity
Operational Validity

C. COMPUTER MODEL VERIFICATION

D. MAINTAINABILITY

Updating
Review

E. USABILITY

Figure 19: Gass & Thompson's Validation Framework (Gass & Thompson, 1980)

The documentation of the model is called out first and the connotation is different from the above. Gass and Thompson describe two kinds of documentation: descriptive of the assumptions and theories behind the model, and technical description of the modeling methods and its software implementation. Although this does not callout gaining agreement at the start, it specifies details that need to be agreed upon in the early stages of model development.

The validity and verification sections also provide more guidance; not to the detail of Oren or Balci, but they provide the modeler with a better sense of the segmentation of steps required in these expansive disciplines of verification and validation.

3.4.2.3 Assessment Processes with Numeric Ratings

Several frameworks have been presented thus far that show varying levels of detail in model validation and begin to introduce the communication aspect necessary for assessment processes. The following three frameworks introduce the idea of assigning a numeric rating to the model at the conclusion of the assessment. Gass (1993) presents four advantages when using a numeric scale during the model assessment process:

1. Decision-makers going through the process of determining weighting for the evaluated criteria require additional attention to those criteria and their value to ensure the right measures are being used
2. Sensitivity analyses can be done to understand the effect of criteria weightings to the final score
3. Similarly, sensitivity analyses can be done to understand the effect of the criteria on the final score
4. The rating objectifies the results that are presented in the model documentation to limit potential interpretation issues

However, Sargent (2001) disagrees with the value of a numeric rating system in improving communication indicating that it can lead to misinterpretation of the model:

1. *A model may receive a passing score and yet have a defect that needs to be corrected*
2. *The subjectiveness of this approach tends to be hidden and thus this approach appears to be objective*
3. *The passing scores must be decided in some (usually) subjective way*
4. *The score(s) may cause over confidence in a model or be used to argue that one model is better than another.* (p. 107)

It is most important to remember that "the accreditation score has no meaning by itself, it has to be combined with a written report, along with related sensitivity studies, so that the user can make a better judgment call as to whether to accredit the model" (Gass, 1993, p. 252).

Jain and McLean (2009) present the idea of a rating scale for attributes related to model confidence. They first present attributes related to increasing model confidence similar to those described earlier:

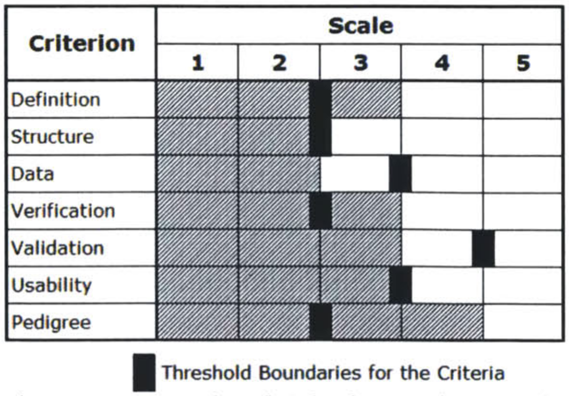
- *Software engineering practices/ software reliability*
- *Modeling practice*
- *Model confidence/ verification, validation, and accreditation*
- *Standards*
- *Interoperability*
- *User friendliness and accessibility*
- *Performance*
- *Innovation*

Within the model confidence / verification, validation, and accreditation attribute, Jain and McLean do not offer their own take on an assessment process, but rather offer other frameworks presented by other authors, some of which discussed here. In addition, however, they present the predictive capability maturity model (PCMM) (Oberkampf, Pilch, & Trucano, 2007). This model presents a zero to three scale to judge six elements as to their level of assessment completed, the objectiveness of the reviewers (internal or external) and level of documentation. The six elements are:

- Representation and Geometric Fidelity
- Physics and Material Model Fidelity
- Code Verification
- Solution Verification
- Model Validation
- Uncertainty Quantification and Sensitivity Analysis

This scale presents an easy communication vehicle to quickly portray the rigor used in the verification and validation process.

Gass and Joel (1981) introduce an assessment framework with a numeric rating scale and highlight the visualization that can be achieved. The framework itself is similar to that of Gass and Thompson presented earlier, but the resulting bar graph (Figure 20) shows an example of how a rating is applied to each criterion represented by the hash bar, which is easily compared to the minimum



Criterion	Scale				
	1	2	3	4	5
Definition	▨	▨	■▨		
Structure	▨	▨	■		
Data	▨	▨		■	
Verification	▨	▨	■▨		
Validation	▨	▨	▨		■
Usability	▨	▨	▨	■	
Pedigree	▨	▨	■▨	▨	

■ Threshold Boundaries for the Criteria

Figure 20: Example of Criterion Rating Matrix where the criteria are listed and a black box shows the minimum threshold required for each criterion for the problem being addressed. A bar is then drawn that compares the actual model rating to the acceptance threshold. [Adapted from (Gass & Joel, 1981)]

threshold required in order to confidently use the model results (black boxes). From this example, it can be quickly seen that although the model plotted in this example has a high quality pedigree, it lacks data and validation efforts. This allows modelers to know what to build upon to create a better model and tells the decision maker what to watch out for when using this model.

The final framework presented is a best practice with regard to the various issues noted above. It provides detailed guidance for building confidence against modeling criteria, documentation guidelines, and a rating scale to improve the objectivity of the model review.

Following the accident with the Columbia space shuttle in 2003, NASA requested the development of a Model and Simulation (M&S) standard [that] would:

1. Include a standard method to assess the credibility of the M&S presented to the decision maker when making critical decisions ... using results from M&S.
2. Assure that the credibility of M&S meets the project requirements.
3. Establish M&S requirements and recommendations that will form a strong foundation for disciplined (structure, management, control) development, validation and use of M&S within NASA and its contractor community.

(Bertch, et al., 2008; NASA, 2008; Thomas, Joiner, Lin, Lowry, & Pressburger, 2010)

There are two primary parts to the resulting standard: NASA-STD-7009. The first gives a series of requirements for the documentation process for the model or simulation. The second introduces a credibility assessment and scale with which to evaluate models and simulations.

There are forty-nine requirements for documentation which are summarized into eight sections (Bertch, et al., 2008; NASA, 2008)

The credibility assessment scale introduced in the second part of the standard (Figure 21) contains eight factors grouped into three categories:

a. M&S Development
    (1) Verification: Were the models implemented correctly, and what was the numerical error/uncertainty?
    (2) Validation: Did the M&S results compare favorably to the reference data, and how close is the reference to the real-world system?
b. M&S Operations
    (1) Input Pedigree: How confident are we of the current input data?
    (2) Results Uncertainty: What is the uncertainty in the current M&S results?
    (3) Results Robustness: How thoroughly are the sensitivities of the current M&S results known?

c. Supporting Evidence
   (1) Use History: Have the current M&S been used successfully before?
   (2) M&S Management: How well managed were the M&S processes?
   (3) People Qualifications: How qualified were the personnel?

(NASA, 2008)

Requirement Sections:	Description:
Programmatics	This section links the model to the program or project being addressed by it. It includes the objective of the M&S, risk assessments, acceptability criteria etc.
Models	The details of the models such as data, structure, assumptions etc. are maintained in this section.
Simulation and Analyses	This section begins with documenting whether the simulation was performed within the limitations of the model and continues with including data used for the simulation, errors obtained, processes used, etc.
Verification, Validation, and Uncertainty Quantification	This section includes documentation of verification and validation techniques used and resulting data. Of particular interest is the uncertainty quantification that shows the processes and results for evaluating uncertainty not only from the model, but from the real system being explored.
Identification and Use of Recommended Practices	In many cases, there are more detailed practices to model verification, validation, etc. that are specific to a certain type of model. Those practices are documented in this section to go above and beyond what the NASA standard is intended to cover.
Training	This section discusses necessary training for the M&S, but also the level of training required for the various roles related to the model.
Assessing the credibility of M&S Results	This section provides the detail of the credibility assessment scale covered in the second part of the standard.
Reporting Results to Decision Makers	This is more of an executive summary – pulling out the primary concerns, recommendations and results that decision makers need to focus on.

Table 2: Summary of eight sections of documentation requirements per the NASA standard NASA-STD-7009. (NASA, 2008)

Each of the eight factors is given a rating from zero to four depending on its level of credibility where zero is generally insufficient evidence and four represents the best case relative to that factor. The standard contains a detailed description of the rating for each of the factors and those details have been left out of this thesis. The rating is generally determined using a short paragraph description of the qualities that are associated with a given rating and factor. Each description varies by factor.

Five of the factors contain two sub-ratings. The first is the evidence based on ratings discussed above. The second factor, technical review, has the same criteria for each of the five categories containing it. It is also a zero to four scale based on the level of peer review conducted, whether it be internal or external with supporting documentation. The technical review is weighted up to 30% of the total score for its corresponding factor with the evidence review being the remaining 70% or more and the addition of the two then become the score for that factor.

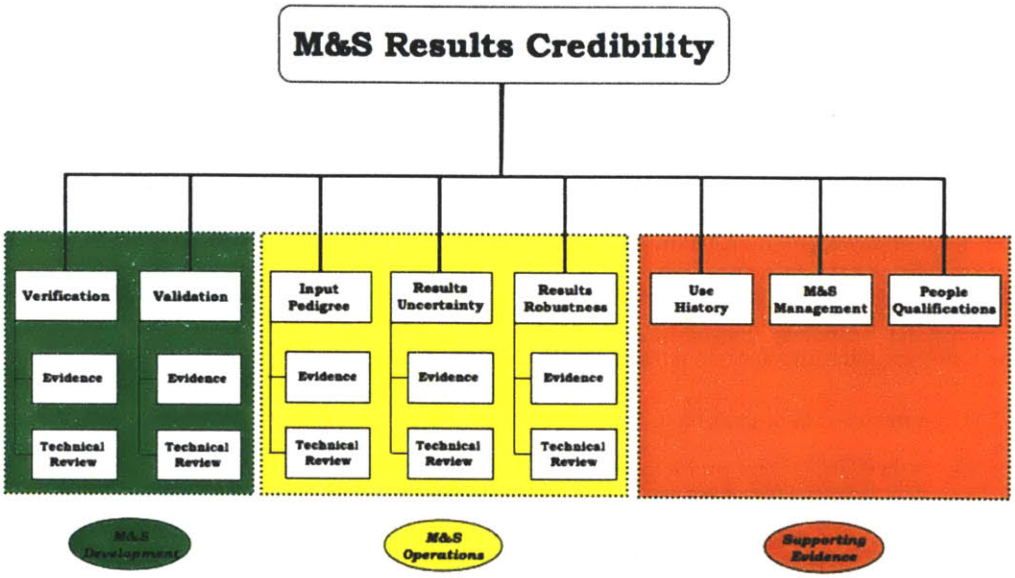


Figure 21: NASA-STD-7009 Credibility assessment scale showing the eight factors and how they are categorized into M&S development, operations, and supporting evidence (NASA, 2008)

The final total credibility score given to a model using this assessment is then the minimum value of the eight factors. This final score can then be compared against the predetermined desired level of credibility to determine if the model is acceptable to use for this application. Ahn and de Weck (2007) for example conducted an assessment of a space logistics simulation M&S called SpaceNet using this process and an interim version of the standard NASA-STD-(I)-7009 using a two round anonymous Delphi process and arrived at an aggregate mean credibility score of 2.4 and standard deviation of 0.6 in round 1 and a mean credibility score of 2.2 and standard deviation of 0.3 in round 2 on the 0 to 4 credibility scale. In the interim standard it was allowed to aggregate the category scores into an overall score.

This M&S standard has many advantages. First, it provides a complete checklist for requirements to be included in the model's documentation. This should help achieve more consistent documentation across models in the organization and is generally more complete. It also places focus on some critical areas that need improvement. Finally, the final rating representation of the model is the minimum rating as opposed to a sum or an average. This places emphasis on the weaknesses in the model's credibility and does not allow them to be averaged out by the model's strengths.

3.4.3 CONCLUSIONS OF MODELING ASSESSMENT

The assessment frameworks presented in this section barely scratches the surface of what is available in the literature on this topic (Balci, 2004). However, it demonstrates some of the breadth in available frameworks such as (1) the scope of the assessment being intrinsic to the model or including factors in the modeling environment (2) assessments done with an emphasis on improving the communication between modeler and decision maker and finally (3) both qualitative and numeric scales.

Besides the quantity and diversity of available frameworks, another observation revolves around the target audience for these frameworks. They are focused on the modeler and developers and how this demographic can work to improve communication and build confidence in the users and decision makers. However, "model confidence [is] not ... an attribute of a model, but of the model user" (Gass & Joel, 1981)

The following questions are raised from the observations above:

1. Given the breadth of available frameworks available to help modelers build credibility in their models, why is there still a large problem in the field with model misuse as introduced in the case studies?
2. The frameworks focus heavily on providing guidance to the modeler in building a better model, but little attention is given to the decision maker's perspective of the model or related processes.

3.5 BUILDING CONFIDENCE IN MODELS

Forrester (1961) differentiates between good and bad managers by their ability to selectively choose information and by how they convert that information into action. "The manager sets the stage for his accomplishments by his choice of which information sources to take seriously and which to ignore" (p. 93). The last section presented frameworks for how modelers can help distinguish their model as one to be taken seriously. However, Brennan and Elam (1986) indicate that "detail itself makes the decision maker's task of pinpointing important model relationships difficult" (p. 50), and also the details may not always be relevant to building confidence.

Confidence is an attribute of the decision maker, and not an attribute of the model (Gass, 1993); therefore, although the frameworks presented earlier do a lot to show whether a model has or does not have sufficient quality for the task at hand, they do not address building confidence on the part of the decision maker. Referring back to the decision support system, the interface with the user has not been addressed. To better understand this problem of perception, it is necessary to define what it means:

> *"Perceived product quality is an idiosyncratic value judgment with respect to the fitness for consumption which is based upon the conscious and/or unconscious processing of quality cues in relation to relevant quality attributes within the context of significant personal and situational variables."*
>
> (Steenkamp, 1990, p. 317)

According to this definition, much of the focus of the assessment frameworks is developing the quality cues and attributes of the product, or model in this case. The cues include the correctness of the model code relative to the conceptual model and the accuracy with which the model represents the real system. The quality attributes include the model's pedigree and experience in solving similar problems or its robustness to changing conditions. Each of these was represented in some fashion by the assessment frameworks presented earlier.

Steenkamp's definition of quality perception introduces the context within which these model-centric cues and attributes are interpreted and aggregated by the decision maker. There are personal or situational factors that can impact the decision maker's perception of the model itself and of the output produced by the model, even if done so unknowingly. This context was not addressed by the earlier frameworks and may help to explain situations where model perception does not match its actual quality (quadrants I and IV). There are a number of examples of context variables that can be provided from the case studies reviewed earlier. For example, in the industrial case regarding the EGR measurement venturi implementation on a diesel engine, the consequences of the situation around the design problem affected how the decision makers viewed the problem. In other words the risk tolerance of decision makers was such that they were not comfortable accepting the model's predictions – despite the model's solid pedigree – without also obtaining experimental data that later turned out to be essentially redundant information. Similarly, immense financial pressure at the $ billion level was created during the airspace closure following Eyjafjallajökull's eruption and this brought uncertainty to the atmospheric model (that had been extensively validated) lowering its perceived confidence level while raising confidence in an engine resilience model that was less mature or at least less accessible but whose results would allow the airspace to approach normal operations again. In the case of the Columbia tragedy, there were also a number of contextual factors affecting perception. Besides consequences related to the space shuttle program, there were documented organizational issues in the hierarchy and a lack of effective safety culture.

Steenkamp further describes the process around quality perception (Figure 22) that shows how these factors interact for a decision maker. There are three stages of the process; first acquiring data about the problem, followed by gathering information on attributes such as use history as well as critical aspects surrounding the problem such as its criticality or deadlines, and finally integrating all the information to generate a perception of the product.

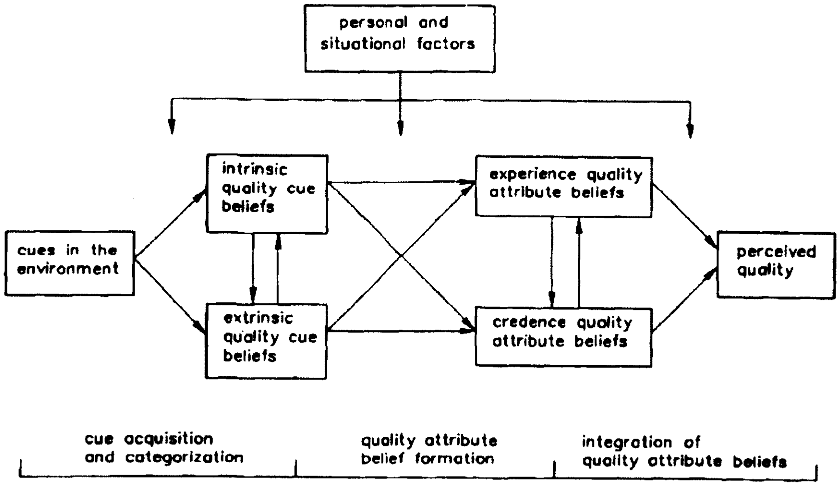

personal and situational factors
cues in the environment
intrinsic quality cue beliefs
extrinsic quality cue beliefs
experience quality attribute beliefs
credence quality attribute beliefs
perceived quality
cue acquisition and categorization
quality attribute belief formation
integration of quality attribute beliefs

Figure 22: Conceptual model of the quality perception process. This process is divided into three stages that begin with acquiring information about the product, applying belief attributes before integrating that with external factors before resulting in a perception of quality (Steenkamp, 1990).

3.6 ROOT CAUSE DETERMINATION

There are two open questions after reviewing modeling assessment frameworks. The first relates to the underlying motivation of the thesis and how, after so much work in the vast field of model validation and assessment processes, there remain numerous examples of model misuse. The second question revolves around the focus of model validation lacking from the perspective of the decision maker.

In the previous section, a framework was presented that focuses entirely on the point of view of the decision maker as opposed to the point of view of the modeler(s) that was discussed earlier. It raised a new question: what are the contextual variables that can impact a decision maker by means of the perception process?

These questions highlight the root cause of the problems associated with validation and perception. With regard to validation, there is much guidance in the literature to help modelers demonstrate the validity of a model. However, what is needed is a means to help the decision makers understand the validation process and its results for particular models. It is crucial for them to know what the areas of validity and what the model limitations are as they input the results into their decision making process. In addition to that, however, there are contextual variables that need to be understood and managed appropriately. These have the potential of completely undermining the impact of a modeling activity may increase the risk of making a decision that is based less on data and model outputs of good models and more on the context and instinct alone.

4 FRAMEWORK

At this stage, the initial problems uncovered from the case study examples – that of perception and validation - have been further analyzed in the previous chapter to identify their root causes: 1) ensuring proper model assessment, 2) understanding the problem from the perspective of the decision maker and 3) awareness of potential contextual variables that may override the impact of model output on a decision-making process.

Organizations trying to integrate model-based design within their processes can address these three problem areas in order to improve the implementation; organizations can help ensure that models are of sufficient quality to use for decision making by applying aspects of the assessment frameworks presented earlier. For instance, the NASA standard for rating model credibility addresses many of the critical areas of a model's validation and helps to create documentation for communication across the organization and for long term traceability. This addresses the first of the root causes, but neglects the second two.

Using lessons from the case studies, the literature, and the author's experience, a series of factors are proposed that can potentially help to explain the behavior of decision makers, particularly in quadrants I and IV. Awareness of these factors during the decision making process can help to better inform the process and open the possibilities for new solutions that may be best alternatives given all the constraints of the problem be they design, performance, or contextual environmental variables:

- Effect of Consequences
- Effect of Schedule Pressure
- Availability of Mature Models
- Purpose of Analysis
- Uncertainty and its communication
- Lack of Requirements and their validation
- Source and transparency of the model
- Fragmented Knowledge in the organization

In this section, these eight factors will be introduced. With each factor, a discussion will be provided that relates how the factor potentially impacts the four-box model and thus impacts the potential for optimal model-based design in organizations. Recall from the four-box model, that the optimum operation point is quadrant II, having sufficient quality models available and a culture that will place confidence in those models by using their results to help make good decisions.

4.1 EFFECT OF CONSEQUENCES

The first factor to consider is the consequences of the decision to be made. Consequences often relate to financial results such as profits and losses, human health including life or death as well as impact on the natural environment. The greater these consequences, the more pressure it will place on the model that is being used to make or inform a decision. This can cause the model to come under a lot of scrutiny to make sure there is as little uncertainty as possible in the result before succumbing to the consequences. The pressure placed on the decision maker in the event of large financial risk or potential loss of life may cause the decision maker to consider other sources of information as opposed to relying solely on the results of the model. In contrast, potentially beneficial consequences that are supported by poor models may be given more credibility as they support a favorable outcome. As shown in Figure 23, the potential impact of this factor is mainly on perception. The hypothesis is that if the outcome predicted by the model is favorable, perception is improved thereby risking a poor decision if the model is actually bad (moving from quadrant III to quadrant I). On the other hand perception of model quality may be lowered if the predicted consequences are detrimental, thereby causing decision makers to move from quadrant II to IV.

Each of these scenarios, both under predicted beneficial and detrimental consequences, are apparent in the case studies reviewed earlier. The EGR measurement venturi design problem had significant consequences relative to program schedules. If the model was wrong, the product could not be released to production before the mandated regulation dates. This would make the product non-compliant and unable to be sold in the domestic market. Similarly, the perception of the quality of the atmospheric model of the ash cloud ejected from Eyjafjallajökull's eruption was lowered as financial pressures grew. In this same case study, the relatively immature model of an engine's resilience to ash supported the possibility of reopening much of European airspace sooner, thus adding to its acceptance or perception despite its non-existence or immaturity.

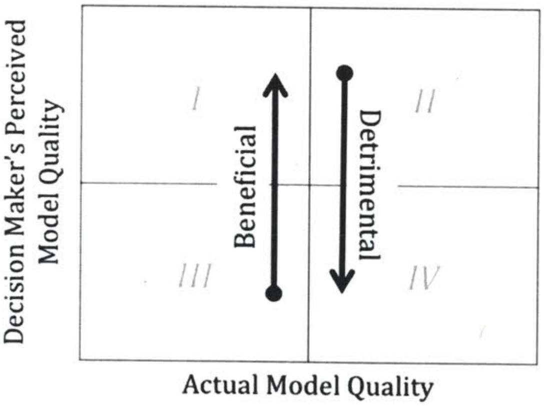


Figure 23: Effect of Consequences represented on four-box model. Consequences affect perception based on whether they are beneficial or detrimental. They risk pulling a bad model from quadrant III into I when consequences are beneficial, or a good model from quadrant II to IV if the consequences are detrimental.

As decision makers are posed with the problem of dealing with large consequences, they do not have the freedom to change those consequences and therefore are left only to confront them. This is especially true in cases where a decision cannot be simply delayed indefinitely. To do so requires first, recognition of the impact of consequences on the decision making process and second, to understand potential alternative actions that may optimize the design within the overall program constraints. Flexibility in engineering design is a emerging field providing a framework that "enable[s] us to avoid downside risks and exploit opportunities" (de Neufville & Scholtes, 2011).

A good example of this in practice is with regard to the final legislation following Eyjafjallajökull's eruption. The tiered approach to airspace closure from ash concentration has allowed for the necessary flexibility for airlines to operate with reduced cost and logistical impacts while still maintaining safety for the passengers. This relieved much of the pressure from the atmospheric model to where its use continued despite the scrutiny placed upon it. However, the conceptual model the tiered legislation is based upon requires proper validation to ensure the continued success of this flexible plan.

Similar strategies might have been taken in the case of the EGR measurement venturi where a number of potential alternatives could have saved testing efforts and schedule delays while providing added confidence that the model results were believable. These range from digging further into past validation work of the model to prove its competence, to running partial physical tests to prove the functionality of the model's results in order to move forward more quickly and with less impact to the program budget and schedule. Alternatively, a flexible design could be implemented immediately, based on guidance from the model. Doing so would impact the final prototype builds less in the case that the model is ineffective rather than having to start over.

4.2 EFFECT OF TIME PRESSURE

Time pressure is related to the effect of consequences, as schedule pressure is often a consequence of a decision itself, as was the case with the Columbia tragedy and the EGR measurement venturi. However, this factor speaks more to the time limit within which decision makers must act. Consequence and time pressure are thus somewhat decoupled variables. If schedule delays are a significant risk, that implies that a decision must be made in even less time to alleviate the schedule and to allow sufficient time for implementation of the action.

When decision makers are asked to take action in a limited amount of time, we hypothesize that they will value inputs to the decision process differently. The impact of this time pressure can be uncertain. It could result in implicit trust in the data from the decision support system, as there is no time to debate and no other sources of data available, or alternatively it could cause the decision maker to rely more on other factors such as the consequences, or past experiences, which may or may not be relevant.

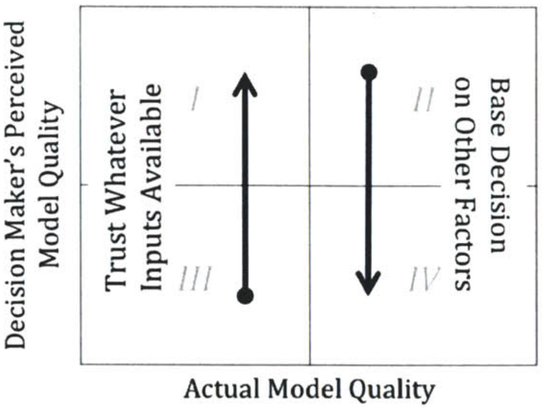

Decision Maker's Perceived Model Quality
Trust Whatever Inputs Available
Base Decision on Other Factors
I
II
III
IV
Actual Model Quality

Figure 24: Effect of Time Pressure represented on four-box model. Time pressure affects perception relative to model-based design based on how the decision-maker relies on the model in this situation. If decision-makers use whatever inputs available, then they are more likely to use a bad model (quadrant I). Conversely, if they choose to use other factors of the problem, the model may not be used effectively.

Depending on where the model belongs on the x-axis in Figure 24, the result to the four-box model is shown. Here, the perception will change depending on how the time available influences the value of inputs. Time pressure could reinforce the positive belief in an already good model, or it could raise the credibility of a poor model with the reverse situations being true. Options while under time pressure can be limited when it comes to providing more understanding of the details around the model to ensure its results are being perceived properly. With regard to the Volcano case, the question was raised: "is rational decision-making possible in a situation of emergency?" (Alemanno, 2011a, p. xxii).

When under time pressure the effect of the other factors may be amplified. In the case study examples, each case demonstrates this where the consequences carry even more weight in the decision-making process. For both the EGR and volcano cases, a risk-averse strategy was chosen in order to take action quickly. However, each of those cases, as discussed above with the consequences factors, could have potentially chosen more flexible solutions that would alleviate some of the risk in the decision.

In the case of the Columbia space shuttle, there was no time to certify a physics-based model that may have predicted more reliable estimates of damage from the foam projectile at launch; there was little time to choose any action before the space shuttle would run out of resources while orbiting. There is significant debate to this day whether it would have been possible to launch a rescue mission with a 2nd space shuttle waiting on the ground. Therefore time pressure during the mission became paramount.

4.3 AVAILABILITY OF MATURE MODELS

Another factor that has an impact on both the validation and perception processes is whether mature models are available to use for the decision-making process. This factor looks not only for the existence of a model to use, but whether that model has been used successfully in the past and is robust to varying conditions. The more of a track record and successful history of use a model has, the more likely it is to be believed. In general models begin as empirical models (e.g. using regression models, kriging, neural networks etc…) when the understanding phenomena are poorly understand and gradually become more sophisticated and credible as the underlying physics and causalities of the problem become better understood.

The impact to validation from this factor is related to having more robust models available that will, over time, gain credibility. With more use, more validation will be done which will move models to the upper right in the four-box model (Figure 25). The eventual impact of this factor is also on perception. Decision makers will become more accustomed to using a model-based decision support system and with experience will come trust.

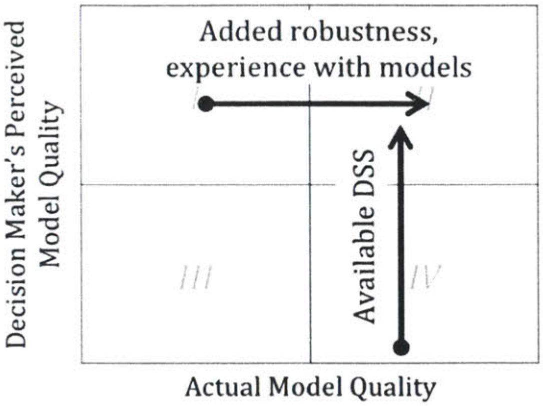

Added robustness, experience with models
I
II
Decision Maker's Perceived Model Quality
III
Available DSS
IV
Actual Model Quality

Figure 25: Availability of Mature Models represented on four-box model. Model availability will improve model quality over time drawing from quadrant I to II as more robust models become available. This practice will also change perceptions from generally risk-averse behaviors in quadrant IV to II.

The example from the Columbia space shuttle case shows the resulting validation problem from not having available mature models. The Crater algorithm had existed for some time, but was based on empirical data, not on a physics base, thereby it was not suitable to a problem that required extrapolation beyond its validation region thus resulting in poor model quality.

The perception effect can be seen from the EGR measurement venturi example where the organization was not yet accustomed to making decisions based on models, but was instead still heavily oriented towards physical testing as most problems in the organization did not have models available to use for decision-making. The prospect of using only a model to make a critical decision was therefore foreign to the decision makers as they lacked the experiential element, having not in the past made successful model-only based decisions.

Addressing this factor first takes investment of resources to develop and manage models within the organization. Impacting the perception from the point of view of the decision maker often takes time before the value of model-based decision making becomes universally accepted. Insisting on the use of models in as many problems as possible can accelerate this process. This will not only improve the validation of the model itself, but will begin to increase the visibility of a model's value with decision makers.

4.4 PURPOSE OF ANALYSIS

The purpose of analysis is a factor intended to capture if there is alignment between the decision makers and the model being used to solve a problem. For instance, it determines if the right questions are being asked to solve the problem at hand, and whether the model can answer those questions. This factor is related to both validation and perception.

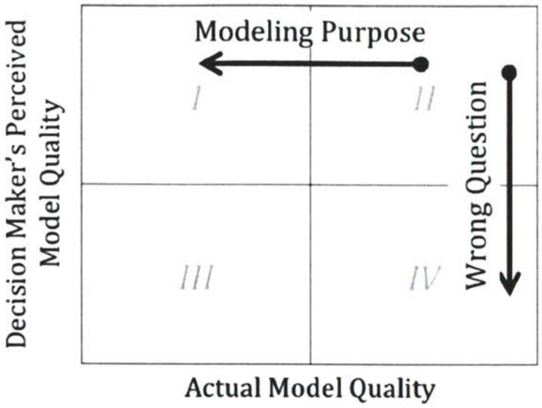


Figure 26: Purpose of Analysis represented in the four-box model. A different purpose of analysis may draw good models from quadrant II to I by using it for problems it was not intended to solve. This factor can also impact perception by taking a good model and asking the wrong question, thereby potentially moving it from quadrant II to IV.

It was established earlier that model validation is performed around a purpose of the model, in other words for a specific set of use cases or range of inputs. Therefore, if the model is not being used for its intended purpose that may essentially invalidate the model (Figure 26). By the same token, if the model is not addressing the right problem, the decision-makers will have less confidence in the model.

The Crater and Crater-like algorithms from the Columbia case best describe the validation problem related to the purpose of analysis. The algorithms were not designed to address the problem at hand. Not only were the models asked to predict RCC panel damage using parameter values well outside their validated region, they were not originally intended to answer the question of penetration depth, particularly in the case of the Crater-like algorithm for the RCC panels. Using the models for this purpose invalidated their results.

The perception effect can be seen from the volcano case related to the atmospheric model. In the days following the initial airspace closure, the validity of the atmospheric model was called into question when in reality key stakeholders were asking the wrong question. The problem was not whether the ash cloud was in one location or another; it was how to get the airspace flyable again with an acceptable level of risk. Once this greater problem was realized, it was clear that the atmospheric model was not the right (or only) source of information to find the ultimate solution.

The purpose of the analysis is a critical communication aspect between the modelers and decision makers in both directions. Without knowing the purpose of the model, decision-makers risk misusing the model as it was intended. Without knowing the purpose of the question, the modeler risks building or selecting the wrong model. Some of the assessment frameworks presented earlier seek to address this communication gap by providing complete documentation of the model's intended purpose and assumptions. By accepting these practices in the modeling process, the modeler has a means to make this more known to the decision maker. However, the problem does not begin with documentation. In fact it begins when the decision maker communicates the intent of the modeling activity to the modeler. A clear, solution-neutral problem statement needs to be provided to the modelers that allow for the best model to be selected or created. In some cases a model developed for a different purpose may indeed – after some extra validation – turn out to be adequate for the new purpose, in some cases the model may require adaptation and, finally, in other cases a model should not be used for the new purpose at all and – assuming sufficient time and resources are available – an entirely new model should be created.

4.5 UNCERTAINTY AND ITS COMMUNICATION

Uncertainty is potentially a very important factor influencing the confidence in a model. Often uncertainty has a bad connotation in that it implies little is known about the system and the model's ability to represent that system. However, knowing the uncertainty allows for a better understanding of the likelihood that the system falls within a range of performance levels predicted by the model. Often uncertainty is not generally determined, particularly not at a system level where the value is directly attributable to a requirement and includes the variability from all pieces of the system it is describing. When uncertainty information is not available or communicated, decision makers are required to use their own mental perception of the model and system to judge within what accuracy the predicted model output is. This concept will vary based on an individual's experiences, biases, and situational factors and therefore will not be consistent.

Uncertainty can impact both a model's validated quality as well as its perception (Figure 27). Reducing uncertainty of the model output against the real system behavior helps the case for improved model quality, but is not a necessity to be sufficient for a problem, for all models will have some level of uncertainty, the key is knowing what level is acceptable. Uncertainty impacts perception in different ways. If uncertainty from the model relative to the real system is known by the decision maker and is at acceptable levels for the problem, perception will improve, as the decision maker will understand the likelihood of making the right decision.

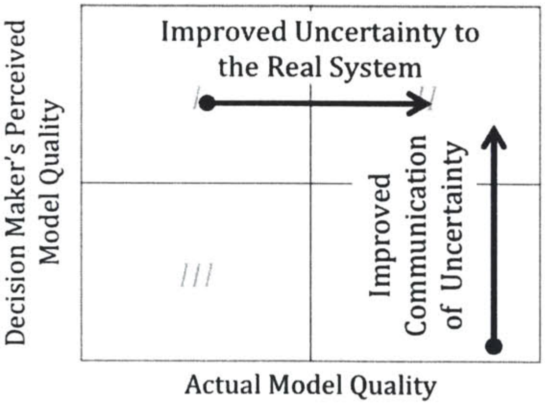

Improved Uncertainty to the Real System
I
II
III
Improved Communication of Uncertainty
Decision Maker's Perceived Model Quality
Actual Model Quality

Figure 27: Uncertainty and its communication represented in the four-box model. Uncertainty has the potential to move models to the optimum for model-based design: quadrant II. It does so by improving perception of good models by indicating its capabilities and limitations. By recognizing uncertainty, it can also drive model improvement to reduce uncertainty relative to the real system thereby allowing bad models to move from quadrant I to II.

Besides determining the level of uncertainty, its communication is perhaps even more critical. As discussed, a decision maker not knowing at all about the uncertainty inherent in a model may let other perception factors such as past experiences or the state of the situation dictate what they sense as being the right decision. Just as important as this, however, is that during the initial problem formulation that the decision maker communicates the level of acceptable uncertainty for a problem to be solved or model output to be produced. For example, early in a development program, a simple answer of the behavioral trend (increasing or decreasing?) may be sufficient without care for the absolute value until later in the program. A common frustration in organizations implementing model-based design is that decision makers generate a problem statement and modelers take too long building a model that has much higher fidelity than is needed for that stage of decision making. As a consequence, decision makers are often required to make a decision without input from a model, since the model is still under development. Once the model is ready, its output is no longer needed since the decision has already been made.

From the case studies there were several discussions about uncertainty that impacted their outcomes. In the Columbia tragedy, there was a very high level of uncertainty associated with the results produced by the Crater and Crater-like algorithms because they were being used so far beyond their validated region. However these uncertainties were not communicated well to the decision makers who were making engineering judgments based on these models. Had these uncertainties been better communicated, perhaps the decision makers would have sought other sources of information with which to reduce the uncertainty, such as the in-orbit pictures of the shuttle that had been requested on numerous occasions.

The atmospheric model used to forecast the ash cloud propagation from Eyjafjallajökull's eruption was also impacted by uncertainty. There were many known uncertain inputs to the model from data based on visual observations and meteorological data. Although these uncertain inputs were recognized, the resulting impact on model output uncertainty was not communicated and perhaps this provoked increased model scrutiny as test flights were being conducted.

The key point related to this factor is that uncertainty is not bad per se, but no knowledge of model uncertainty or lack of communication of it can lead to model misuse. For model-based design to be effective in organizations, a clear understanding of allowable uncertainty for a problem needs to be established early and communicated to the modelers. In return, modelers need to provide model uncertainty information that is aligns well with the requirements of the real system.

4.6 LACK OF REQUIREMENTS AND THEIR VALIDATION

This factor refers to the level of upfront work to a system and model to generate requirements and validate those requirements against the real system. The stated requirements govern the desired performance and behavior of the system and therefore it is not possible to judge the quality of a model if the details of its behavior are unknown or ill defined. Therefore, this factor impacts the model validation and resulting quality (Figure 28).

In the case of the EGR measurement venturi example, lacking requirements early in the program prohibited early system-level modeling and limited the effectiveness of the engine cycle simulation. Although high-level requirements for the product were available as mandates from the various stakeholders such as emissions limits from the EPA, engine performance expectations from the customers, and sizing constraints from the vehicle applications, the applicability of these requirements to subsystems was not determined. For example, knowing the allowable error in EGR measurement as a function of required NOx emissions may have helped to understand earlier what an optimum design solution might be.

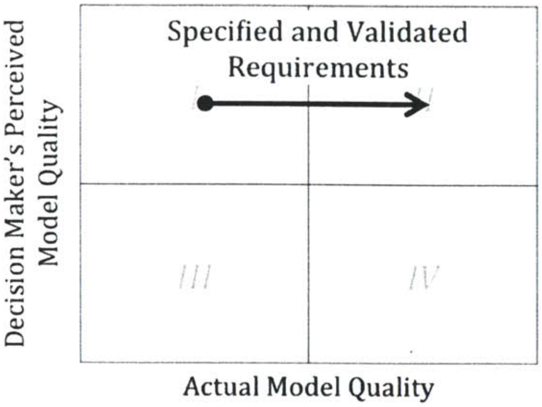

Specified and Validated Requirements
I
II
III
IV
Decision Maker's Perceived Model Quality
Actual Model Quality

Figure 28: Requirements and their validation represented in the four-box model. Improving requirement specification and validation yields the opportunity to improve model quality as it is better defined.

A similar situation can be seen from the Columbia accident. Early space shuttle program requirements had indicated that the RCC panels would not be subject to impact. As a consequence RCC panels were not designed specifically to withstand foam impacts during ascent. After numerous impacts during launch did occur in reality, it became clear that this was a missing requirement. Had this requirement been properly stated and subsequently validated, it may have helped to develop more mature models to assess foam impact earlier.

There is much to read in the literature about requirements of engineering processes. As part of systems engineering implementation, a crucial step is setting up full requirements early and decomposing them to subsystem and component levels. In addition, they need to be properly managed during the course of the program as changes to requirements often occur. Besides generating good requirements, model-based design applied to systems engineering can help to validate those requirements early while continually developing models for use later in the program.

4.7 SOURCE AND TRANSPARENCY OF THE MODEL

The source of a model refers to who created and programmed it or who used it to generate results. Transparency is similar, but refers to how accessible the concepts, assumptions and governing equations of the model are to the decision makers. The more trust-worthy the source and the more transparent the model, the more confidence a decision maker will likely have in it (Figure 29). According to this logic a "black box" model from an unknown source would result in a low level of confidence. This is not to say that proprietary models or those from less trust-worthy or unknown sources are wrong, just that they are not likely to be perceived with as high a level of regard as transparent models from reputable sources.

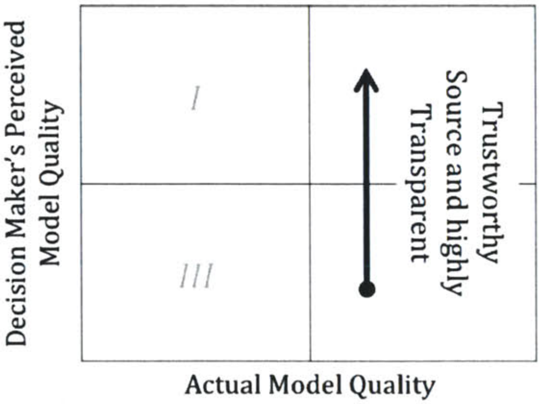

Decision Maker's Perceived Model Quality
I
III
Trustworthy Source and highly Transparent
Actual Model Quality

Figure 29: Model Source and Transparency represented in the four-box model. A trustworthy source combined with a model that is transparent will improve how people perceive the model allowing a good model to potentially move from quadrant IV to II.

From the volcano case, the engine model is a good example of lack of transparency. As the model does not exist in the public domain and may be held confidentially by engine manufacturers, it is difficult to validate its credibility particularly as it is used for public legislation related to actions during volcanic eruptions that affect the flying public. The effect of the legal system and availability of "discoverable" information during lawsuits may have a strong impact on model transparency and willingness to share model details.

The EGR measurement venturi represents an example of a trusted source gone awry. In this case, a conceptual model of the measurement venturi was accepted with few limitations because it was what competitive engines in the on-highway diesel engine market had already adopted. However, as the designers were in competition with each other, transparency into the details of their implementation was not available to understand potential difficulties.

It is difficult to overcome personal biases related to a model's source and transparency. However, as model-based design initiatives and benchmarking competitions in industry increase, there are more modeling packages and third-party modeling consultants emerging to help address the level of work and expertise required. Therefore, this has the potential to become a growing issue in industry. It is important for decision makers to remain as objective as possible in the face of these concerns and to develop standard validation plans that will help to ensure that the models, regardless of their origins, are still representative of the system.

4.8 FRAGMENTED KNOWLEDGE IN THE ORGANIZATION

It is not possible for every person in an organization to know, in detail, all aspects of every technical problem. What is critical is making sure the necessary information required to make a decision is properly communicated and aggregated; lack of this can lead to fragmented knowledge and an information gap between the decision makers and the modelers. This can happen because of the structure of the organization, obstacles to communication, or differing levels of technical understanding of the system and model. The impact of this factor is similar to the purpose of analysis where if the decision maker does not have full knowledge of the problem and surrounding situation, the perception will be affected more so by other factors. Similarly, if the modeler does not have sufficient knowledge of the problem, the model may not meet needed requirements.

In the case of airspace closure from the Eyjafjallajökull volcanic eruption, much has been documented on the segmentation of knowledge and responsibilities and how that played a role in the impact the volcano had. Macrae presented the problem in a succinct manner (Figure 31). Here, the capability of making a decision is plotted against the knowledge of the problem. Many of the modelers and scientists had the most knowledge about the problem such as details from the atmospheric model and information about the engine's resilience to ash at the engine manufacturers; but these groups did not make the decisions. Moving up the decision capability or decision authority axis, knowledge is lost about the details of the problem. It is imperative that the knowledge the people at the top of the decision chain have is what is needed for the problem. This is also an acute challenge in the intelligence community.

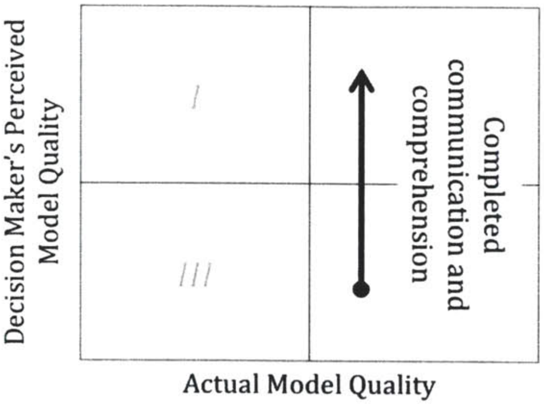

Decision Maker's Perceived Model Quality
I
III
Completed communication and comprehension
Actual Model Quality

Figure 30: Fragmented Knowledge represented in the four-box model. As decision-makers have a better understanding of the problem and the model being used, a good model is more likely to be utilized.

Model assessment frameworks presented earlier show different documentation techniques for the model intended to help decision makers become more knowledgeable about the model. However, the contextual variables suggested in the previous factors are critical to share as well. This forces both modeler and decision maker to be cognizant of the factors, some of which they cannot control, and can then work to mitigate their effects in order to get the optimal use out of the model and to ultimately make the best decision.

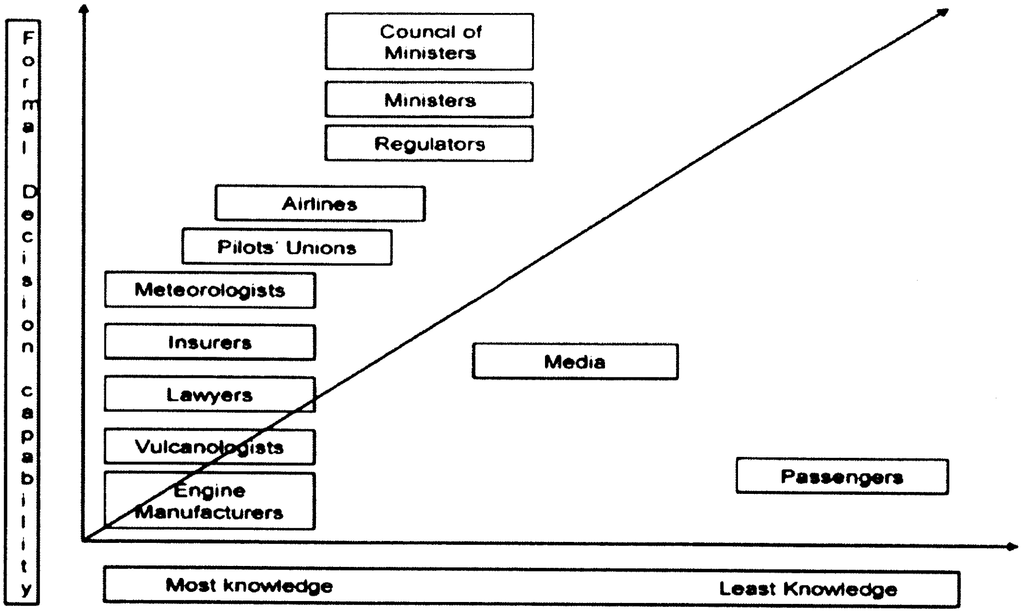


Figure 31: Decision-Making and Knowledge Map of Actors in the Volcano Case. The x-axis shows the degree of knowledge about the case where the y-axis illustrates the capability or authority in the decision process of the various actors. Engine manufacturers had the most knowledge of the situation, but no capability in making decisions on airspace. The decision-makers at the top of the chart had much less knowledge about the models on which to base a decision (Macrae, 2011).

5 TESTING THE FACTORS

The framework presented in chapter 4 included eight factors that help to describe how contextual factors can impact decisions within an organization trying to implement effective model-based design principles and decision-making. The factors were formulated based on research, case study analysis, and industry experience. By drawing from these sources, there are risks that additional complexities in the cases may have skewed the conclusions. Therefore, it is necessary to test the hypotheses discussed in the last section by means of an experiment where the complexities can be better controlled. This section will present such an experiment that used a website with a simple model of a catapult that had been developed by Dr. Troy Savoie and Dr. Daniel Frey (Savoie & Frey, 2012). People were asked to interact with the model in order to make a design decision on the catapult. They were asked to provide direction for the design decision, and then complete a survey based on their experiences.

As part of the experiment, there were different problem details and model descriptions that intended to alter the contextual factors from the framework and measure how people responded. This data could then be used to determine the impact of each factor discussed in sections 4.1 through 4.8 and also help inform methods that may optimize the impact of these factors relative to the four-box model thereby driving behavior of organizations preferentially towards quadrants II and III.

5.1 HYPOTHESIS

The overall hypothesis of the experiment is that the factors presented in the framework (4.1 through 4.8) do indeed impact decision-making processes in model-based design. It was also of interest to test each factor against the four-box model used in the previous section to describe its impact because doing so would help to identify methods for managing these factors. Therefore, the detailed hypotheses are shown by factor in Table 3.

Factor	Hypothesis
Effect of Consequences	Prediction of beneficial consequences will improve model perception while detrimental consequences will reduce perception of model quality
Effect of Time Pressure	Applying time pressure will cause people to use whatever inputs are available in the process whether validated or not
Purpose of Analysis	By asking questions of the model for which it was not originally created, the quality perception of that model will be reduced
Uncertainty and its communication	Improving information about uncertainty will lead people to better recognize a good or bad model
Source and transparency of the model	A more trustworthy model author and transparent governing equations will improve model perception
Fragmented Knowledge	Better knowledge of the entire problem will cause the perception of the model to increase
Setting Requirements and Model Availability	Could not be easily tested in this experiment

Table 3: Summary of experimental hypotheses by factor

5.2 METHODOLOGY

5.2.1 DESIGN PROBLEM

The experiment revolved around a single binary decision related to a simple design problem. The following text was given to the user describing the objective of the problem:

"Your company is releasing a catapult to the market that will be used for educational purposes... You are responsible for approving the design before releasing the product. Your team was on track to release the product until there were some last-minute changes in requirements. You now have to decide whether to approve the design for release, or delay the product launch to either redesign the catapult or wait until prototypes have arrived to test the new catapult design based on the new specifications."

The changes in requirements involved 1) Changing the ball type, 2) reducing the number of rubber bands from two to one and 3) addition of a constraint where the launch and pullback angles together could not be more than 90 degrees (see Figure 32). The performance objective of the catapult was unchanged. The catapult had to be able to get a ball into a cup located at a given distance from the base of the catapult (see Figure 33).

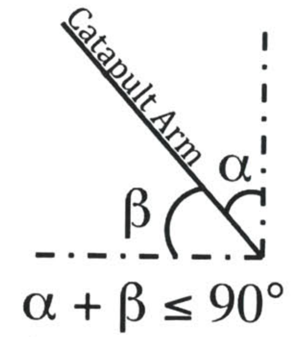


Figure 32: Launch plus Pullback Angle Design Constraint

After interacting with the model for a limited time, the user had to choose between two options:

- Proceed with Product Launch: use the current design proposal to meet requirements
- Delay Product Launch: wait for additional testing with the proposed design with potential for redesign needed

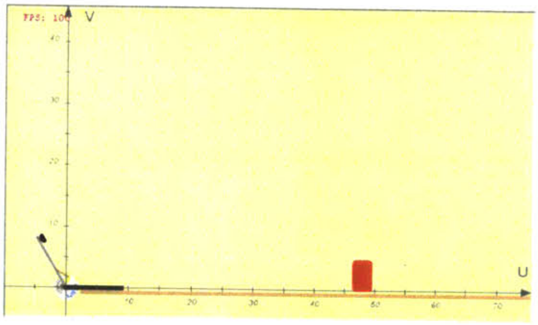


Figure 33: Performance Objective for the Catapult

The user was provided an online model of the system to use to help understand if the new design was going to work and meet its requirements. Some of the test subjects were also given test data from the real system to validate if the model was truly representative. The users were told they would have limited time to check the model against the validation data, test the design scenario, and make a decision.

5.2.2 DESIGN OF EXPERIMENTS AND FACTORS

The experiment conducted was based on a one factor at a time (OFAT) experiment design. In this method, a reference case is run, and each factor is changed one at a time while the remaining factors remain at a reference level (Savoie & Frey, 2012). This design was chosen for the reduced number of experiments required to test the factors, as there was concern about getting a large enough sample size for each case. The OFAT method loses the ability to quantify any potential interactions in the factors, but this was not a primary focus of the experiment.

Not all factors presented in the framework in chapter 4 could be experimented on in the context of a website, therefore the following factors were chosen to vary in this experiment:

- Effect of Consequences
- Effect of Time Pressure
- Purpose of Analysis
- Uncertainty
- Model Source

In order to validate the 4-box model (Figure 1), it was necessary to also run models that were both of good and poor quality. This resulted in the following test matrix (Table 4) with 12 experiments with each of the five factors tested with good and bad model quality.

The reference and test cases will be discussed in the following sections specific to each factor, however, the reference was intended to be run with more realistic settings of the factors, not necessarily the best or worse case for model perception. A full factorial experiment would have required 2^6 = 64 test cases and with approximately 20 test subjects per test case to obtain reliable statistics a total of about 1,280 test subjects would have been required. With the OFAT design the number of required test subjects was closer to 240, which was more realistic given the time constraints of this thesis.

After the respondents each completed the experiment by responding with a design decision (go for production now or wait for additional information), they were directed to a survey to collect additional data about their thoughts and opinions of the model. The full survey instrument is provided in the Appendix, and the questions were designed to elicit the respondents' confidence in the model in cases where that confidence may have differed from their choice to proceed or not as that decision may have been driven by factors other than model confidence. In addition, questions were asked in the survey to determine if the test factors were important to the users in their considerations, and therefore would be effective in their respective test cases. For example, in the case of model source, the respondents were asked if they considered the model's source as important when making their decision. For other factors, like uncertainty, the respondents were asked to rate the capability of the model and catapult with respect to launch distance repeatability. In this case, those presented with uncertainty bands from the test case knew the answer to this question, whereas those without uncertainty bands were left to guess.

	Model Quality	Consequences	Time Pressure	Purpose of Analysis	Uncertainty	Model Source
1	Good	Reference	Reference	Reference	Reference	Reference
2	Bad	Reference	Reference	Reference	Reference	Reference
3	Good	TEST	Reference	Reference	Reference	Reference
4	Bad	TEST	Reference	Reference	Reference	Reference
5	Good	Reference	TEST	Reference	Reference	Reference
6	Bad	Reference	TEST	Reference	Reference	Reference
7	Good	Reference	Reference	TEST	Reference	Reference
8	Bad	Reference	Reference	TEST	Reference	Reference
9	Good	Reference	Reference	Reference	TEST	Reference
10	Bad	Reference	Reference	Reference	TEST	Reference
11	Good	Reference	Reference	Reference	Reference	TEST
12	Bad	Reference	Reference	Reference	Reference	TEST

Table 4: OFAT Design of Experiments. The definitions for "Reference" and "Test" levels for each factor will be discussed in sections 5.2.3 through 5.2.8.

5.2.3 MODEL AND VALIDATION DATA

The model used for this experiment was a simple catapult model based on a physical catapult, XPULT (Figure 34), used for teaching design of experiments, six sigma, etc. (Peloton Systems LLC, 2010). The model was a java-based model developed by Dr. Troy Savoie for use in his own experimenting (2010). The visual interface was very realistic and intuitive, showing the catapult, it launching a ball and a depiction of the ball's trajectory and impact point. The model allowed the user to change the same factors as in the real catapult (Figure 35):

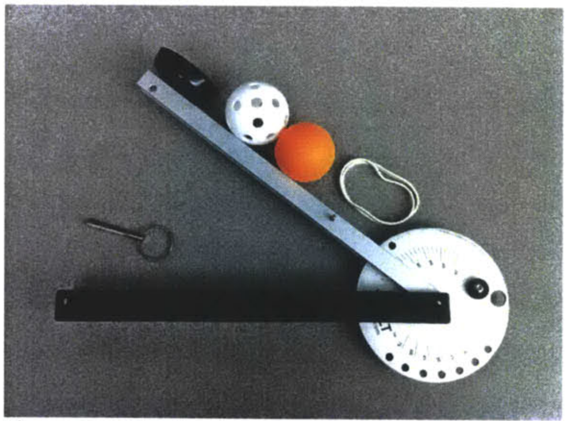
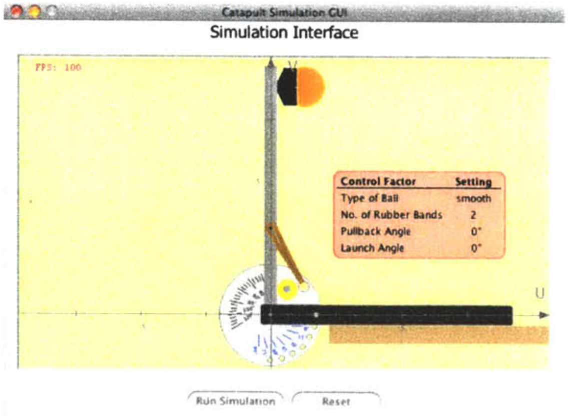


Figure 34: XPULT catapult from Peloton Systems

- Ball (smooth or perforated)
- Number of Rubber Bands (1 or 2)
- Launch Angle (0, 15, 30, 45, 60, 75, 90 degrees)
- Pullback Angle (0 – 120 degrees)

Besides the ability to use the model, some users received a set of validation data. This was data taken using the physical catapult within a given configuration. Generally, the validation data included runs with a single ball and rubber band setup and a series of launch angle and pullback angle combinations were run. The data was presented in both graphical form and tabular form.



Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	0°

Figure 35: Java-based model of the catapult (Savoie, 2008)

It was necessary to have a good and bad model for this experiment. Due to complications with obtaining the source code for the model, it was not feasible to edit the model in order to break it and make it intentionally bad. Therefore, the model was virtually "broken" using the validation data. In the physical experimentation, it was found that the smooth ball data correlated well with the model, while the perforated (whiffle) ball did not, particularly at smaller launch angles. Therefore, in the case of the good model (odd experiment numbers in Table 4: rows 1,3,5,7,9 and 11), the data from the smooth ball was provided and the test subjects were asked to evaluate the model using the smooth ball, while in the bad model experiments (even experiment numbers in Table 4: rows 2,4,6,8,10 and 12) test subjects received validation data from the perforated ball. The difference between the two cases, besides the ball, was primarily in the performance of the model at the critical launch and pullback angle points relative to the new performance requirement (constraint) – or the combination of launch and pullback angles that were equal to 90 degrees. The good model matched its validation data within the uncertainty bounds (about plus or minus 10 inches) at the launch and pullback angle combinations of interest (primarily 30 degree launch with 60 degree pullback angles and 45 degree launch with 45 degree pullback angles). The bad model, however, did not match the validation data at these same pullback angle / launch angle combination points within the same uncertainty bounds. The data was modified some from the physical experimentation so that the "good" model data would be better and the "bad" model data would be worse. The final charts comparing the data between the model and the validation data provided are given in the Appendix (Figure 39 and Figure 40).

5.2.4 CONSEQUENCES FACTOR

In order to represent the effect of serious consequences on the perception of a model, a scenario was devised that would place economic pressure on the user. The test subjects were told that there was a potential reward for early sales contracts if the design was released on time, which would require proceeding with the design now. The value of these contracts was $1,000,000. However, there was a risk of $3,000,000 in warranty claims if the design was sent to market early and was unable to meet its performance objective once deployed. If the user chose to delay the design for further testing, there would be a cost of $500,000 in testing and redesign efforts. The following table summarized the consequences factor (Table 5):

	Catapult Design is GOOD	Catapult Design is BAD
Launch Product	$1,000,000 potential for securing first-to-market contracts	$3,000,000 lost in warranty costs
Delay Launch	Lost opportunity for $1,000,000 in early contracts Additional $500,000 in testing efforts to verify design	Saved from $3,000,000 in warranty costs Additional $500,000 in redesign and testing efforts

Table 5: Consequences of Design Problem Decision

The reference case for the Consequences factor was chosen to be representative of reality in which case the scenario presented in Table 5 was provided to the user. In the test case, the users were told this was purely an academic exercise. The implication of the test case was that there was no real consequence whether the test subject made a good or bad decision.

5.2.5 TIME PRESSURE FACTOR

The time pressure factor was implemented by changing the time users were given to spend with the validation data and when they were supposed to make a decision. When the user went to "run the model" from the website, the next page included the validation data, a link to download the data, and a summary of the design problem they were trying to solve. At the top of the page was a timer to remind them how much time they had left. After the timer expired, they were automatically sent to another page without option to return and they were forced to make a decision right there and then.

The reference case for the time pressure was not chosen based on what is most common in reality, which would generally be a short amount of time to make a decision. During pilot testing of the experiment, it was found that the short time created such significant pressure on the user they were not paying attention to any other factors in the experiment. Continuing with a short amount of time as a reference may have resulted in inconclusive results for the remaining factors. Therefore, the reference case was made to be 15 minutes (long time = little or no time pressure) and the test case was 4 minutes (short time = lots of time pressure).

5.2.6 PURPOSE OF ANALYSIS

For the purpose of analysis factor, the performance objective of the catapult was changed. The model is very well suited to provide a prediction of the horizontal distance traveled by the ball from the catapult. The model visually shows the launch and ballistic trajectory of the ball, therefore, it is possible to judge a maximum vertical height reached by the ball, although this value is not an explicit numerical output from the model and therefore it requires careful visual split-second inspection while the model is running to determine it.

The performance objective of the catapult was to get a ball into a cup (Figure 33). Test subjects in the reference case were told the cup was located at 48 inches from the base of the catapult. If they could confirm that the catapult could launch the ball greater than this distance, then it could be configured by the customer to land in the cup.

The test case used the height as the objective. In this case, users were told the cup was located at a distance from the base of the catapult. The engineering team had confirmed that the ball could travel far enough under the new requirements to reach the cup, however, the ball needed to also go high enough in order to fall into the cup as opposed to hitting the side or rim. Therefore, the ball needed to reach a minimum height of 14 inches during its travel. The validation data did not change for the case and was still focused on horizontal distance.

5.2.7 UNCERTAINTY

The uncertainty in this case was communicated via the validation data. Error bars were either presented or not presented depending on the experiment. The charts shown in Figure 36 show examples of what the test subjects saw. On the left of Figure 36, the uncertainty information is not available and it was up to the user to decide how accurate the validation data was. On the right, the error bars were provided showing approximately plus or minus two standard deviations from the actual measurements.

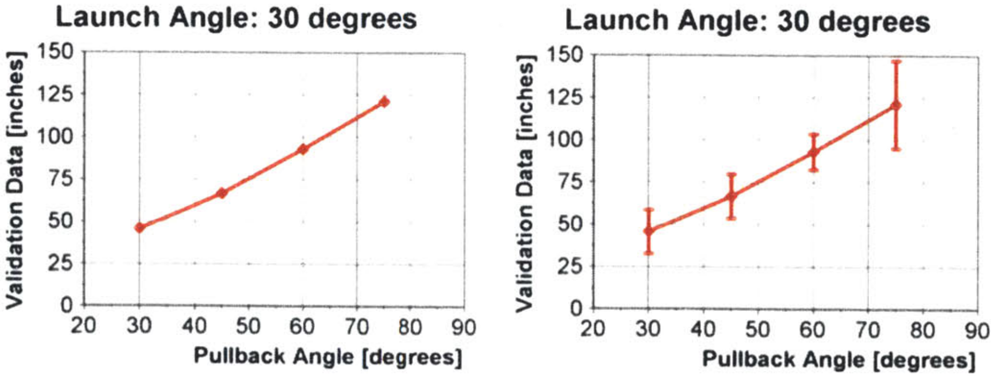


Figure 36: Example Validation without Uncertainty (left) and with Uncertainty (right)

The reference case for uncertainty was to not communicate uncertainty information (no error bars), as this is typical of reality. The test case presented the error bars in the validation data.

5.2.8 MODEL SOURCE

The model source used in this experiment was meant to provide more or less credibility to the model based on who developed it and what sort of code was used. This was provided in the description of the design problem as well as a reminder when the user was inputting their decision. The reference case was to be realistic and truthful about the source of the model:

"The model was developed by a PhD student in the Mechanical Engineering department at MIT. This model uses principles of physics to estimate the landing position of the ball. The PhD student has modified the model to meet the new design specifications."

Users with the test case saw the following:

"The base model was found online and was calibrated based on data from the original prototypes. This model is proprietary and therefore it is unknown how the model is calculating distance. The editable parameters in the model have been modified to meet the new design specifications."

5.2.9 REMAINING FACTORS

The remaining factors: Lack of requirements and their validation, availability of models, and fragmented knowledge were not directly included in this experiment primarily due to the difficulty in testing them in this environment. However, questions were included in the survey to help collect potential datasets that could determine patterns based purely on the a priori experiences of the respondents. The survey questions asked at the end of the experiment are shown in Appendix C section 8.3.1.4.

5.3 IMPLEMENTING THE EXPERIMENT

The website for this experiment was developed by Spinutech (2012) with direction and content provided by the author. The website included each of the 12 experiments described in Table 4. As a user went to the website, they were assigned an experiment ID number and saw the details related to that experiment. Subsequent users received the next experiment in the table and so on. So if a test subject started the experiment and was assigned to row 5, then the next test subject would automatically be assigned to row 6. After row 12 the table index reset to row 1. There was no need to randomize the runs as the order people went the website was essentially random. This was also done in an attempt to provide even sample sizes for each of the experiments. Test subjects were only allowed to do the experiment once so that learning effects should not be a factor.

The website link was distributed by a variety of means. The intent was to target "students and professionals in a technical field." This was a rather broad classification. Emails were sent to several hundred colleagues both in professional and academic environments. In the email, the recipients were encouraged to forward the email to their own network. In addition, posts were made to the LinkedIn website (LinkedIn Corporation, 2012) specifically posting to various technically focused group to which the author belongs.

The experiment was open for about four weeks. During that time, there were about 400 hits to the website. However, only results that were tagged as "completed" were used to analyze as these results included a final decision on the design problem. There were 252 responses that completed the experiment and this slightly exceeded the target of 240 that had been set for this OFAT experimental design. The responses were distributed across the experiment rows as shown in Table 6. By inspection, there was a fairly even distribution of responses by experiment number.

Experiment #	Number of Respondents	Model Quality	Consequences	Time Pressure	Purpose of Analysis	Uncertainty	Model Source
1	**23**	Good	Reference	Reference	Reference	Reference	Reference
2	**22**	Bad	Reference	Reference	Reference	Reference	Reference
3	**19**	Good	TEST	Reference	Reference	Reference	Reference
4	**23**	Bad	TEST	Reference	Reference	Reference	Reference
5	**21**	Good	Reference	TEST	Reference	Reference	Reference
6	**22**	Bad	Reference	TEST	Reference	Reference	Reference
7	**20**	Good	Reference	Reference	TEST	Reference	Reference
8	**21**	Bad	Reference	Reference	TEST	Reference	Reference
9	**20**	Good	Reference	Reference	Reference	TEST	Reference
10	**18**	Bad	Reference	Reference	Reference	TEST	Reference
11	**22**	Good	Reference	Reference	Reference	Reference	TEST
12	**21**	Bad	Reference	Reference	Reference	Reference	TEST

Table 6: Number of Respondents by Experiment Number

5.4 RESULTS

5.4.1 STATISTICAL TEST

The data obtained from the experiment required regression analysis in order to determine relationships between dependent and independent variables. However, the variables collected from this experiment were non-continuous and usually dichotomous therefore requiring logistic regression as the primary method for analysis. Logistic regression is similar to the more familiar linear regression but is designed specifically for use with discrete outcomes by employing the logistic distribution with the logit transformation (Hosmer & Lemeshow, 1989). Details of the logistic regression derivation will not be covered here. This section will provide an introduction to some of the types of results taken from the regression analysis used in the data results and interpretation to follow.

There were two important things to extract from the experimental data being analyzed: first, do relationships exist between input variables and outputs and second, are those relationships statistically significant. The resulting relationship from logistic regression following the logit transformation is of the form of linear regression where:

$$g(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

The resulting coefficients (β) are more useful for independent variables that are continuous in nature. Therefore, the relationship of interest for discrete independent variables is an odds ratio. If the independent variable in the model has options A or B, the odds ratio is defined as the odds of A over those of B for a given outcome (Hosmer & Lemeshow, 1989). The odds ratio (Ψ) is related to the regression coefficients through the following:

$$\Psi_1 = e^{\beta_1}$$

For the example where the odds ratio is the odds of A over B, the resulting value can be interpreted by the following with respect to a given outcome:

$$\begin{cases} \Psi = 1.0 & \text{A is as likely as B to produce the outcome} \\ \Psi < 1.0 & \text{A is less likely than B} \\ \Psi > 1.0 & \text{A is more likely than B} \end{cases}$$

Based on the above ratio, it is key to understand which level of the independent variable is A and which is B as provided here to interpret the ratio properly.

Knowing the odds ratio will determine the impact a variable has on an outcome, however, it is important to understand if the variables in the relationship are statistically significant to the model and therefore are reasonable predictors of outcomes. As in the case with other regression techniques, the model is compared against the data when it includes the variable in question, and when it does not (Hosmer & Lemeshow, 1989). A statistic describes this comparison for logistic regression as:

$$G = -2\ln\left[\frac{\text{(likelihood without the variable)}}{\text{(likelihood with the variable)}}\right]$$

The G statistic follows a chi-squared distribution and can be used to determine probabilities. The resulting p-value is the probability of the G statistic on the chi-squared distribution. This p-value is the probability of the null hypothesis ($\beta = 0$) being true given it was rejected, therefore it is necessary for the p-value to be small, or less than a significance level (Rice, 2007). The significance level for this testing was chosen to be 10% due to the small sample sizes and the nature of testing. Therefore, p-values less than 0.10 would indicate that the independent variable in the modeled relationship would be a significant factor to the outcome.

In the remaining section, the results from the testing will be discussed in terms of these two ideas: the odds ratio or how much more or less likely a characteristic may play over another and the p-value to determine if that ratio is statistically significant.

5.4.2 OVERALL RESULTS

In the results, there are two outcomes that are indicators of perceived model quality relative to what has been presented thus far (i.e. the y-axis in the 4-box model of Figure 1). The first is the design problem decision: whether to proceed with the design or wait for additional testing. This indicates how the respondents viewed the entire design problem, but did not necessarily speak to model perception, particularly if the decision was made with little influence from the model. Therefore, model confidence was another response obtained from the aposteriori survey question that was used to judge simply how the model, itself, was perceived regardless of the design decision that was made.

With this in mind, the first question of the overall data was whether the testing was effective at allowing perception factors to influence respondents' behaviors. If the decision was obvious, then there would be no influence of other factors like time pressure or consequences. To test this, the decision outcome was compared against the quality of the model used for the case (Table 7). If the cases were too obviously related to model quality, then people would always choose to proceed with a good model and to wait with a bad model. In other words we would expect that all test subjects would only fall into quadrants II and III. The results showed close to even distributions between a "proceed" and "wait" decision regardless of model quality. This shows that respondents were influenced by other factors, like those from the framework, when making their decisions.

Design Decision Outcome by Model Quality

Model Quality (Good Model is Baseline)	Decision Outcome Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Proceed})}{p(\text{Wait})}$
	Proceed	Wait			
Good Model	43%	57%	-	-	-
Bad Model	51%	49%	-0.60	0.170	0.55

Table 7: Distribution of decisions based on good or bad models from the cases. The distribution shows about even distributions between proceed and wait regardless of model quality indicating the problem was not obvious.

Knowing that factors contributed to people's decisions, an overall regression analysis was done comparing the factor being tested in each case against the reference. The dependent variable in this analysis was the decision outcome of proceed or wait. The results showed that only the purpose of analysis factor was effective at significantly changing the design problem decision (Table 8). The data was also analyzed for each factor's overall impact on model confidence. Model confidence was rated in the survey on a scale from one to five (five being the highest level of confidence) and the average is plotted by factor in Figure 37. Other measures recorded via the survey could be mined in the future to further refine the dataset and understand to what extent each of the factors impacted the decision process. The remainder of this section will explore further each of the factors individually relative to the additional data obtained from the experiments.

For each of the factors, a similar analysis approach is taken. First, the data was analyzed to be sure the experiment was effective at testing the factor. This was judged using survey questions specifically targeting each factor. Once confirming the experiment was a successful test, the data was analyzed to understand how the factor may have impacted the decision or the model confidence.

Design Decision Outcome by Factor Tested

Factor (Reference is Baseline)	Decision Outcome Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Proceed})}{p(\text{Wait})}$
	Proceed	Wait			
Reference	60%	40%	-	-	-
Consequences	45%	55%	-0.60	0.170	0.55
Purpose of Analysis	39%	61%	-0.85	**0.054**	0.43
Model Source	44%	56%	-0.64	0.139	0.53
Time	49%	51%	-0.45	0.294	0.64
Uncertainty	45%	55%	-0.62	0.167	0.54

Table 8: Analysis of the decision outcome by the factor test case as compared to the reference case. All factors show no significant difference in the decision outcome as compared to the reference except for the Purpose of Analysis factor where there is a higher likelihood of choosing to "wait" with this case as compared to the reference.

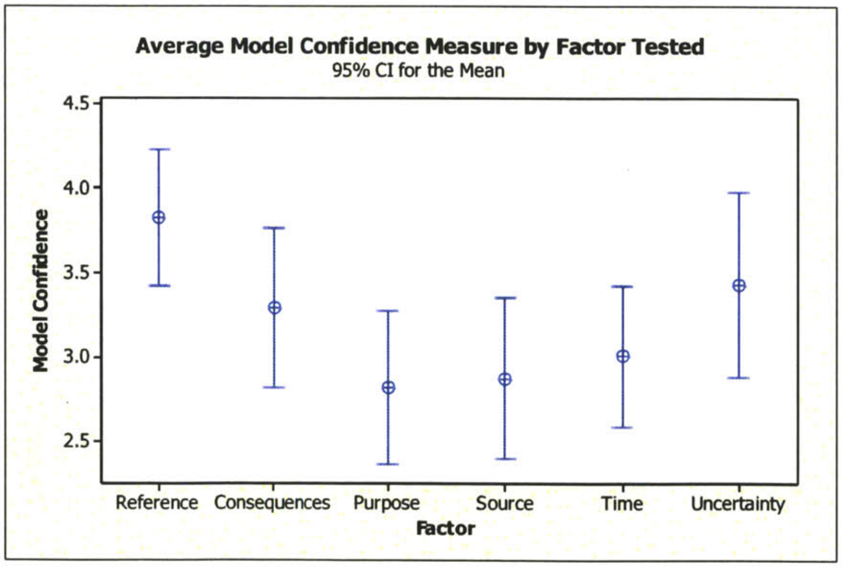


Figure 37: Average of model confidence (dots) plotted against the test factors. Purpose, Source, and Time factors all show a significant reduction in confidence as compared to the Reference case taking into account the 95% confidence interval.

Based on the results it appears that the different factors influenced mean model confidence significantly but did not clearly bias the actual binary decision one way or the other except for the purpose of analysis. Another way to say this is to say that of all the factors tested the one most likely to lower confidence in a model is when decision makers are asked to base a decision on a model that was originally designed for a different purpose. Not only did this lower confidence in the model, it also made decision makers more risk averse (i.e. wait 69% versus accept the current design 31%). The next strongest factors influencing model credibility were the source of the model and the exercising of time pressure.

5.4.3 CONSEQUENCES

Changing whether there was financial risk to the design decision in the experiment tested the effect of consequences. In the case of the reference test, there was money at stake for early contracts and possible penalty of warranty for a defective design; whereas in the consequences test case, there were no risks dependent on the decision. Initial inspection of the full dataset was done by comparing the survey response asking whether the consequence influenced the decision. The consequences test responses showed a statistically significant difference from the reference case indicating those faced with higher consequences and financial downside risk, as in the reference, were more likely to be influenced when making a decision, all else being equal (Table 9). This result corresponds with many of the comments provided in the survey where respondents expressed concern for making it to market first or being subject to potential warranty exposure. This proves that consequences were a contributor to the decision.

Influenced by Consequences by Factor Tested

Factor (Reference is Baseline)	Influenced by Consequences Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Yes})}{p(\text{No})}$
	Yes	No			
Reference	79%	21%	-	-	-
Consequences	**43%**	**57%**	**-1.59**	**0.002**	**0.20**
Purpose of Analysis	84%	16%	0.32	0.592	1.38
Model Source	68%	32%	-0.55	0.300	0.58
Time	76%	24%	-0.14	0.801	0.87
Uncertainty	78%	22%	-0.07	0.910	0.93

Table 9: Analysis of the Test Cases as compared to the Reference for whether the consequences influenced the ultimate decision for the design problem. The consequences test case was the only factor that showed a significant difference from the reference. The odds ratio indicates that with consequences removed, as in the test case, the respondents are 20% less likely to be impacted by the consequences when making their decision.

The hypothesis presented in section 4.1 suggested that the perception of the model would be impacted by consequences depending on whether they were beneficial or detrimental (Figure 23). Because the reference case had both depending on the decision, it was necessary to determine how the respondents felt was a stronger consequence to know how their perception would be impacted. To do this, it is necessary to calculate the expected value (EV) of the decision with consequences as is done in Figure 38. In this calculation, the probabilities of the model being correct are left unknown and shown as p being the probability that the model is correct and 1-p being the model is wrong.

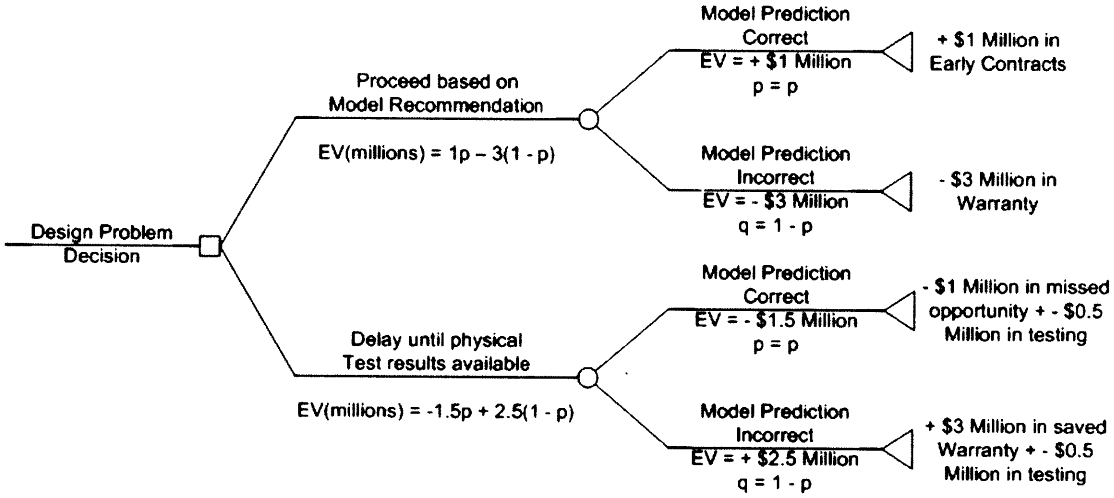


Figure 38: Decision tree for Reference test case including Expected Value calculations

At 50% probability that the model is correct, the EV of proceeding is –$1 million whereas delaying would have an EV of +$0.5 million. Knowing nothing else, the best choice would be to choose to delay as it has a higher EV. However, the respondents knew more about the model than 50/50 odds and its validation allowed to make a better determination of p and therefore know whether the consequences were more beneficial or detrimental with the design. The minimum value of p to make the EV of proceeding greater than waiting can be calculated with the following inequality (Equation 1). Here, if the respondents are more than 68% confident in the model, they will believe the consequences to be beneficial and may perceive the model more positively, whereas anything less than 68% would be considered detrimental and may bring about risk-averse behavior.

$$\begin{aligned}&EV_{proceed} > EV_{delay}\\&4p-3 > -4p+2.5\\&\therefore p > 68.75\%\end{aligned}$$

Equation 1: Expected Value Inequality

For this problem, the model confidence was binned in order to align with indicating a probability greater than 68% or not. Therefore, model confidences of four and five were made "high" and the rest were "low." Only for those that indicated they were impacted by the consequences in their decision for the reference case, which amounted to about 80% of the respondents, a regression analysis was done to determine the impact of confidence on the ultimate decision. The results shown in Table 10, indicate that the reference case showed a significant impact of model confidence on the ultimate decision where the higher the confidence, the more likely the respondents were to choose to proceed with the design, regardless of model quality. Although this seems an obvious conclusion, it is more interesting when compared to the consequences test case analysis. Here, the same test was done and found to be inconclusive, or model confidence did not impact the ultimate decision because there was no influence from beneficial consequences in the case of high confidence, or detrimental consequences in the case of low confidence.

Design Decision Outcome by Factor Tested when Consequences Influenced the Decision

Factor	Model Confidence (High Confidence is Baseline)	Decision Outcome Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Proceed})}{p(\text{Wait})}$
		Proceed	Wait			
Reference	High Confidence	79%	21%	-	-	-
	Low Confidence	18%	82%	-2.83	0.003	0.06
Consequences	High Confidence	43%	57%	-	-	-
	Low Confidence	33%	67%	-0.41	0.697	0.67

Table 10: Analysis results for determining decision outcome against model confidence for the reference and consequences test cases for those respondents who indicated the consequences impacted their decisions. The Reference shows a statistically significant impact indicating that as confidence drops, the respondent is more likely to wait on the design problem regardless of the model's quality. This is in contrast to the consequences test cases where decisions were made to proceed or not regardless of confidence in the model.

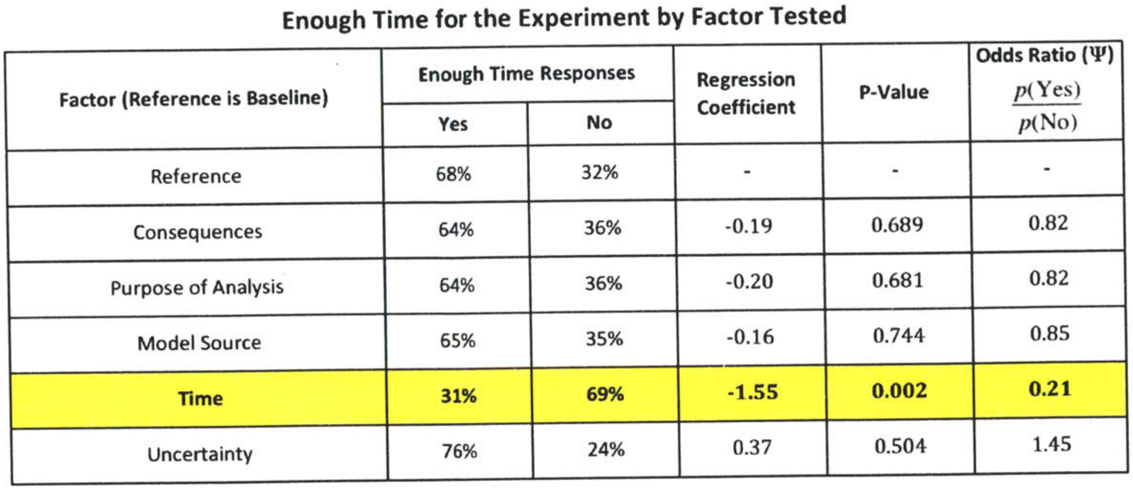
Aligning model quality with its perception is highly correlated with the decision being made. The solution, however, cannot mimic this experiment where the consequences were removed all together. One key to making model-based design more effective in these situations is to raise the value of p, or the probability the model is correct, from the perspective of the decision maker. This includes improving the model's quality itself, but also ensuring adequate communication of the model and its validation to reduce potential uncertainties from the decision maker. Using the experiment as an example, for the good model cases, it was necessary to portray to the user that the model was truly good enough to make a decision to proceed. In the cases with the bad model, there was no opportunity given to improve the model itself, but proper validation communication would still be effective because it would have lowered the confidence in the model, driving people to wait for testing.

Another option for more effective model-based design is to add flexibility where new decision paths may allow for improved beneficial consequences. In the experiment, users were given two options: proceed or wait. However there could be alternatives that would allow for higher EV even in the face of model uncertainty. A third option could be considered where the design would proceed under conditions where the model had been proven valid and parallel testing activities would continue to confirm functionality under the new requirements where the model was less certain. For example, the model provided for the experiment had been tested and validated with two rubber bands and a specific ball. Could the design be released in this configuration at elevated product costs? Once confirmation testing had been completed and valid design was ready, this could be released as a means to lower product costs for the remainder of its life cycle. In this scenario, the early contracts would be awarded and the potential warranty costs would be avoided.

The consequences factor proved to be an important variable in how people approached the design decision. Regardless of a model's quality, people were influenced by whether they believed the consequences to be more beneficial or detrimental. This can hinder the implementation of model-based design in organizations particularly where consequences are large. However, by improving model validation techniques using some of the methods presented in section 3.4, the decision tree can better align with the actual quality of the model. More importantly, by using principles of flexible design, the decisions can be setup to make the best use of scenarios where the model is most certain and still avoid penalties where the model is less certain.

5.4.4 TIME PRESSURE

Time pressure was tested in this experiment by allowing 15 minutes with the model for the reference case and only 4 minutes in the high time pressure test cases. Based on initial inspection of the data, the distinction in time was significant enough to cause a separation between the cases where the time test cases were five times more likely to respond that there was not enough time to make a decision as compared to the reference (Table 11). This time pressure did however not impact whether people chose to proceed with the design or not as compared to the reference and therefore, they did not perceive the problem differently. However, their confidence in the model was significantly reduced (Figure 37) and further analysis was needed to understand the impact.



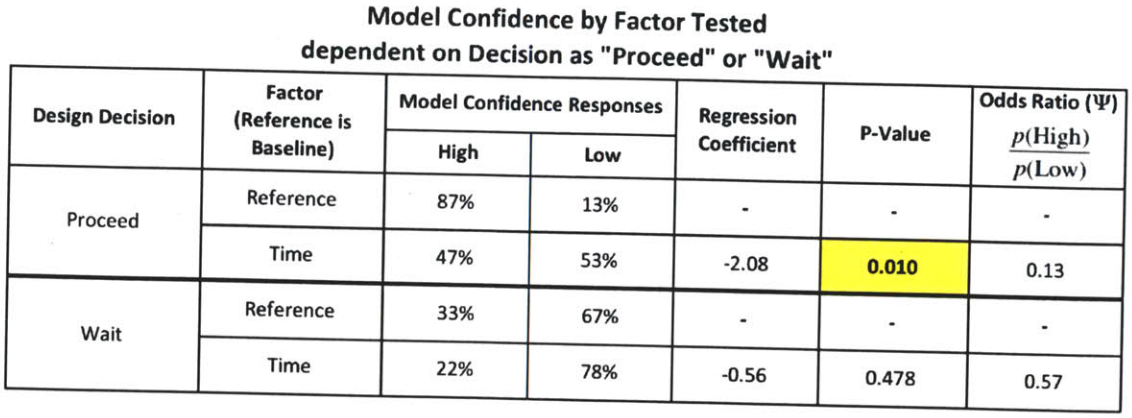
Enough Time for the Experiment by Factor Tested

Factor (Reference is Baseline)	Enough Time Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Yes})}{p(\text{No})}$
	Yes	No			
Reference	68%	32%	-	-	-
Consequences	64%	36%	-0.19	0.689	0.82
Purpose of Analysis	64%	36%	-0.20	0.681	0.82
Model Source	65%	35%	-0.16	0.744	0.85
Time	**31%**	**69%**	**-1.55**	**0.002**	**0.21**
Uncertainty	76%	24%	0.37	0.504	1.45

Table 11: Regression analysis on whether the respondents had enough time to make a decision by the test cases. The Time Pressure test case was the only case where there was a significant difference from the reference where people were more likely to say that there was not enough time.

The hypothesis presented earlier suggested that time pressure would improve perception when the respondents trusted whatever inputs were available or would seek other sources if they did not have confidence in the model. To test this, the confidence was tested against each decision for both test cases. The same binning used in the previous section was upheld for consistency. For those that chose to wait, there was no difference between the reference and time pressure cases where in each case lower confidence in the model generally resulted in a "wait" decision. However, where the respondents chose to proceed with the design, the impact of model confidence was different between the reference and time pressure (Table 12) regardless of model quality. This results in the Time Pressure cases to be more likely to choose to proceed with low model confidence as compared to the reference case. In fact, the results showed that about 85% of the reference case respondents who chose to proceed had high confidence, where the time pressure case was about half and half on confidence regardless of model quality. This suggests that when people do not have time to make a rational decision based on data but are forced to do so, they come out with about the same odds as the decision itself. In the case of a binary decision this is similar to a coin flip.

Further investigation was done, based on the hypothesis, that other factors would force the decision. However, there were no ties to a change in the influence of consequences, for example. Therefore, the hypothesis presented earlier is not fully validated. When there is high confidence in the model, the perception of the model (or the ultimate decision) does not necessarily follow in the case of high time pressure. Similarly, other sources were not sought to make the decision such as the impact of consequences. Instead, people chose a decision and the result came out to about an even number of cases between a proceed or wait decision.



Model Confidence by Factor Tested
dependent on Decision as "Proceed" or "Wait"

Design Decision	Factor (Reference is Baseline)	Model Confidence Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{High})}{p(\text{Low})}$
		High	Low			
Proceed	Reference	87%	13%	-	-	-
	Time	47%	53%	-2.08	**0.010**	0.13
Wait	Reference	33%	67%	-	-	-
	Time	22%	78%	-0.56	0.478	0.57

Table 12: Analysis of the impact on confidence by test case dependent upon the design decision made. When respondents chose to wait, there was no difference in model confidence between the reference and time pressure cases. However, when they chose to proceed, there was a significant difference in confidence between the two test cases. Here, the time pressure case was much more likely to have low confidence in the model yet respondents still choose to proceed.

What was clearly determined, was that under time pressure, model confidence is reduced. Therefore, under time pressure, the full benefit of model-based design in organizations would not be realized and decisions would not be data-based but more akin to a random coin flip. Over time, people would not grow to trust models. This is where flexible design becomes important as discussed in the previous section. There, the example was given to proceed using a case where the models were validated (i.e. two rubber bands) and proceeding with confirmation testing as well under the new requirements. Although there are additional costs for releasing a more expensive product, those can be removed in further iterations and the risk of warranty cost is reduced and benefits of arriving early to market are realized.

5.4.5 PURPOSE OF ANALYSIS

The purpose factor required the respondents to evaluate the maximum height of the ball at its apex as opposed to its horizontal distance. Given that the model did not provide a height output, this was very difficult to determine; only an image of the ball flying through the air against a background grid could be used to judge height achieved. In fact, qualitatively, this factor caused the most distress in the respondents. Several people in this test case expressed frustration in the model's inability to predict height via emails as well as comments in the survey. This distress was evident in the overall results from the experiment where this factor was the only one where a significant difference was noted in the decision outcome as compared to the reference case (Table 8) where people were more likely to wait on the results of the physical design testing regardless of model quality. It follows too that the confidence in the model for the purpose test cases was more likely to be lower as compared to the reference (Figure 37). These findings agree with the hypothesis that asking questions from a model for which it was not originally intended will generally lower quality perception of that model.

What is of most interest, however, is understanding why people chose to proceed anyway with a design even when the model was clearly not suitable for the problem similar to what was found in the case studies. Further investigation of the results may help inform why this continues to be a problem in organizations. Reading through the comments for those that had been selected for the test group and chose to proceed with the design, there were two common themes that described why people chose to move forward anyway:

- The consequence of losing first-to-market advantage was too great
- Assumed the model was correct and interpreted it as best as they could

The first item was discussed in the earlier section on consequences. The second item, however, highlights a weakness. People were provided with a measure of height, however subjective that measure was. This was done in order to improve the fidelity, or the realism of the model where people could pull back the catapult arm, launch the ball and watch the ball in flight. This improves the experience of the decision maker (Grogran, Lee, & de Weck, 2011). However, there is a tradeoff that additional information used in a model should be validated before it is incorporated to protect a model's credibility. Based on examples from this experiment, and lessons from the case studies such as in the Columbia accident, decision makers will seek out any sources of information to better inform themselves on the decision to be made.

5.4.6 UNCERTAINTY

The impact of uncertainty in the experiment was tested by showing error bars on top of the validation data for the test case whereas the reference case had no error bars. The intent was to allow the respondents to ascertain whether their validation runs were valid relative to the provided data. Based on the overall results, there was no overall impact on the decision or model confidence with uncertainty communicated (Table 8 and Figure 37). To measure the effectiveness of the factor in the experiment, the responses were compared between the stated capabilities of a catapult against that of the model. The uncertainty test case respondents knew the catapult's capability to reproduce its results from the error bars in the validation data – the data showed the catapult was fairly poor at reproducing distance. Based on the results (Table 13), the uncertainty test case was more likely to predict poor capabilities for both the catapult and the model as expected. This indicates the factor was effective in testing.

Catapult and Model Capability by Factor Tested

Catapult or Model Capability	Factor (Reference is Baseline)	Capability Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Good})}{p(\text{Bad})}$
		Good	Bad			
Catapult	Reference	75%	25%	-	-	-
	Uncertainty	45%	55%	-1.31	**0.015**	0.27
Model	Reference	59%	41%	-	-	-
	Uncertainty	35%	65%	-0.97	**0.061**	0.38

Table 13: Statistical analysis of stated capabilities of a catapult and the model comparing the Reference to Uncertainty cases. In both cases, there is a strong correlation between people's stated capabilities and the test case.

The hypothesis presented in section 4.5 indicated that with communicated uncertainty, the perception of a model would improve because by presenting uncertainty information decision makers would gain more clarity as to the validity bounds of the model. However, it was found that the decisions did not change significantly between the uncertainty and reference test cases. The confidence in the model was compared between the reference and uncertainty cases to see if the perceived capability of a catapult and model translated to the confidence in the model (Table 14). Again, the case for uncertainty showed no change in trend from the reference case.

Based on these results, the hypothesis for uncertainty could not be validated. Therefore, although the uncertainty was communicated and showed people a reduced level of capability in the model and catapult, it did not impact their decision and therefore their overall perception of the model as a decision-making tool.

This is a very curious result as much of the literature, as discussed earlier, is clear about the value of communicating uncertainty and its value on improving the use of models to make decisions. In this example, data shows that the uncertainty was properly communicated, but made no difference in how people interpreted that knowledge into their decision. Therefore, more needs to be understood with respect to uncertainty. For example, what form does uncertainty need to be in for better communication and reception by the decision makers? In this experiment, the uncertainty was in the distance, but perhaps a better uncertainty would be in the final decision: there is X% of being correct (p-value from the decision tree in Figure 38) resulting in Y expected value. This statement is much more intuitive to interpret and puts the uncertainty in terms that are directly applicable to the decision maker.

Model Confidence by Factor Tested dependent on Model Quality

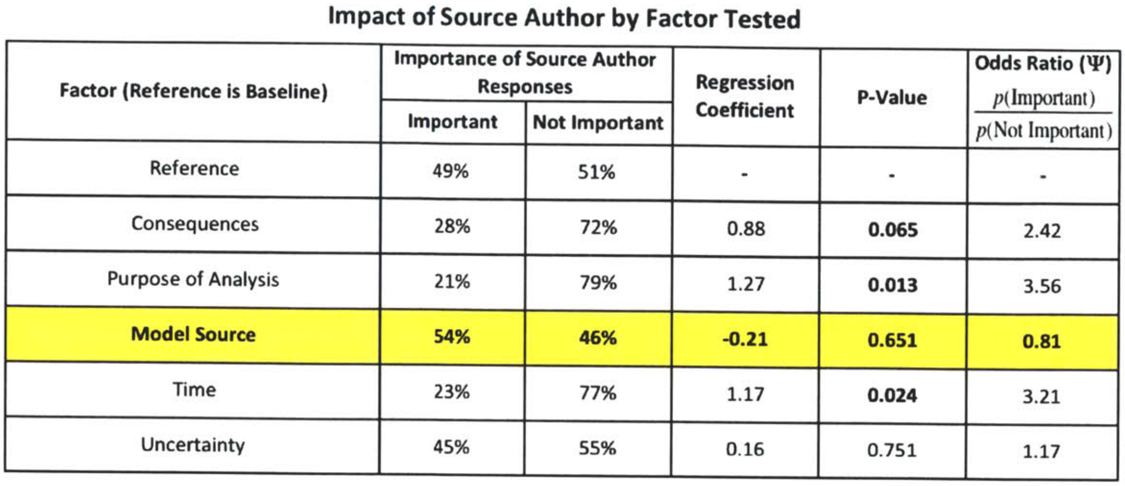
Model Quality	Factor (Reference is Baseline)	Model Confidence Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{High})}{p(\text{Low})}$
		High	Low			
Overall (all model quality cases)	Reference	66%	34%	-	-	-
	Uncertainty	61%	39%	-0.22	0.672	0.80
Good Model Quality	Reference	65%	35%	-	-	-
	Uncertainty	64%	36%	-0.03	0.966	0.97
Bad Model Quality	Reference	67%	33%	-	-	-
	Uncertainty	57%	43%	-0.41	0.582	0.67

Table 14: Tables comparing model confidence between Reference and Uncertainty test cases. Confidence is similar for both cases regardless of model quality.

5.4.7 MODEL SOURCE

The final factor that was explicitly tested was the source of the model. In the reference case, the model was from a PhD student from MIT and was based on physics principles. In the source test cases, the model was downloaded from the internet and the governing principles in the model were unknown (black box model). Using survey responses indicating the importance of the model's source in the decision, the data was able to determine if this factor was effective in the experiment. However, results showed no difference in whether the source author made a difference between the reference and source test cases (Table 15). Interestingly, a few of the other test cases did. The consequences, purpose, and time factors all behaved differently as compared to the reference. The odds ratio suggests these cases were less likely to be impacted by the source author. This is likely because the other test factors in these cases overshadowed the model's source.

To analyze the case of the model source author further, focus was placed on the datasets that indicated importance of the source author between the reference and source test cases. About 50% of the cases for each found the model's source to be important. The design decision outcome was analyzed based on the reference or source case as well as the model quality for the experiment. There was an interesting outcome in this analysis (Table 16). The results showed that when the source of the model was important to the decision maker, then the source test cases were more likely to choose to wait on the design decision as compared to the reference with statistical significance. This was true for the good model, and almost exclusively the case for the bad model. This indicates that the hypothesis presented earlier suggesting that the perception will change depending on the trustworthiness of the source was able to be validated. In the source test cases, the model's source was less trustworthy in having no specified author and no understanding of the concepts comprising the model. This created enough uncertainty with the respondents that they tended to want to wait and not proceed with product launch.



Impact of Source Author by Factor Tested

Factor (Reference is Baseline)	Importance of Source Author Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Important})}{p(\text{Not Important})}$
	Important	Not Important			
Reference	49%	51%	-	-	-
Consequences	28%	72%	0.88	**0.065**	2.42
Purpose of Analysis	21%	79%	1.27	**0.013**	3.56
Model Source	**54%**	**46%**	**-0.21**	**0.651**	**0.81**
Time	23%	77%	1.17	**0.024**	3.21
Uncertainty	45%	55%	0.16	0.751	1.17

Table 15: Analysis results comparing each of the test cases and the impact of the source author on those cases. The source test case showed no significant difference as compared to the reference although consequences, purpose, and time did vary from the reference case. In each case, the odds ratio suggests that these cases were less likely to be impacted by the model source.

To understand this problem further, the model source is broken into two attributes: the trust-worthiness of the author and the proprietary-nature of the code whether it is open or closed to the model users. This experiment tested two extremes relative to these attributes. First, the reference case represented a model from a trustworthy source (MIT PhD in Mechanical Engineering) and an open model (physics-based principles) and the results showed acceptance of this. In the source test case, the opposite scenario was presented: an unknown, therefore potentially untrustworthy source combined with a proprietary model with unknown governing equations. The data showed people's hesitation to this combination. Evaluating the other two cases may help to determine possible solutions to model source concerns in decision-making.

The first of these cases is when the model comes from a potentially untrustworthy or unknown source but its code is open for inspection. In this case, whether a model is used should depend on how the model performs in validation. Because the source code is open, validation is a step that can be performed well in order to understand how well the model can support a decision-making process.

The alternate case is when the model is from a trustworthy source, but the code is proprietary. This happens often in industry where third-party consultants sell models but protect their intellectual property by locking down the code of the model. In these cases, validation can be done around the known problem space, but little can be done to understand how well the model can extrapolate to new regions in the space. In this case, it is important to review the credibility of the source and the experience behind the model and its previous applications.

From these two scenarios, it is clear that it is important to validate the available information whether it be the credibility of the source or the validation and transparency of the model or both.

Design Decision Outcome by Factor Tested
when Source Author is Important for Good and Bad Model Quality

Model Quality	Factor (Reference is Baseline)	Decision Outcome Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Proceed})}{p(\text{Wait})}$
		Proceed	Wait			
Good Model	Reference	73%	27%	-	-	-
	Source	8%	92%	-3.38	**0.007**	0.03
Bad Model	Reference	87%	13%	-	-	-
	Source	44%	56%	-2.17	**0.086**	0.11

Table 16: Design Decision results compared against source cases and model quality where model source was important. For the source test case, with both the good and bad models, the respondent's would more likely choose to wait as compared to the reference.

5.4.8 IMPLICIT FACTORS FROM THE EXPERIMENT

Three remaining factors: availability of models, requirements and their validation, and fragmented knowledge were difficult to test in the experiment form used as they are more dependent upon factors among organizations as a whole. Survey questions were included to try to understand more about the population relative to these factors and those will be shared.

Model availability was examined by three survey questions. First, the users were asked if models were generally available in their organizations, second if those models were designed for the purpose of the problem and finally, a sense for model availability based on whether the users' organizations generally did more physical testing or modeling during their product development activities. The responses for the population generally indicated that models were available and often for the required purpose (Table 17). However, whether people had models available or not did not show significant relationships with any responses such as the decision, model confidence, or capabilities in the model or catapult. This factor requires testing in a more disciplined manner in an environment specific to an organization where more than qualitative responses could be used to judge model availability.

Requirements and their validation was another difficult factor to test using this experiment methodology. This factor was surveyed by asking the respondents to rate their organizations with regard to setting requirements and validating them. As before, the responses showed that generally this practice was done well within organizations and there was, again, no relationship in this response to other outcomes from the experiment. Similar to the model availability, this factor requires a different survey method with quantitative inputs.

Are models available?	
Yes	No
66%	34%

Do they match the purpose?	
Yes	No
77%	23%

Does your organization depend more on physical testing? Or modeling?	
Physical Testing	Modeling
53%	47%

Table 17: Survey responses relative to model availability

The final implicit factor tested was fragmented knowledge within the organization. Given the experimental method, it was not feasible to understand how well knowledge was managed within an organization and attempting to do so would likely lead to results similar to those for the model availability and requirements. Therefore, the respondents were asked if they:

a. Were familiar with catapults and the physics-based concepts

b. Were familiar with creating and using mathematical models

Using the responses from the above two questions, four categories were created indicating the level of knowledge: familiar with models and catapults, neither, or only one or the other. These categories were then tested against the design decision and model confidence. Table 18 shows there was no impact to the overall decision as a result of knowledge. Although the case with model familiarity and without knowledge of a catapult had response percentages that trended towards more waiting in the decision. However, the sample size of this group was very small compared to the others which explains the statistical insignificance of this trend. Table 19 shows the same analysis with model confidence. Here, there is a significant difference in being familiar with catapults but not models. These cases proved to be more likely to have confidence in the model as compared to those who either had no knowledge or were more familiar with modeling practices. This is an interesting result that suggests parallels with MacKenzie's certainty trough (1990). Those who are far from the problem (no familiarity) have high levels of uncertainty (indicated by low model confidence). Similarly, those are very close to the problem (model familiarity) have high levels of uncertainty as well. Where those familiar with the real system and not models are in the trough, or have high model confidence.

More testing is required to fully understand the impact of this factor as these preliminary analyses suggest important findings, but they lack enough detail to make specific recommendations.

Design Decision Outcome by System Familiarity

System Familiarity (No Familiarity is Baseline)	Decision Outcome Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{Proceed})}{p(\text{Wait})}$
	Proceed	Wait			
No Familiarity with Cataults or Models	41%	59%	-	-	-
Familiar with Catapults NOT Models	52%	48%	0.42	0.209	1.53
Familiar with Models NOT Catapults	27%	73%	-0.65	0.301	0.52
Familiar with BOTH Catapults and Models	45%	55%	0.16	0.640	1.18

Table 18: Analysis of the design decision by familiarity of the system. This analysis showed no significant difference between system knowledge levels and the final decision outcome.

Model Confidence by System Familiarity

System Familiarity (No Familiarity is Baseline)	Model Confidence Responses		Regression Coefficient	P-Value	Odds Ratio (Ψ) $\frac{p(\text{High})}{p(\text{Low})}$
	High	Low			
No Familiarity	46%	54%	-	-	-
Familiar with Catapults NOT Models	61%	39%	-0.59	**0.084**	0.55
Familiar with Models NOT Catapults	40%	60%	0.25	0.667	1.28
Familiar with BOTH Catapults and Models	44%	56%	0.09	0.801	1.09

Table 19: Analysis of the model confidence by familiarity of the system. This analysis showed that familiarity with catapults but not models had significantly higher confidence in the model as compared to the other cases.

6 CONCLUSIONS

6.1 SUMMARY

Through the course of this thesis, the goal has been to understand the challenges in model-based design within organizations as related to model confidence and validation. Using case study analysis from three different real world cases, two primary challenges emerged: the problems of model validation and model perception. A thorough literature review further analyzed these problems to uncover the underlying questions behind these problems: 1) ensuring proper model assessment 2) understanding the problem from the perspective of the decision-maker and 3) awareness of potential contextual variables that may overwhelm the input data to a decision-making process.

Much has been done in the literature to address the first question regarding model assessment; NASA's standard NASA-STD-7009 used to assess model credibility is a good example of an assessment to help ensure that models are good for use. However, the latter two issues relate more to the decision-makers and how they perceive the model, a realm not addressed well by literature.

Using lessons from the case studies, literature, and industry experience, eight factors are proposed that, when merged with a mature model assessment process, may positively influence decision-makers. By positively influence we mean that decision makers should be able to recognize bad models and not use their output during decision-making (quadrant III) and that they should have confidence in good models and use them for decision-making (quadrant II). The goal is not to convince them to use a model's result in every design problem, but rather to better align their perception with the actual quality of the model result relative to the design problem.

The 8 factors influencing perception of model quality and potentially decision outcomes are as follows:

- Effect of a decision's consequences
- Effect of time pressure under which to make a decision
- Availability of mature models to use for design problems
- Purpose for which the model is being used relative to its originally intended purpose
- Uncertainty and its communication
- Lack of system-level requirements and their validation
- Model's source author and transparency
- Fragmentation of knowledge through organizations.

These factors were tested for their true impact using a simple model in a web-based experiment. Results from the experiment showed interesting findings that proved that the framework proposed is a good start toward a further understanding these problems. It illustrated some particular areas of research and suggested some potential methods that could help alleviate the negative impact these factors might have on a decision-making process. A summary of the findings from the experiment is shown in Table 20.

Factor	Hypothesis	Hypothesis Validated	Summary of Findings
Effect of Consequences	Prediction of beneficial consequences will improve model perception while detrimental consequences will reduce perception of model quality	Yes	The expected value of the options (based on model confidence) drove the decision when consequences were present. When consequences were removed, model confidence no longer impacted the decision outcome.
Effect of Time Pressure	Applying time pressure will cause people to use whatever inputs are available in the process whether validated or not	Yes	The threshold of confidence in the model needed to proceed with the design was significantly reduced when applying time pressure.
Purpose of Analysis	By asking questions of the model for which it was not originally created, the quality perception of that model will be reduced	Yes	This factor significantly changed the design decision from the reference to more likely wait.
Uncertainty and its communication	Improving information about uncertainty will lead people to better recognize a good or bad model	No	Results showed this factor was properly tested based on responses of catapult capability changing from the reference, but there was no significant difference in the decision outcome or the model confidence
Source and transparency of the model	A more trustworthy model author and transparent governing equations will improve model perception	Yes	For the cases where the source was important to the decision, there was a significant difference in the decision outcome where untrustworthy sources caused reduced confidence.
Fragmented Knowledge	Better knowledge of the entire problem will cause the perception of the model to increase	Inconclusive	Initial results showed possible connection between knowledge and model confidence however this factor requires more testing to explore this factor further.

Table 20: Summary table of factors tested in the experiment, their hypotheses and a review of the findings from the experiment results.

6.2 RECOMMENDATIONS

There were some overriding themes that came from the analysis that can serve as recommendations to organizations looking to improve their model-based design practices. First, ultimately the decision involved in a design problem can be mapped on a decision tree where the probabilities of model correctness can be used to provide a quantitative assessment in terms of the expected value of the decision options. This then allows evaluation of flexible design options where the risks of acting on a bad model can be mitigated with minimized cost involved in rejecting a good model. Therefore, whatever the consequences of a decision or the scheduled deadline under which a decision must be made, the organization can still take advantage of model-based design practices with higher confidence and lower risk.

Communication across an organization is another recommendation from the results. This is also covered in the literature as an important aspect, although few solutions are available and proven to improve this. However, this thesis illustrates the impact it can have when the purpose of a model is miscommunicated either from the decision makers in their request for a model or from them modelers in their intended purpose of a model. Besides model purpose, general understanding of the problem space, either of the real system itself or the modeling principles, can vary the tendencies of the decision-makers. Therefore it is necessary to ensure all parties in an organization have the needed background information of the system, not just the model data, before making a decision.

Communication is also raised as a point in the literature as necessary with regard to model uncertainty information. This experiment showed that knowledge of model uncertainty provided no benefit to the decision. However, it also touched on another point that may improve how information is communicated. The uncertainty in this case was not provided as a probability related to the decision, but as input variables to the decision. Putting the uncertainty in terms of system-level outcomes may aide in decision-makers understanding of the impact of uncertainty.

Models can be a useful tool for design and communication. The latter, in particular, is improved through added fidelity or realism. However, the trade off lies with the added validation work required for all the parameters in a model. Results showed people would use this information, even if not validated. Therefore, thought and care needs to go into how information is communicated, whether that information is in the right form relative to system requirements, and if all pieces of information are properly validated.

Implementing these recommendations could help alleviate some of the challenges in model-based design because they were shown, by means of the experiment, to have an impact on model confidence.

6.3 FUTURE RESEARCH

This thesis highlights a new approach to addressing model-based design challenges by focusing less on the validation steps and more on how contextual factors can influence decision-makers. There is much more that can be done under this new analysis technique as new case studies present challenges to the proposed frameworks and as more confirmation under this new premise is completed. Of particular interest are cases in quadrants I and IV.

One area of future research to help address this new approach is to understand the interactions of the factors in the framework. Due to time constraints, a one-factor-at-a-time DOE was run to learn about the main effects of the framework, but it was not possible to glean how multiple interacting factors might alter behavior. These interactions would be more representative of real product development environments and could add more fidelity to this model. It is estimated that a full factor experiment might require about 1,280 test subjects, about five times more than the 252 respondents that participated in the experiment presented in this thesis.

Other areas of future research are based on findings from the experiment conducted. First, improved communication is highlighted as a strong recommendation, which is consistent with volumes of other literature. The question then remains, why has this problem not yet been addressed? And what could be done to improve organizations communication and knowledge management.

The final point is in regards to the findings specific to model uncertainty communication. Literature continues to highlight the benefit of communicating uncertainty about a model and the design problem. However, results from the online experiment showed no benefit of it, despite recognition of the uncertainty. Therefore, more needs to be understood here. First, after numerous recommendations in the literature, why is uncertainty still a problem? Second, how can uncertainty be better communicated so as not to overwhelm decision-makers, but rather make it digestible to them? Would referencing uncertainty relative to system-level outputs as opposed to a single model's output improve this behavior as suggested in the recommendations?

This page intentionally left blank

7 REFERENCES

AAAS. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Ahn, J., & de Weck, O. (2007, May 15, 2007). [Pilot Study: Credibility Assessment of SpaceNet 1.3 with NASA-STD-(I)-7009].

Alemanno, A. (2011a). Introduction. In A. Alemanno (Ed.), Governing Disasters: The Challenges of Emergency Risk Regulation (pp. xx-xxxv): Edward Elgar Publishing.

Alemanno, A. (2011b). What Happened and Lessons Learned: A European and International Perspective. In A. Alemanno (Ed.), Governing Disasters: The Challenges of Emergency Risk Regulation (pp. 1-10): Edward Elgar Publishing.

Aviation Week and Space Technology. (1990). Volcanic Ash Cloud Shuts Down All Four Engines Of Boeing 747-400, Causes $ 80 Million in Damage. *Aviation Week and Space Technology, 132*, 93.

Balci, O. (1989). *How to assess the acceptability and credibility of simulation results*. Paper presented at the Proceedings of the 21st conference on Winter simulation, Washington, D.C., United States.

Balci, O. (2001). A methodology for certification of modeling and simulation applications. *ACM Trans. Model. Comput. Simul., 11*(4), 352-377. doi: 10.1145/508366.508369

Balci, O. (2003). *Verification, validation, and certification of modeling and simulation applications: verification, validation, and certification of modeling and simulation applications*. Paper presented at the Proceedings of the 35th conference on Winter simulation: driving innovation, New Orleans, Louisiana.

Balci, O. (2004). *Quality assessment, verification, and validation of modeling and simulation applications*. Paper presented at the Proceedings of the 36th conference on Winter simulation, Washington, D.C.

Balci, O., Nance, R. E., Arthur, J. D., & Ormsby, W. F. (2002). *Improving the model development process: expanding our horizons in verification, validation, and accreditation research and practice*. Paper presented at the Proceedings of the 34th conference on Winter simulation: exploring new frontiers, San Diego, California.

Balci, O., Ormsby, W. F., John T. Carr, I., & Saadi, S. D. (2000). *Planning for verification, validation, and accreditation of modeling and simulation applications*. Paper presented at the Proceedings of the 32nd conference on Winter simulation, Orlando, Florida.

Balci, O., & Saadi, S. D. (2002). *Simulation standards: proposed standard processes for certification of modeling and simulation applications*. Paper presented at the Proceedings of the 34th conference on Winter simulation: exploring new frontiers, San Diego, California.

Banks, J. (1998). *Handbook of Simulation - Principles, Methodology, Advances, Applications, and Practice*: John Wiley & Sons.

Bertch, W., Zang, T., & Steele, M. (2008). *Development of NASA's Models and Simulations Standard*. Paper presented at the Spring Simulation Interoperability Workshop (SIW), Pasadena, CA.

Bolić, T., & Sivčev, Ž. (2011). Eruption of Eyjafjallajökull in Iceland. *Transportation Research Record: Journal of the Transportation Research Board, 2214*(-1), 136-143. doi: 10.3141/2214-17

Bonczek, R. H., Holsapple, C. W., & Whinston, A. B. (1980). THE EVOLVING ROLES OF MODELS IN DECISION SUPPORT SYSTEMS*. [10.1111/j.1540-5915.1980.tb01143.x]. *Decision Sciences, 11*(2), 337-356.

Boston Museum of Science. (2001). Science Thinking Skills: Making Models (description) Retrieved 14 March 2012, from http://www.mos.org/exhibitdevelopment/skills/models.html

Box, G. E. (1979). Robustness in the Strategy of Scientific Model Building. In R. Launer & G. Wilkinson (Eds.), *Robustness in Statistics*. New York: Academic Press.

Brennan, J. J., & Elam, J. J. (1986). Understanding and validating results in model-based decision support systems. *Decision Support Systems, 2*(1), 49-54. doi: 10.1016/0167-9236(86)90120-x

Brooker, P. (2010). Fear in a handful of dust: aviation and the Icelandic volcano. [10.1111/j.1740-9713.2010.00436.x]. *Significance, 7*(3), 112-115.

Browning, T. R., Deyst, J. J., Eppinger, S. D., & Whitney, D. E. (2002). Adding value in product development by creating information and reducing risk. *Engineering Management, IEEE Transactions on, 49*(4), 443-458.

Brugnach, M., Tagg, A., Keil, F., & de Lange, W. (2007). Uncertainty Matters: Computer Models at the Science-Policy Interface. *Water Resources Management, 21*(7), 1075-1090. doi: 10.1007/s11269-006-9099-y

CAIB. (2003). Columbia Accident Investigation Board: Report Volume 1. NASA.

Cho, Y., Karkatsouli, I., Rasheed, B., & Saad, F. M. (2012). *Volcanic Ash and Aviation Safety: The Case of Eyjafjallajökull*. Paper presented at the MIT class ESD. 864 Modeling and Assessment for Policy, Cambridge, MA.

Clark, W. C., & Majone, G. (1985). The Critical Appraisal of Scientific Inquiries with Policy Implications. *Science, Technology, & Human Values, 10*(3), 6-19.

de Neufville, R., & Scholtes, S. (2011). *Flexibility in Engineering Design*. Cambridge, MA: The MIT Press.

Diaz, A. (2004). A renewed commitment to excellence: An assessment of the NASA agency-wide applicability of the Columbia accident investigation board report. NASA.

Dobrila, L. (August 8, 2011). *Personal interview*. [Technical Lead of System Integration].

DoD. (2006). VV&A Recommended Practices Guide. Available from U.S. Department of Defense Modeling and Simulation Coordination Office. (RPG Build 3.0). Retrieved April 1, 2012 http://vva.msco.mil/

Eckley, N. (2001). Designing effective assessments: The role of participation, science and governance, and focus. European Environment Agency.

Control of emissions from new and in-use nonroad compression-ignition engines, 40 C.F.R. § Part 1039 (2011).

Eurocontrol. (21 May 2010). European measures to minimise disruption caused by volcanic ash Retrieved April 6, 2012, from http://www.eurocontrol.int/press-releases/european-measures-minimise-disruption-caused-volcanic-ash

Forrester, J. W. (1961). *Industrial Dynamics*. Waltham, MA: Pegasus Communications.

Forrester, J. W., & Senge, P. M. (1980). Tests for Building Confidence in System Dynamics Models. In A. A. Legasto, J. W. Forrester & J. M. Lyneis (Eds.), *System Dynamics* (pp. 209-228). New York: North-Holland.

Gass, S. I. (1983). Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. *Operations Research, 31*(4), 603-631.

Gass, S. I. (1993). Model accreditation: A rationale and process for determining a numerical rating. *European Journal of Operational Research, 66*(2), 250-258. doi: 10.1016/0377-2217(93)90316-f

Gass, S. I., & Joel, L. S. (1981). Concepts of model confidence. *Computers & Operations Research, 8*(4), 341-346. doi: 10.1016/0305-0548(81)90019-8

Gass, S. I., & Thompson, B. W. (1980). Guidelines for Model Evaluation: An Abridged Version of the U.S. General Accounting Office Exposure Draft. *Operations Research, 28*(2), 431-439.

Grogran, P. T., Lee, C., & de Weck, O. L. (2011). *Comparative Usability Study of Two Space Logistics Analysis Tools*. Engineering Systems Division. Massachusetts Institute of Technology.

Guffanti, M., Casadevall, T. J., & Budding, K. (2010). *Encounters of Aircraft with Volcanic Ash Clouds: A Compilation of Known Incidents, 1953-2009*. Reston, VA: U.S. Geological Survey.

Gulwadi, S. (August, 8 2011). *Personal interview*. [Staff Engineer].

Gulwadi, S., & Hruby, E. (2011). Process Document IT4 MBC Development.

Harmon, S. Y., & Youngblood, S. M. (2005). A Proposed Model for Simulation Validation Process Maturity. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 2*(4), 179-190. doi: 10.1177/154851290500200402

Henderson, G. H., Marthaler, M. J., Braun, J. J., Thomas, V., Pan, G., & McKinley, T. L. (2004). U.S. Patent No. 6,810,725. Washington, DC: U.S. Patent and Trademark Office.

Highland, H. J. (1973). A taxonomy of models. *SIGSIM Simul. Dig., 4*(2), 10-17. doi: 10.1145/1102659.1102660

Hosagrahara, A., & Smith, P. (2005). *Measuring Productivity and Quality in Model-Based Design.* Paper presented at the SAE World Congress, Detroit, MI.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: John Wiley & Sons.

ICAO. (2007). Manual on Volcanic Ash, Radioactive Material and Toxic Chemical Clouds. Doc 9691 AN/954.

INCOSE. (2011). Systems Engineering Handbook: A guide for system life cycle processes and activities C. Haskins, K. Forsberg, M. Krueger, D. Walden & R. D. Hamelin (Eds.), *SE Handbook Working Group* Retrieved from http://www.incose.org

Jain, S., & McLean, C. R. (2009, 13-16 Dec. 2009). *Recommended practices for homeland security modeling and simulation.* Paper presented at the Simulation Conference (WSC), Proceedings of the 2009 Winter.

Johnson, C., & Jeunemaitre, A. (2011). Risk and the Role of Scientific Input for Contingency Planning: A Response to the April 2010 Eyjafjallajökull Volcano Eruption. In A. Alemanno (Ed.), Governing Disasters: The Challenges of Emergency Risk Regulation (pp. 49-62): Edward Elgar Publishing.

Kleb, B. (2007). Toward Scientific Numerical ModelingComputational Uncertainty in Military Vehicle Design: RTO-MP-AVT-147, Paper 17 (pp. 17-11 - 17-14). Neuilly-sur-Seine, France: RTO (http://www.rta.nato.int/).

Kleindorfer, G. B., & Geneshan, R. (1993). *The philosophy of science and validation in simulation.* Paper presented at the Proceedings of the 25th conference on Winter simulation, Los Angeles, California, United States.

Lahsen, M. (2005). Seductive Simulations? Uncertainty Distribution around Climate Models. *Social Studies of Science, 35*(6), 895-922.

Leveson, N., & Cutcher-Gershenfeld, J. (2004). What System Safety Engineering Can Learn from the Columbia Accident

Lewandowski, A. (1981). Issues in Model Validation. WP-81-32. Laxenburg, Austria, International Institute for Applied Systems Analysis.

Lindblom, C. E., & Cohen, D. K. (1979). *Usable Knowledge: Social Science and Social Problem Solving.* New Haven, CT: Yale University Press.

LinkedIn Corporation. (2012). LinkedIn Retrieved April 06, 2012, from http://www.linkedin.com

Lujan, J. M., Galindo, J., Vera, F., & Climent, H. (2007). Characterization and dynamic response of an exhaust gas recirculation venturi for internal combustion engines. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 221*(4), 497-509.

Lujan, J. M., Payri, F., Vera, F., & Guardiola, C. (2001). Modelling, Effect and Behaviour of the EGR Venturi in a Heavy-Duty Diesel Engine. [10.4271/2001-01-3227]. doi: 10.4271/2001-01-3227

MacKenzie, D. (1990). *Inventing Accuracy: A Historical Sociology of Nuclear Missle Guidance.* Cambridge, MA: MIT Press.

Macrae, D. (2011). Which Risk and who decides when there are so many players? In A. Alemanno (Ed.), Governing Disasters: The Challenges of Emergency Risk Regulation (pp. 11-24): Edward Elgar Publishing.

Masys, A. J., Roza, M., Giannoulis, C., & Jacquart, R. (2008). Verification, validation, and accreditation (VV&A): The GM-VV contribution and roadmap. NATO.

NASA. (2008). *Standard for Models and Simulations.* (NASA-STD-7009).

National Research Council (U.S.). Committee on Models in the Regulatory Decision Process. (2007). *Models in environmental regulatory decision making.* Washington, D.C.: National Academies Press.

Naylor, T. H., Finger, J. M., McKenney, J. L., Schrank, W. E., & Holt, C. C. (1967). Verification of Computer Simulation Models. *Management Science, 14*(2), B92-B106.

Oberkampf, W. L., Pilch, M., & Trucano, T. (2007). Predictive Capability Maturity Model for Computational Modeling and Simulation. Albuquerque, New Mexico, Sandia National Laboratories.

Oren, T. I. (1977). Simulation - as it has been, is, and should be. *SIMULATION, 29*(5), 182-183. doi: 10.1177/003754977702900531

Oren, T. I. (1981). Concepts and criteria to assess acceptability of simulation studies: a frame of reference. *Commun. ACM, 24*(4), 180-189. doi: 10.1145/358598.358605

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science, 263*(5147), 641-646.

Pace, D. (2004). Modeling and Simulation Verification and Validation Challenges. *Johns Hopkins APL Technical Digest, 25*(2), 163-172.

Payri, F., Benajes, J., Galindo, J., & Serrano, J. R. (2002). Modelling of turbocharged diesel engines in transient operation. Part 2: Wave action models for calculating the transient operation in a high speed direct injection engine. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 216*(6), 479-493.

Peloton Systems LLC. (2010). XPULT Catapult Retrieved April 06, 2012, from http://www.xpult.com/

Ragona, M., Hansstein, F., & Mazzocchi, M. (2011). The Financial Impact of the Volcanic Ash Crisis on the European Airline Industry. In A. Alemanno (Ed.), Governing Disasters: The Challenges to Emergency Risk Regulation (pp. 25-46): Edward Elgar Publishing.

Refsgaard, J. C., & Henriksen, H. J. r. (2004). Modelling guidelines--terminology and guiding principles. *Advances in Water Resources, 27*(1), 71-82. doi: 10.1016/j.advwatres.2003.08.006

Rice, J. (2007). *Mathematical Statistics and Data Analysis* (Third Edition ed.). Belmont, CA: The Thomson Corporation.

Romero, V. (2007). *Validated Model? Not So Fast. The Need for Model "Conditioning" as an Essential Addendum to Model Validation.* Paper presented at the AIAA Non-Deterministic Approaches Conference, Honolulu, HI.

Ross, J. (August 10, 2011, August 10, 2011). *Personal Interview*. [Model Based Systems Development Leader].

Sargent, R. G. (1987). *An overview of verification and validation of simulation models.* Paper presented at the 1987 Winter Simulation Conference Proceedings, 14-16 Dec. 1987, San Diego, CA, USA.

Sargent, R. G. (2001). *Verification and validation: some approaches and paradigms for verifying and validating simulation models.* Paper presented at the Proceedings of the 33nd conference on Winter simulation, Arlington, Virginia.

Sargent, R. G. (2005). *Verification and validation of simulation models.* Paper presented at the Proceedings of the 2005 Winter Simulation Conference, 4-7 Dec. 2005, Piscataway, NJ, USA.

Savoie, T. (2008). Catapult Simulation Interface. Cambridge, MA: Robust Design Group Department of Mechanical Engineering MIT, System Engineering Division Draper Laboratory.

Savoie, T. (2010). *Human Detection of Computer Simulation Mistakes in Engineering Experiments.* Doctor of Philosophy in Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Savoie, T., & Frey, D. (2012). Detecting mistakes in engineering models: the effects of experimental design. *Research in Engineering Design, 23*(2), 155-175. doi: 10.1007/s00163-011-0120-y

Sclove, R. (2010). Science and technology innovation program: Reinventing technology assessment *Using Citizen Participation, Collaboration and Expert Analysis to inform and improve decision-making on issues involving science and technology*

SCS. (1979). Terminology for model credibility. *SIMULATION, 32*(3), 103-104. doi: 10.1177/003754977903200304

Smith, P. F., Prabhu, S. M., & Friedman, J. (2007). *Best Practices for Establishing a Model-Based Design Culture.* Paper presented at the SAE World Congress & Exhibition, Detroit, MI.

Software Engineering Institute. (2012). Capability Maturity Model Integration(CMMI) Retrieved April 05, 2012, from http://www.sei.cmu.edu/cmmi/

Spinutech Inc. (2012). Web Design, Web Development & Web Strategy Company in Cedar Falls & Des Moines Iowa Retrieved April 06, 2012, from http://www.spinutech.com/

Steenkamp, J.-B. E. M. (1990). Conceptual model of the quality perception process. *Journal of Business Research, 21*(4), 309-333. doi: 10.1016/0148-2963(90)90019-a

Stohl, A., Prata, A. J., Eckhardt, S., Clarisse, L., Durant, A., Henne, S., . . . Weinzierl, B. (2011). Determination of time- and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: the 2010 Eyjafjallajökull eruption. *Atmospheric Chemistry and Physics, 11*(9), 4333-4351.

Thomas, D., Joiner, A., Lin, W., Lowry, M., & Pressburger, T. (2010). *The unique aspects of simulation verification and validation.* Paper presented at the 2010 IEEE Aerospace Conference, March 6, 2010 - March 13, 2010, Big Sky, MT, United states.

Ulfarsson, G., & Unger, E. (2011). Impacts and Responses of Icelandic Aviation to the 2010 Eyjafjallajökull Volcanic Eruption. *Transportation Research Record: Journal of the Transportation Research Board, 2214*(-1), 144-151. doi: 10.3141/2214-18

Utting, M., Pretschner, A., & Legeard, B. (2006). A Taxonomy of Model-Based Testing. 04/2006. Hamilton, New Zealand, Department of Computer Science, The University of Waikato.

van der Sluijs, J. P., Risbey, J. S., Kloprogge, P., Ravetz, J. R., Funtowica, S. O., Quintana, S. C., . . . Huijs, S. W. F. (2003). RIVM/MNP guidance for uncertainty assessment and communication. Copernicus Institute for Sustainable Development and Innovation.

Voas, J. (1998). The Software Quality Certification Triangle. *CrossTalk - The Journal of Defense Software Engineering, 11*(11), 12-14.

Weiss, C. H., & Bucuvalas, M. J. (1977). The Challenge of Social Research in Decision Making. In C. H. Weiss (Ed.), *Using Social Research in Public Policy Making* (pp. 213-230). Lexington, MA: Lexington Books.

Whitner, R. B., & Balci, O. (1989). *Guidelines for selecting and using simulation model verification techniques.* Paper presented at the Proceedings of the 21st conference on Winter simulation, Washington, D.C., United States.

Woods, D. (2005). Creating Foresight: Lessons for Enhancing Resilience from *Columbia*. In W. Starbuck & M. Farjoun (Eds.), *Organization at the Limit: Lessons from the Columbia Disaster*: Blackwell Publishing.

This page intentionally left blank

8 APPENDICES

8.1 APPENDIX A: MODEL VALIDATION DATA

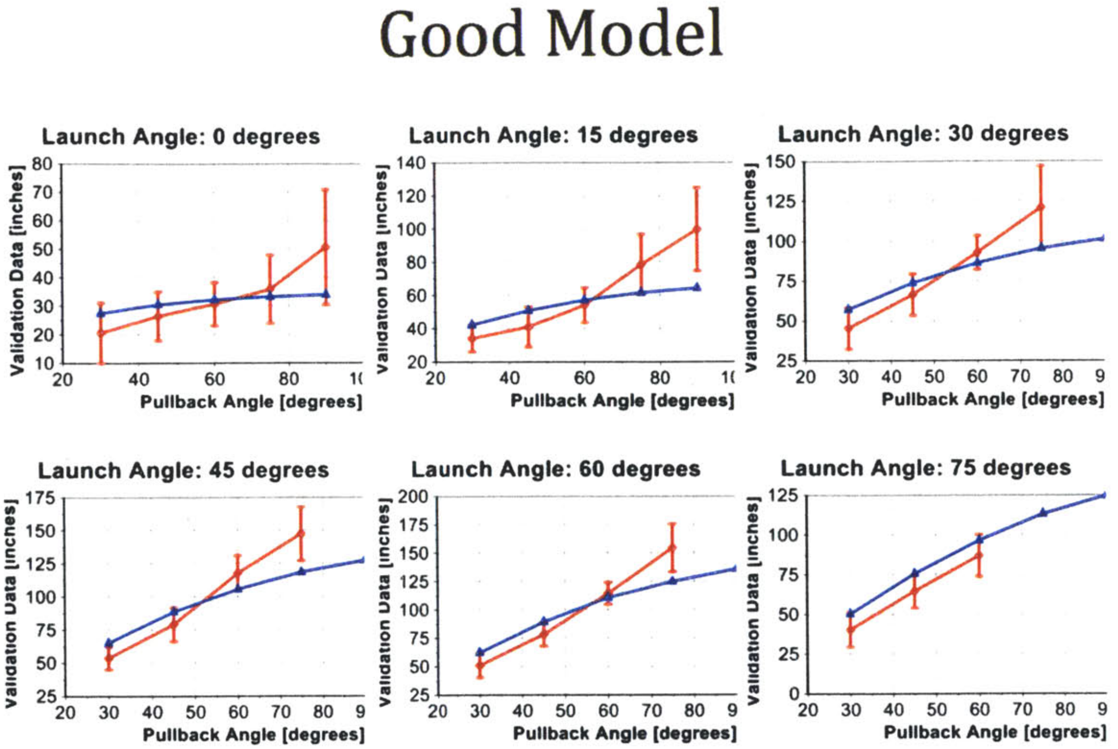


Figure 39: Validation Data for the Good Model case. This data used the smooth ball with two rubber bands. Each plot shows the distance traveled for varying launch and pullback angles. The Red curve with error bands shows the validation data the user received; the Blue curve shows the results the model would give for those same conditions. This data was not available to the user, but could be reproduced using the model.

Bad Model

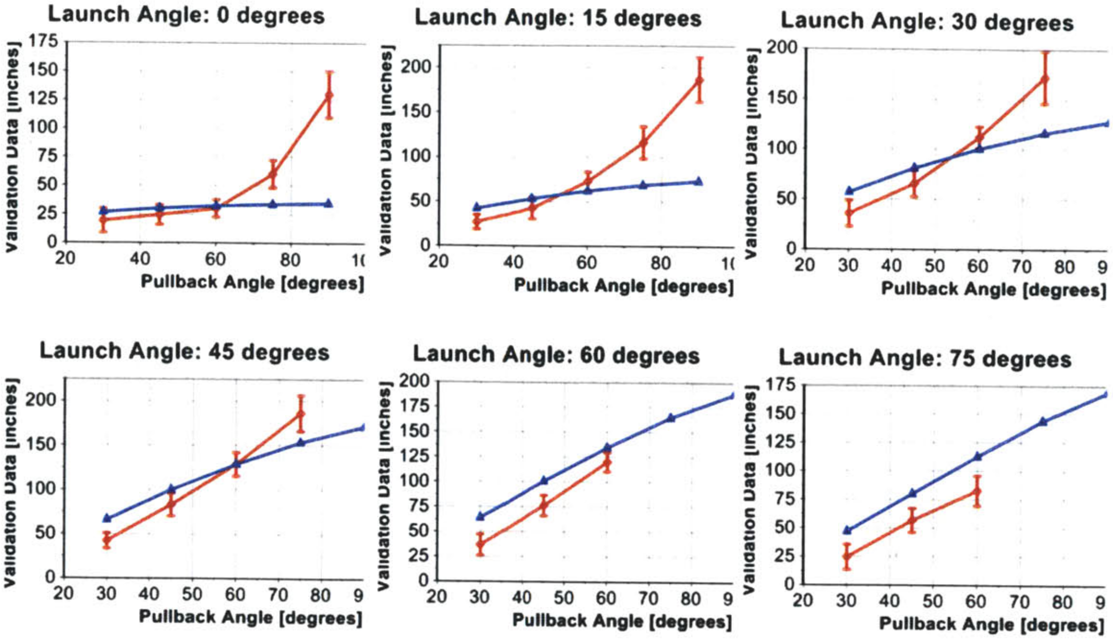
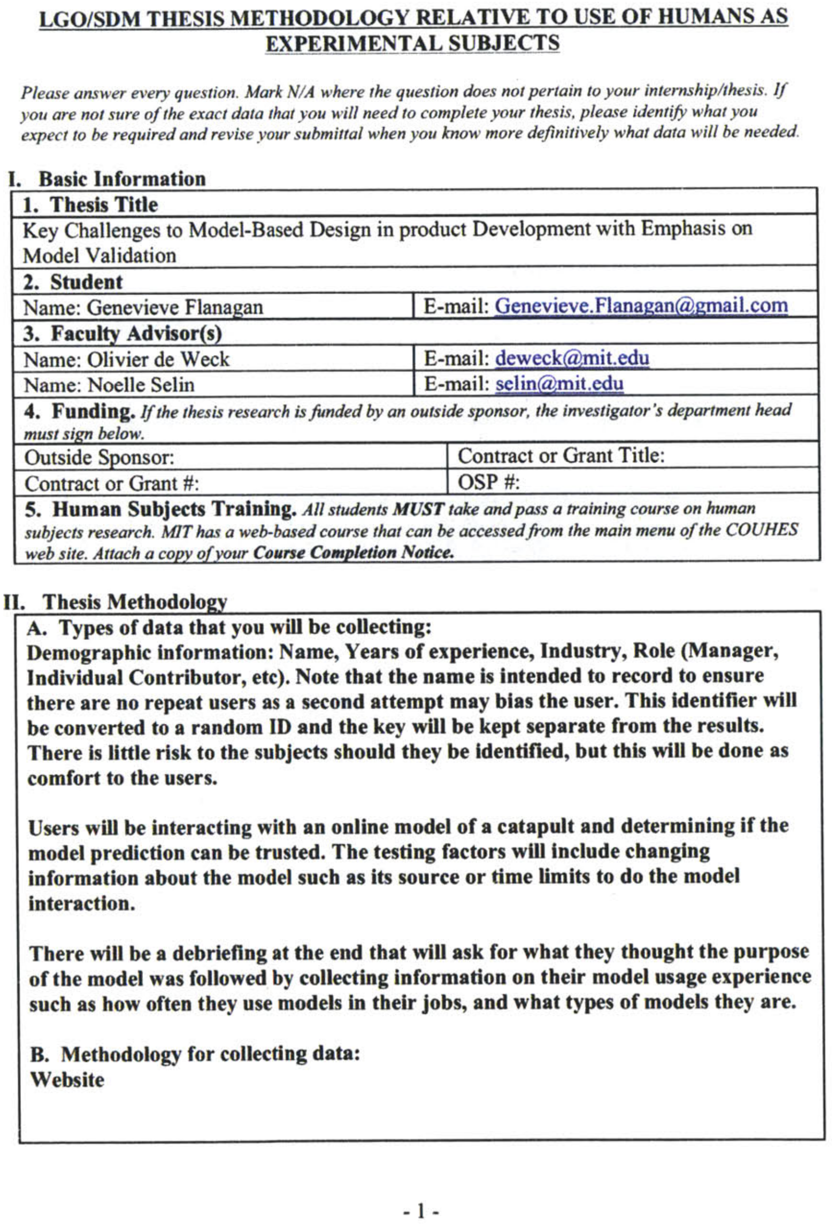


Figure 40: Validation Data for the Bad Model case. This data used the perforated ball with two rubber bands. Each plot shows the distance traveled for varying launch and pullback angles. The Red curve with error bards shows the validation data the user received; the Blue curve shows the results the model would give for those same conditions. This data was not available to the user, but could be reproduced using the model.

8.2 APPENDIX B: COUHES APPROVAL



LGO/SDM THESIS METHODOLOGY RELATIVE TO USE OF HUMANS AS EXPERIMENTAL SUBJECTS

Please answer every question. Mark N/A where the question does not pertain to your internship/thesis. If you are not sure of the exact data that you will need to complete your thesis, please identify what you expect to be required and revise your submittal when you know more definitively what data will be needed.

I. Basic Information

1. Thesis Title	
Key Challenges to Model-Based Design in product Development with Emphasis on Model Validation	
2. Student	
Name: Genevieve Flanagan	E-mail: Genevieve.Flanagan@gmail.com
3. Faculty Advisor(s)	
Name: Olivier de Weck	E-mail: deweck@mit.edu
Name: Noelle Selin	E-mail: selin@mit.edu
4. Funding. *If the thesis research is funded by an outside sponsor, the investigator's department head must sign below.*	
Outside Sponsor:	Contract or Grant Title:
Contract or Grant #:	OSP #:
5. Human Subjects Training. *All students **MUST** take and pass a training course on human subjects research. MIT has a web-based course that can be accessed from the main menu of the COUHES web site. Attach a copy of your **Course Completion Notice.***	

II. Thesis Methodology

A. Types of data that you will be collecting:
Demographic information: Name, Years of experience, Industry, Role (Manager, Individual Contributor, etc). Note that the name is intended to record to ensure there are no repeat users as a second attempt may bias the user. This identifier will be converted to a random ID and the key will be kept separate from the results. There is little risk to the subjects should they be identified, but this will be done as comfort to the users.

Users will be interacting with an online model of a catapult and determining if the model prediction can be trusted. The testing factors will include changing information about the model such as its source or time limits to do the model interaction.

There will be a debriefing at the end that will ask for what they thought the purpose of the model was followed by collecting information on their model usage experience such as how often they use models in their jobs, and what types of models they are.

B. Methodology for collecting data:
Website

C. If you plan to interview/survey people, how will those people be identified?
It is preferred to use professionals in a technical field as the subjects. I plan to use the following recruiting strategy:
e-mails to John Deere employees via our Systems Engineering and Model-based design communities of practice
e-mails to local INCOSE chapter seeking assistance
e-mails to suppliers we interact with
e-mails to MIT SDM groups
promotion in advisors' classes
potential to post on linked in groups

D. If you plan to use interviews and/or surveys, outline the types of questions you will include:
Demographic Questions
Interaction with a model with final questions based on their experience
Understanding the users' experiences with model usage in the technical field
Understanding the users' experiences with decision making using models

E. Location of the research:
I am located in Iowa, research is online, therefore could be global

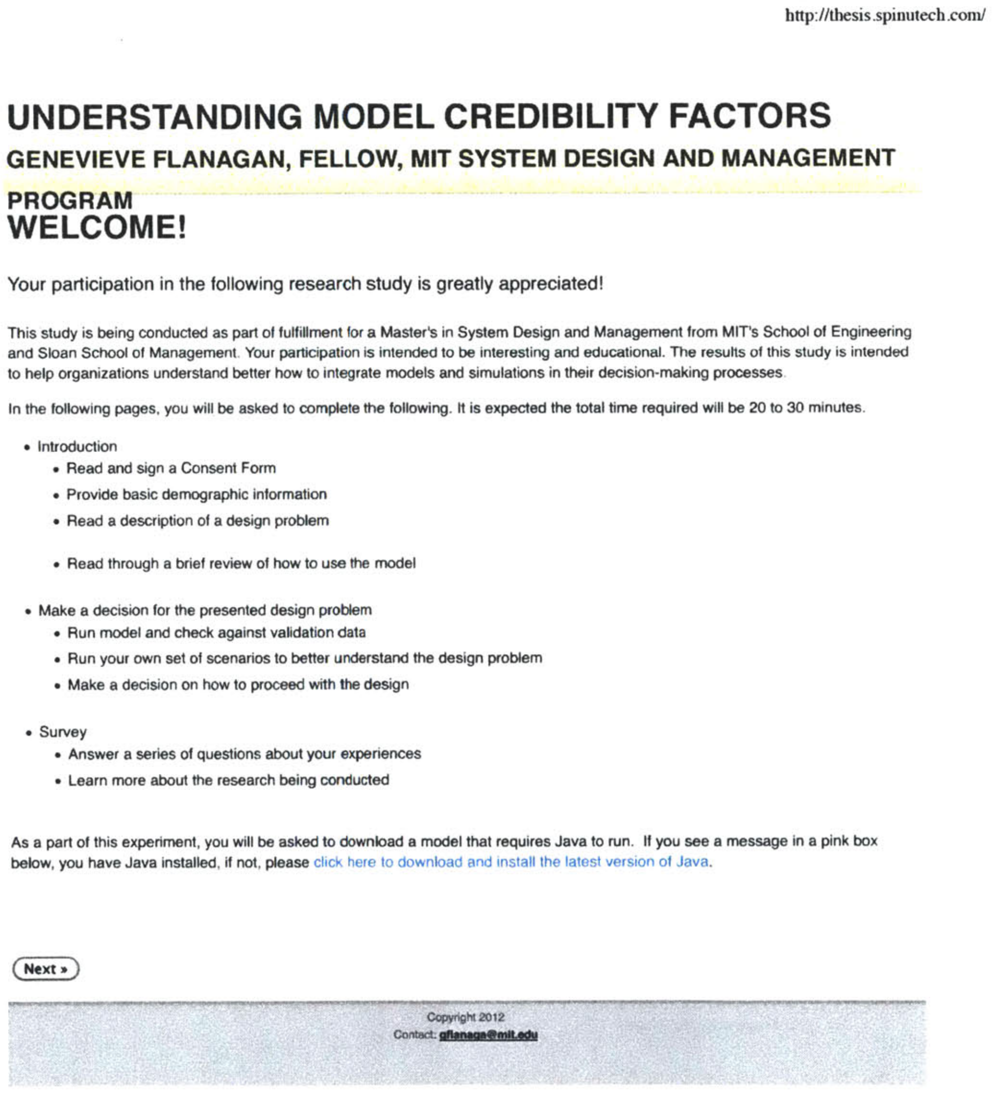
G. Procedures to ensure confidentiality:
Data being collected has minimal risk to the users. However, name information will be converted to a random identifier and the key will be kept separately from the rest of the data to ensure confidentiality. Name information is required only to ensure that there are no repeat runs of the test.

Signature____________________________ **Date _2/28/2012**__________
Director, SDM Fellows Program

8.3 APPENDIX C: SCREENSHOTS FROM THE EXPERIMENT

8.3.1 COMMON SCREENS TO ALL USERS

8.3.1.1 Introduction



http://thesis.spinutech.com/

UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

WELCOME!

Your participation in the following research study is greatly appreciated!

This study is being conducted as part of fulfillment for a Master's in System Design and Management from MIT's School of Engineering and Sloan School of Management. Your participation is intended to be interesting and educational. The results of this study is intended to help organizations understand better how to integrate models and simulations in their decision-making processes.

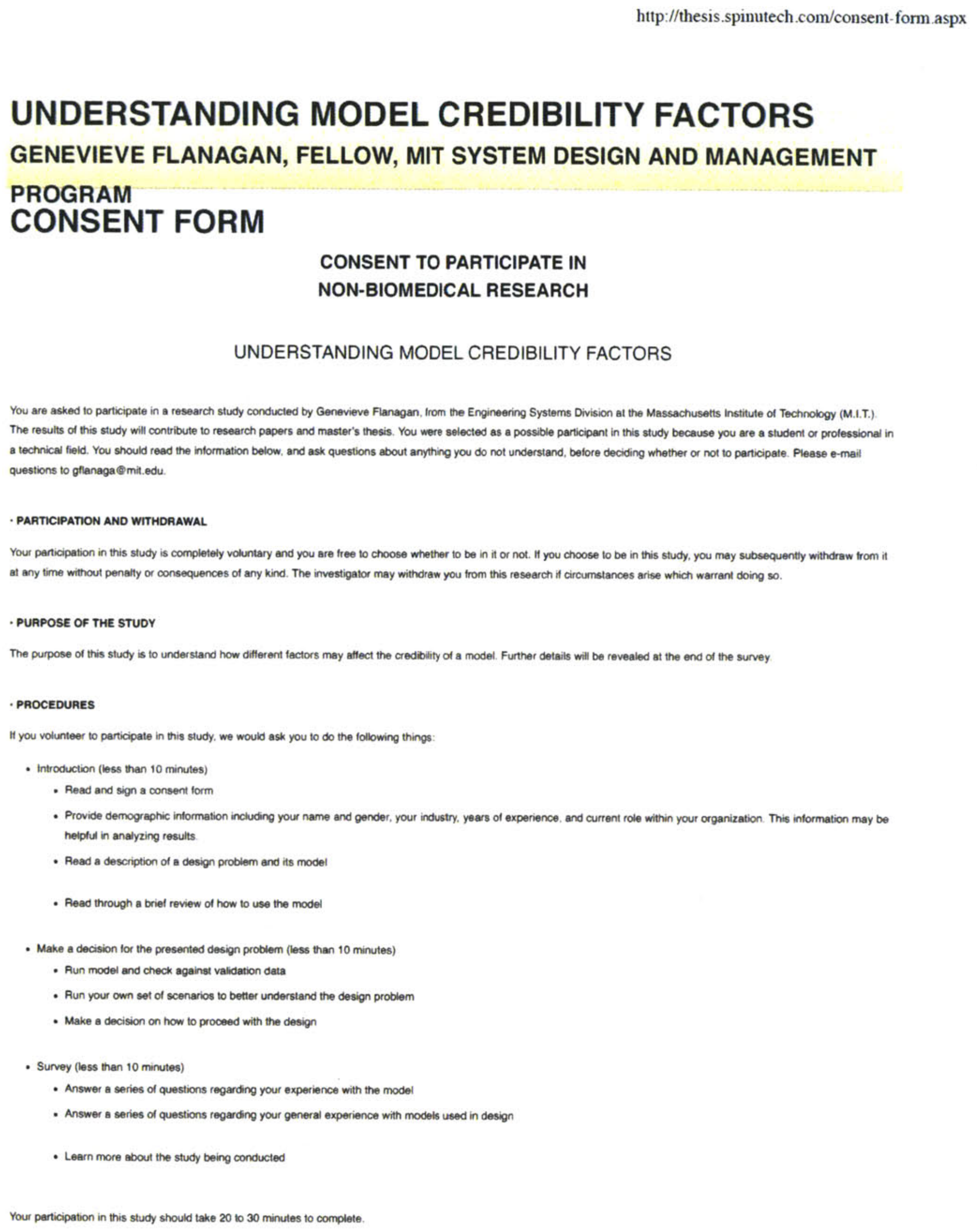
In the following pages, you will be asked to complete the following. It is expected the total time required will be 20 to 30 minutes.

- Introduction
  - Read and sign a Consent Form
  - Provide basic demographic information
  - Read a description of a design problem
  - Read through a brief review of how to use the model
- Make a decision for the presented design problem
  - Run model and check against validation data
  - Run your own set of scenarios to better understand the design problem
  - Make a decision on how to proceed with the design
- Survey
  - Answer a series of questions about your experiences
  - Learn more about the research being conducted

As a part of this experiment, you will be asked to download a model that requires Java to run. If you see a message in a pink box below, you have Java installed, if not, please click here to download and install the latest version of Java.

Copyright 2012
Contact: gflanaga@mit.edu

8.3.1.2 Consent Form



http://thesis.spinutech.com/consent-form.aspx

UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

CONSENT FORM

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

UNDERSTANDING MODEL CREDIBILITY FACTORS

You are asked to participate in a research study conducted by Genevieve Flanagan, from the Engineering Systems Division at the Massachusetts Institute of Technology (M.I.T.). The results of this study will contribute to research papers and master's thesis. You were selected as a possible participant in this study because you are a student or professional in a technical field. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate. Please e-mail questions to gflanaga@mit.edu.

· PARTICIPATION AND WITHDRAWAL

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

· PURPOSE OF THE STUDY

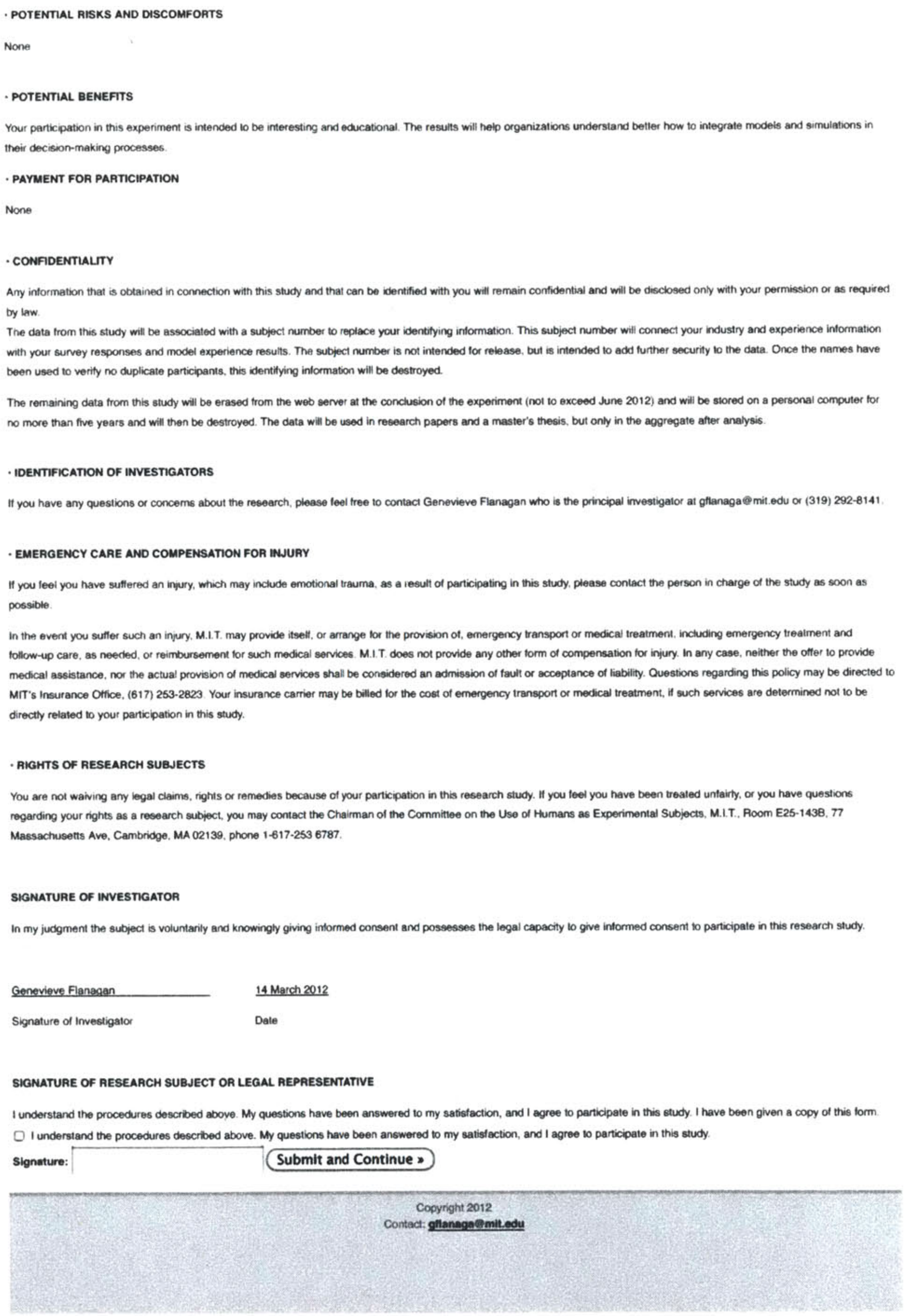
The purpose of this study is to understand how different factors may affect the credibility of a model. Further details will be revealed at the end of the survey

· PROCEDURES

If you volunteer to participate in this study, we would ask you to do the following things:

- Introduction (less than 10 minutes)
    - Read and sign a consent form
    - Provide demographic information including your name and gender, your industry, years of experience, and current role within your organization. This information may be helpful in analyzing results.
    - Read a description of a design problem and its model
    - Read through a brief review of how to use the model
- Make a decision for the presented design problem (less than 10 minutes)
    - Run model and check against validation data
    - Run your own set of scenarios to better understand the design problem
    - Make a decision on how to proceed with the design
- Survey (less than 10 minutes)
    - Answer a series of questions regarding your experience with the model
    - Answer a series of questions regarding your general experience with models used in design
    - Learn more about the study being conducted

Your participation in this study should take 20 to 30 minutes to complete.



· **POTENTIAL RISKS AND DISCOMFORTS**

None

· **POTENTIAL BENEFITS**

Your participation in this experiment is intended to be interesting and educational. The results will help organizations understand better how to integrate models and simulations in their decision-making processes.

· **PAYMENT FOR PARTICIPATION**

None

· **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.
The data from this study will be associated with a subject number to replace your identifying information. This subject number will connect your industry and experience information with your survey responses and model experience results. The subject number is not intended for release, but is intended to add further security to the data. Once the names have been used to verify no duplicate participants, this identifying information will be destroyed.

The remaining data from this study will be erased from the web server at the conclusion of the experiment (not to exceed June 2012) and will be stored on a personal computer for no more than five years and will then be destroyed. The data will be used in research papers and a master's thesis, but only in the aggregate after analysis.

· **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concems about the research, please feel free to contact Genevieve Flanagan who is the principal investigator at gflanaga@mit.edu or (319) 292-8141.

· **EMERGENCY CARE AND COMPENSATION FOR INJURY**

If you feel you have suffered an injury, which may include emotional trauma, as a result of participating in this study, please contact the person in charge of the study as soon as possible.

In the event you suffer such an injury, M.I.T. may provide itself, or arrange for the provision of, emergency transport or medical treatment, including emergency treatment and follow-up care, as needed, or reimbursement for such medical services. M.I.T. does not provide any other form of compensation for injury. In any case, neither the offer to provide medical assistance, nor the actual provision of medical services shall be considered an admission of fault or acceptance of liability. Questions regarding this policy may be directed to MIT's Insurance Office, (617) 253-2823. Your insurance carrier may be billed for the cost of emergency transport or medical treatment, if such services are determined not to be directly related to your participation in this study.

· **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

Genevieve Flanagan ______________ 14 March 2012

Signature of Investigator Date

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

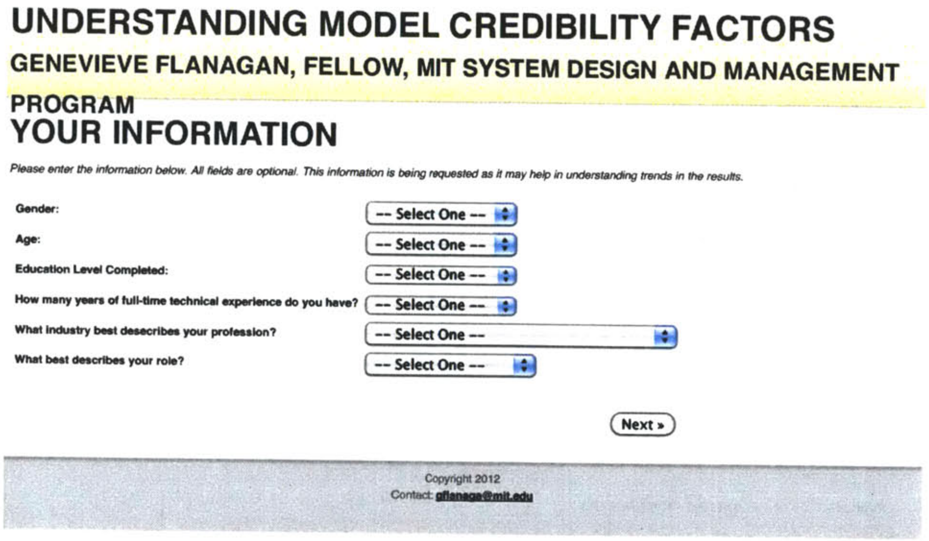
□ I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study.

Signature:

Copyright 2012
Contact: gflanaga@mit.edu

8.3.1.3 Demographics Questionnaire

http://thesis.spinutech.com/demographic-information.aspx



UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

YOUR INFORMATION

Please enter the information below. All fields are optional. This information is being requested as it may help in understanding trends in the results.

Gender: -- Select One --

Age: -- Select One --

Education Level Completed: -- Select One --

How many years of full-time technical experience do you have? -- Select One --

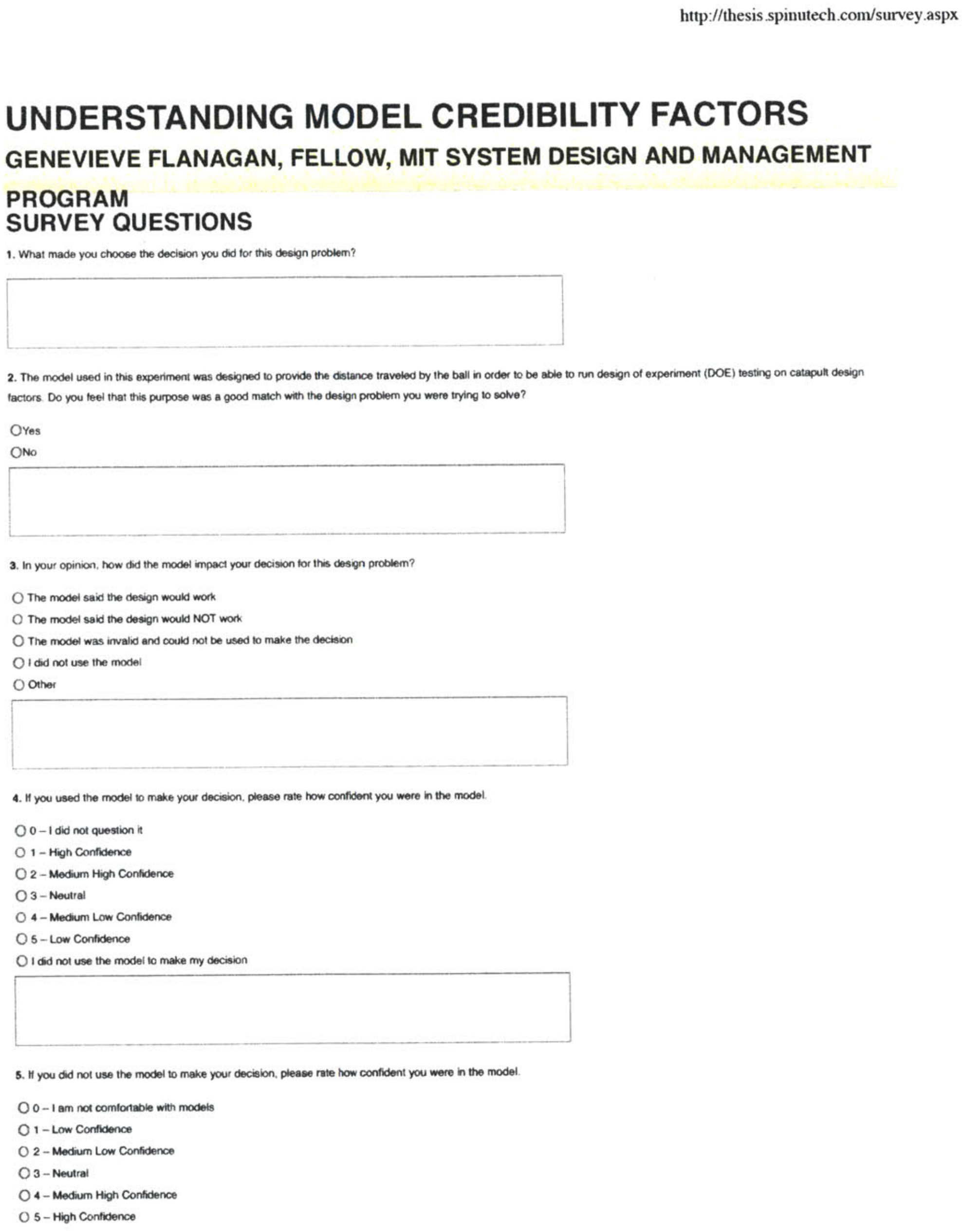
What industry best desecribes your profession? -- Select One --

What best describes your role? -- Select One --

Copyright 2012

Contact: gflanaga@mit.edu

8.3.1.4 Survey Questionnaire



http://thesis.spinutech.com/survey.aspx

UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

SURVEY QUESTIONS

1. What made you choose the decision you did for this design problem?

2. The model used in this experiment was designed to provide the distance traveled by the ball in order to be able to run design of experiment (DOE) testing on catapult design factors. Do you feel that this purpose was a good match with the design problem you were trying to solve?

○Yes
○No

3. In your opinion, how did the model impact your decision for this design problem?

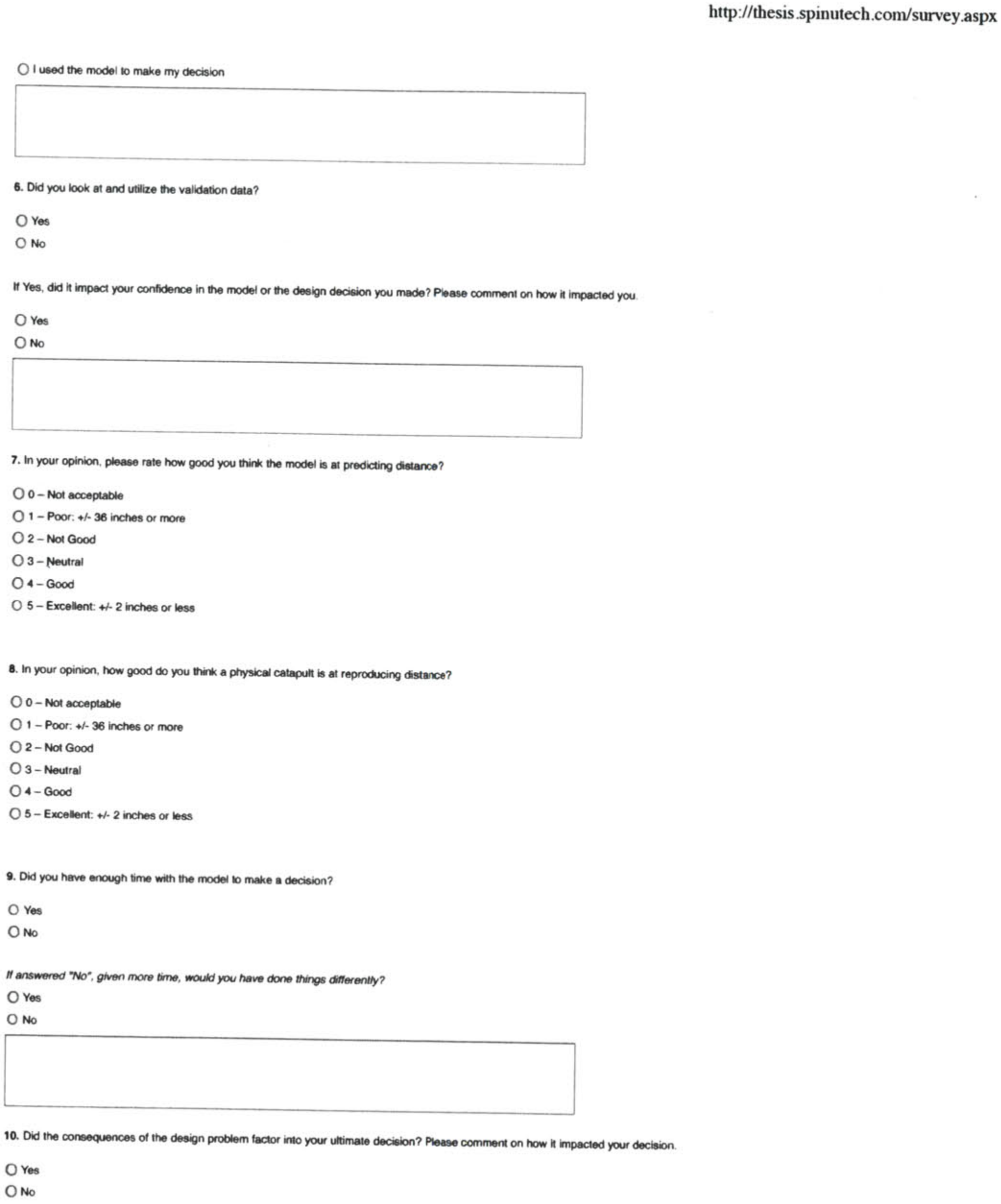
○ The model said the design would work
○ The model said the design would NOT work
○ The model was invalid and could not be used to make the decision
○ I did not use the model
○ Other

4. If you used the model to make your decision, please rate how confident you were in the model.

○ 0 – I did not question it
○ 1 – High Confidence
○ 2 – Medium High Confidence
○ 3 – Neutral
○ 4 – Medium Low Confidence
○ 5 – Low Confidence
○ I did not use the model to make my decision

5. If you did not use the model to make your decision, please rate how confident you were in the model.

○ 0 – I am not comfortable with models
○ 1 – Low Confidence
○ 2 – Medium Low Confidence
○ 3 – Neutral
○ 4 – Medium High Confidence
○ 5 – High Confidence



http://thesis.spinutech.com/survey.aspx

○ I used the model to make my decision

6. Did you look at and utilize the validation data?

○ Yes
○ No

If Yes, did it impact your confidence in the model or the design decision you made? Please comment on how it impacted you.

○ Yes
○ No

7. In your opinion, please rate how good you think the model is at predicting distance?

○ 0 – Not acceptable
○ 1 – Poor: +/- 36 inches or more
○ 2 – Not Good
○ 3 – Neutral
○ 4 – Good
○ 5 – Excellent: +/- 2 inches or less

8. In your opinion, how good do you think a physical catapult is at reproducing distance?

○ 0 – Not acceptable
○ 1 – Poor: +/- 36 inches or more
○ 2 – Not Good
○ 3 – Neutral
○ 4 – Good
○ 5 – Excellent: +/- 2 inches or less

9. Did you have enough time with the model to make a decision?

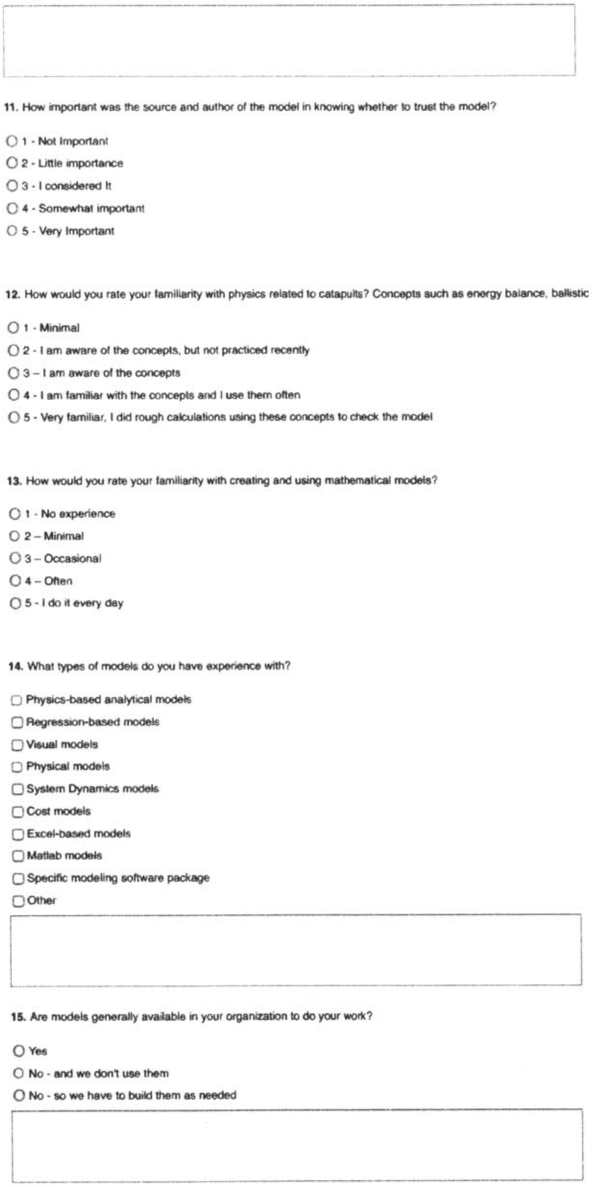
○ Yes
○ No

If answered "No", given more time, would you have done things differently?

○ Yes
○ No

10. Did the consequences of the design problem factor into your ultimate decision? Please comment on how it impacted your decision.

○ Yes
○ No



11. How important was the source and author of the model in knowing whether to trust the model?

○ 1 - Not Important
○ 2 - Little importance
○ 3 - I considered It
○ 4 - Somewhat important
○ 5 - Very Important

12. How would you rate your familiarity with physics related to catapults? Concepts such as energy balance, ballistic effects and drag.

○ 1 - Minimal
○ 2 - I am aware of the concepts, but not practiced recently
○ 3 – I am aware of the concepts
○ 4 - I am familiar with the concepts and I use them often
○ 5 - Very familiar, I did rough calculations using these concepts to check the model

13. How would you rate your familiarity with creating and using mathematical models?

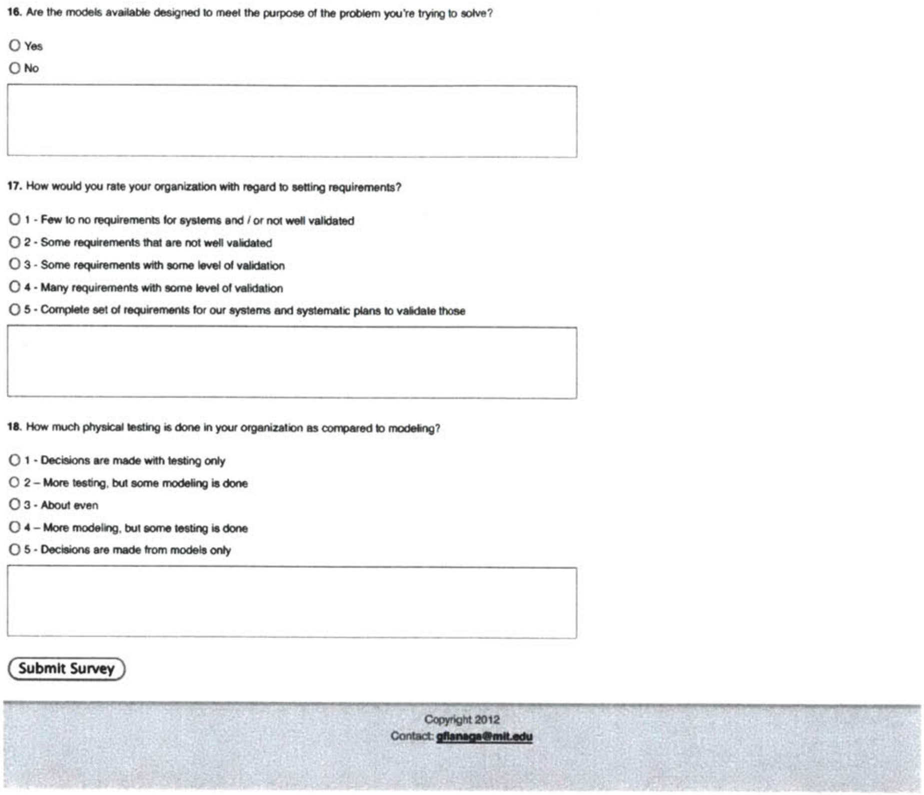
○ 1 - No experience
○ 2 – Minimal
○ 3 – Occasional
○ 4 – Often
○ 5 - I do it every day

14. What types of models do you have experience with?

□ Physics-based analytical models
□ Regression-based models
□ Visual models
□ Physical models
□ System Dynamics models
□ Cost models
□ Excel-based models
□ Matlab models
□ Specific modeling software package
□ Other

15. Are models generally available in your organization to do your work?

○ Yes
○ No - and we don't use them
○ No - so we have to build them as needed



16. Are the models available designed to meet the purpose of the problem you're trying to solve?

○ Yes

○ No

17. How would you rate your organization with regard to setting requirements?

○ 1 - Few to no requirements for systems and / or not well validated

○ 2 - Some requirements that are not well validated

○ 3 - Some requirements with some level of validation

○ 4 - Many requirements with some level of validation

○ 5 - Complete set of requirements for our systems and systematic plans to validate those

18. How much physical testing is done in your organization as compared to modeling?

○ 1 - Decisions are made with testing only

○ 2 – More testing, but some modeling is done

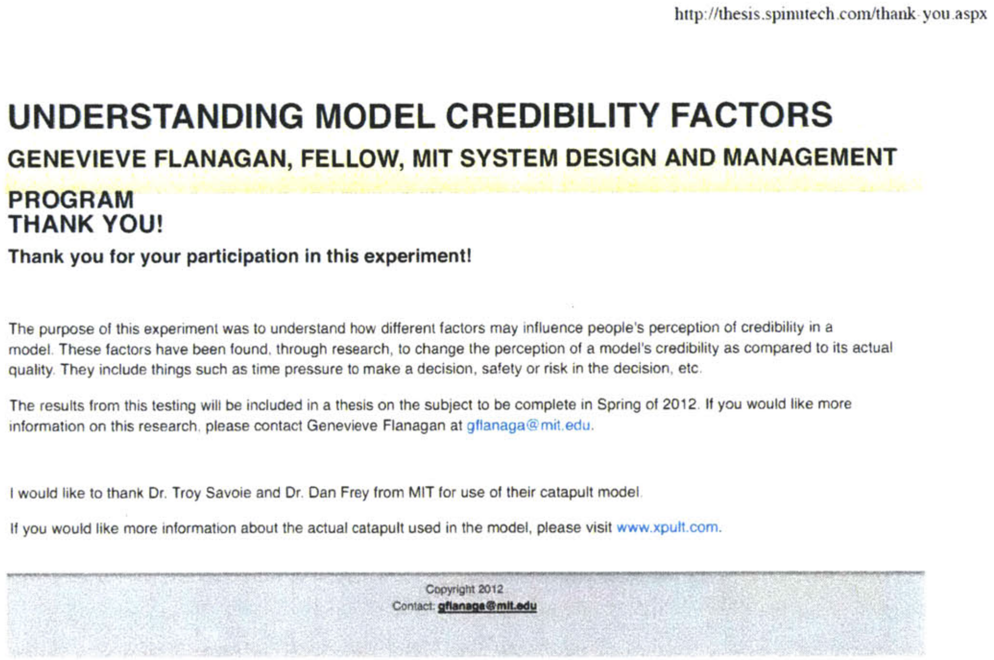
○ 3 - About even

○ 4 – More modeling, but some testing is done

○ 5 - Decisions are made from models only

Copyright 2012
Contact: gflanaga@mit.edu

8.3.1.5 Conclusion



http://thesis.spinutech.com/thank-you.aspx

UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

THANK YOU!

Thank you for your participation in this experiment!

The purpose of this experiment was to understand how different factors may influence people's perception of credibility in a model. These factors have been found, through research, to change the perception of a model's credibility as compared to its actual quality. They include things such as time pressure to make a decision, safety or risk in the decision, etc.

The results from this testing will be included in a thesis on the subject to be complete in Spring of 2012. If you would like more information on this research, please contact Genevieve Flanagan at gflanaga@mit.edu.

I would like to thank Dr. Troy Savoie and Dr. Dan Frey from MIT for use of their catapult model.

If you would like more information about the actual catapult used in the model, please visit www.xpult.com.

Copyright 2012
Contact: gflanaga@mit.edu

8.3.1.6 Detailed Help Page

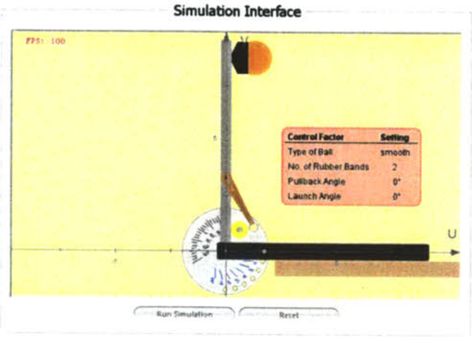
http://thesis.spinutech.com/model-help.aspx

UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

MODEL HELP

GUIDE TO USING THE CATAPULT MODEL



Simulation Interface

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	0°

DESIGN FACTORS:

Type of Ball:

- Smooth (Orange)
- Perforated (Blue with holes)

Number of Rubber Bands:

- 1
- 2

Launch Angle:

Angle at which the arm stops and releases the ball

- 0°
- 15°
- 30°
- 45°
- 60°
- 75°
- 90°

Pullback Angle:

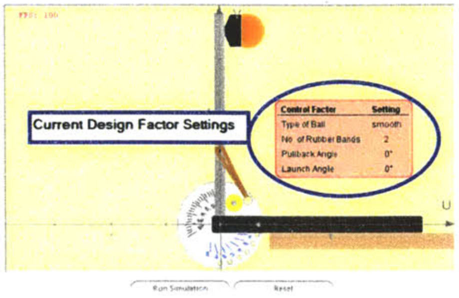
Angle at which the arm of the catapult is pulled back prior to launch

- Variable 21° - 120°

ADJUSTING DESIGN FACTORS:

The design factors used in the current set-up are shown in the red box in the upper right corner of the simulation as shown below. The instructions below describe how to change them for your experimentation.

Simulation Interface



Current Design Factor Settings

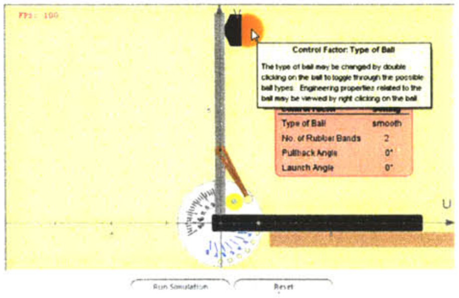
Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	0°

Changing the Ball:

Place your mouse cursor over the ball in the simulation and wait until you see the dialog box as shown: "Control Factor: Type of Ball."

At this point, double-click on the ball to switch between the balls

Simulation Interface



Control Factor: Type of Ball

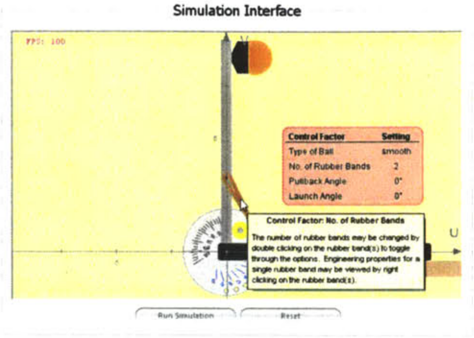
The type of ball may be changed by double clicking on the ball to toggle through the possible ball types. Engineering properties related to the ball may be viewed by right clicking on the ball

Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	0°

Changing the Number of Rubber Bands:

Place your mouse cursor over the rubber bands in the simulation and wait until you see the dialog box: "Control Factor: No. of Rubber Bands."

At this point, double-click on the rubber bands to change between 1 and 2.



Simulation Interface

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	0°

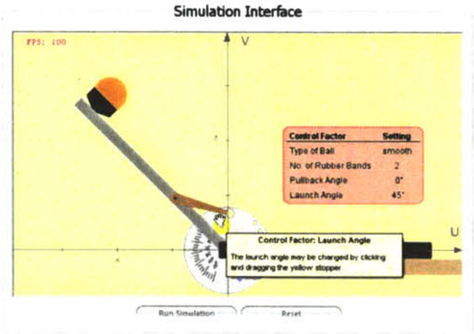
Control Factor: No. of Rubber Bands

The number of rubber bands may be changed by double clicking on the rubber band(s) to toggle through the options. Engineering properties for a single rubber band may be viewed by right clicking on the rubber band(s).

Changing the Launch Angle:

Place your mouse cursor over the yellow pin at the base of the catapult and wait until you see the dialog box: "Control Factor: Launch Angle."

At this point, drag the pin around the pivot of the catapult. The catapult arm will move as this pin changes value.



Simulation Interface

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	45°

Control Factor: Launch Angle

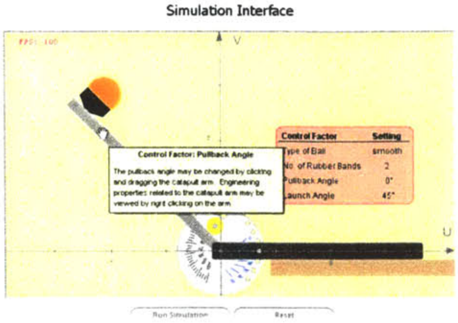
The launch angle may be changed by clicking and dragging the yellow stopper

Changing the Pullback Angle:

Place your mouse cursor anywhere over the arm of the catapult and wait until you see the dialog box: "Control Factor: Pullback Angle."

At this point, drag the arm back from its launch position to whatever pullback is desired. Note a green area will highlight showing the pullback region as compared to the launch angle.

Note: you must have at least a 21° pullback angle in order to run the simulation.

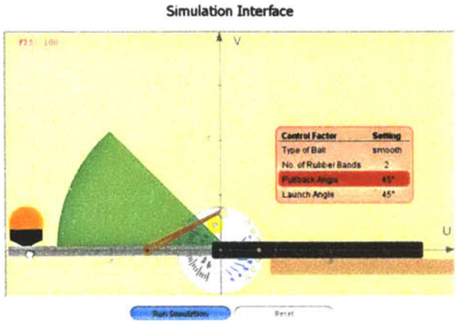


Simulation Interface

Control Factor: Pullback Angle

The pullback angle may be changed by clicking and dragging the catapult arm. Engineering properties related to the catapult arm may be viewed by right clicking on the arm.

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	45°



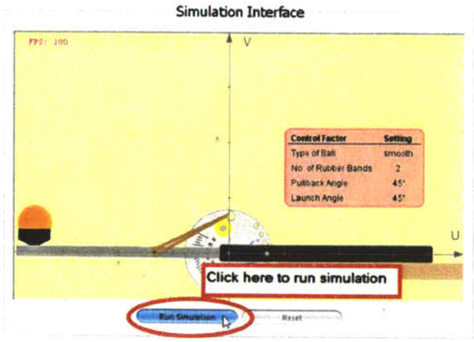
Simulation Interface

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	45°
Launch Angle	45°

RUNNING THE SIMULATION:

Once you have set-up your design factors, the "Run Simulation" button found at the bottom of the simulation will be highlighted in blue. If it is not, most likely you have not set-up an appropriate pullback angle. Click this button to run the simulation. You will see a countdown timer and then the catapult will launch.

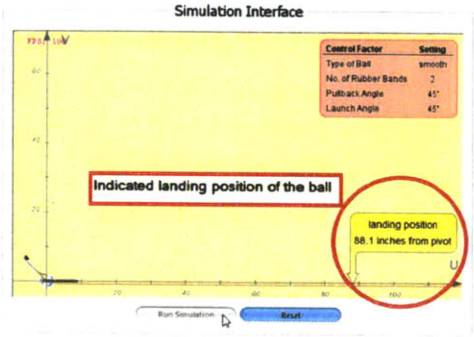
Upon completing the simulation, a box will appear at the landing location of the ball. This box will indicate the distance the ball traveled.



Simulation Interface

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	45°
Launch Angle	45°

Click here to run simulation



Simulation Interface

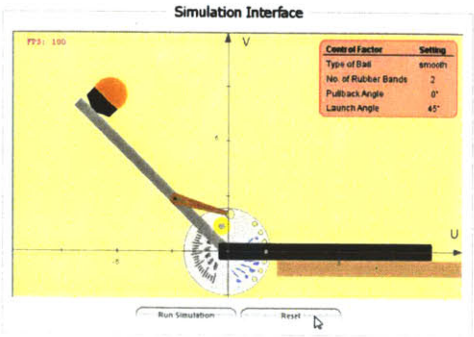
Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	45°
Launch Angle	45°

Indicated landing position of the ball

landing position
88.1 inches from pivot

RESETTING THE SIMULATION TO RUN AGAIN

Once the simulation is complete, the "Reset" button found at the bottom of the simulation will be highlighted in blue. Click on this button to return to the catapult to set-up a new experiment. The catapult will be in the same configuration as was previously run with the exception of the pullback angle.



Simulation Interface

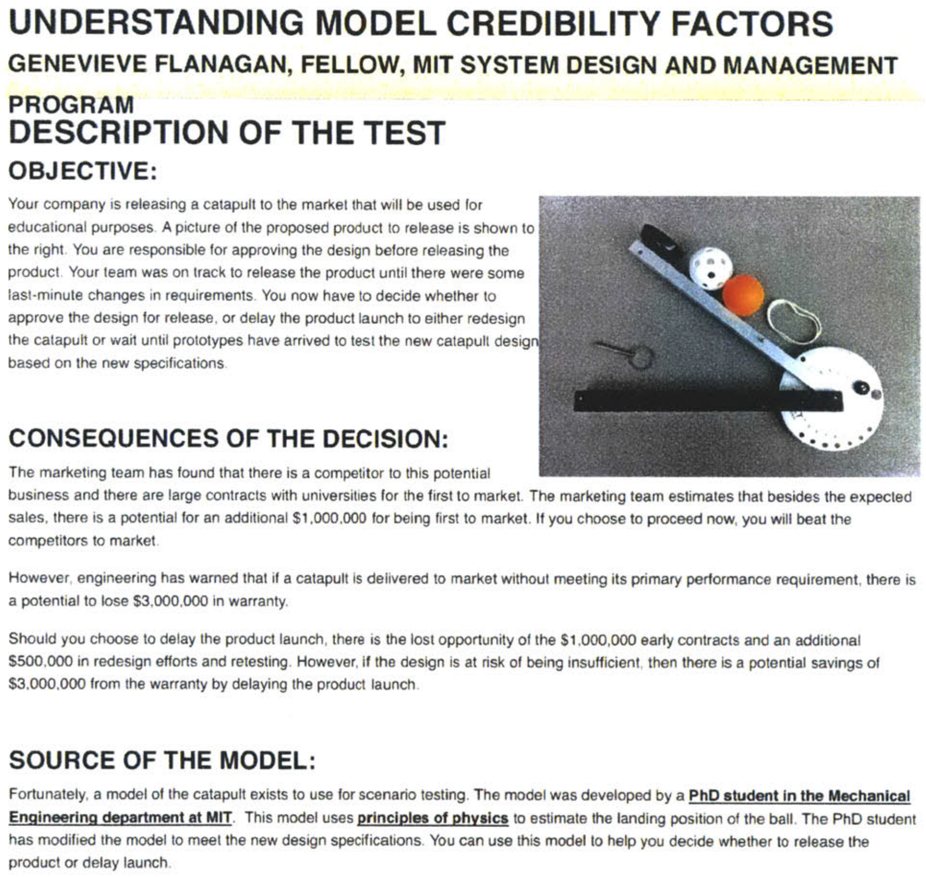
Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	45°

8.3.2 SCREENS FOR REFERENCE TEST CASE WITH GOOD MODEL

The following screenshots show those related to the reference test case with a good model. The other test cases have a similar look, but information on each page relative to the case was changed as what was described in section 5.2.

8.3.2.1 Experiment Overview Page

http://thesis.spinutech.com/description.aspx



UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

DESCRIPTION OF THE TEST

OBJECTIVE:

Your company is releasing a catapult to the market that will be used for educational purposes. A picture of the proposed product to release is shown to the right. You are responsible for approving the design before releasing the product. Your team was on track to release the product until there were some last-minute changes in requirements. You now have to decide whether to approve the design for release, or delay the product launch to either redesign the catapult or wait until prototypes have arrived to test the new catapult design based on the new specifications.

CONSEQUENCES OF THE DECISION:

The marketing team has found that there is a competitor to this potential business and there are large contracts with universities for the first to market. The marketing team estimates that besides the expected sales, there is a potential for an additional $1,000,000 for being first to market. If you choose to proceed now, you will beat the competitors to market.

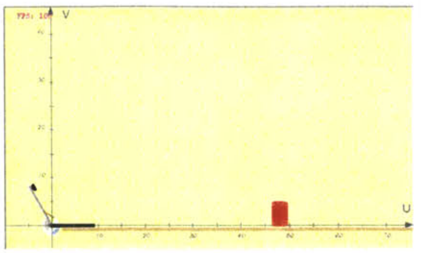
However, engineering has warned that if a catapult is delivered to market without meeting its primary performance requirement, there is a potential to lose $3,000,000 in warranty.

Should you choose to delay the product launch, there is the lost opportunity of the $1,000,000 early contracts and an additional $500,000 in redesign efforts and retesting. However, if the design is at risk of being insufficient, then there is a potential savings of $3,000,000 from the warranty by delaying the product launch.

SOURCE OF THE MODEL:

Fortunately, a model of the catapult exists to use for scenario testing. The model was developed by a **PhD student in the Mechanical Engineering department at MIT**. This model uses **principles of physics** to estimate the landing position of the ball. The PhD student has modified the model to meet the new design specifications. You can use this model to help you decide whether to release the product or delay launch.

DESIGN PROBLEM:



The original prototypes tested used a **SMOOTH ball** and **2 rubber bands**. Due to cost reductions, the ball and number of rubber bands has been changed. You must now release a catapult that uses a **PERFORATED ball** and only **1 rubber band**.

In addition, a constraint has been applied such that the **launch angle plus pullback angle must not exceed 90 degrees**. Customer testing found interference if the arm of the catapult were to pass 90 degrees during pullback. (For example, if you were to set a 60 degree launch angle, you could pullback only 30 degrees for a total of 90 degrees)

Due to time constraints, the remaining factors of the catapult such as the material of the arm, installation set-up, etc are not proposed to change as these are long lead-time items to acquire and verify.

The primary performance metric for the catapult is unchanged. The catapult must be able to get a ball into a cup located 4 feet (48 inches) from the catapult as shown in the graphic below. It is understood by the design team, that as long as the catapult **can achieve at least 48 inches in distance**, it can be set-up by the customer to land in the cup.

VALIDATION DATA:

A series of tests were performed on the actual prototype catapults to create validation data for the model. This validation data will be available to you when you access the model. Although this data does not give results based on the new requirements, you can use this validation data to help you understand if the model is calibrated well and an appropriate tool to make your design decision.

TIME LIMIT:

Given the time constraints, you will have **15 minutes** to make a decision. In that time, you can do the following:

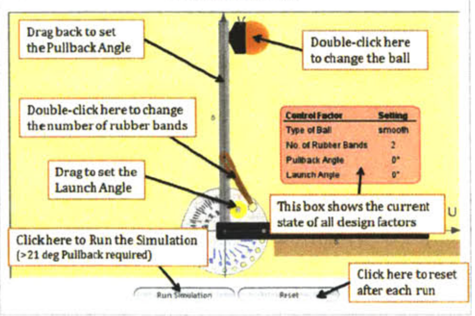
- Download the model from the following page
- Check the model against the validation data
- Run the new design scenarios with the model

Note that after this period of time, the page with the validation data will go away. ***Please be sure to enter a decision as soon as this occurs*** - or earlier if you are finished with your evaluation.

QUICK MODEL USER GUIDE

Please click on the picture below to see a more detailed description if needed. It will open in a new window.

Simulation Interface

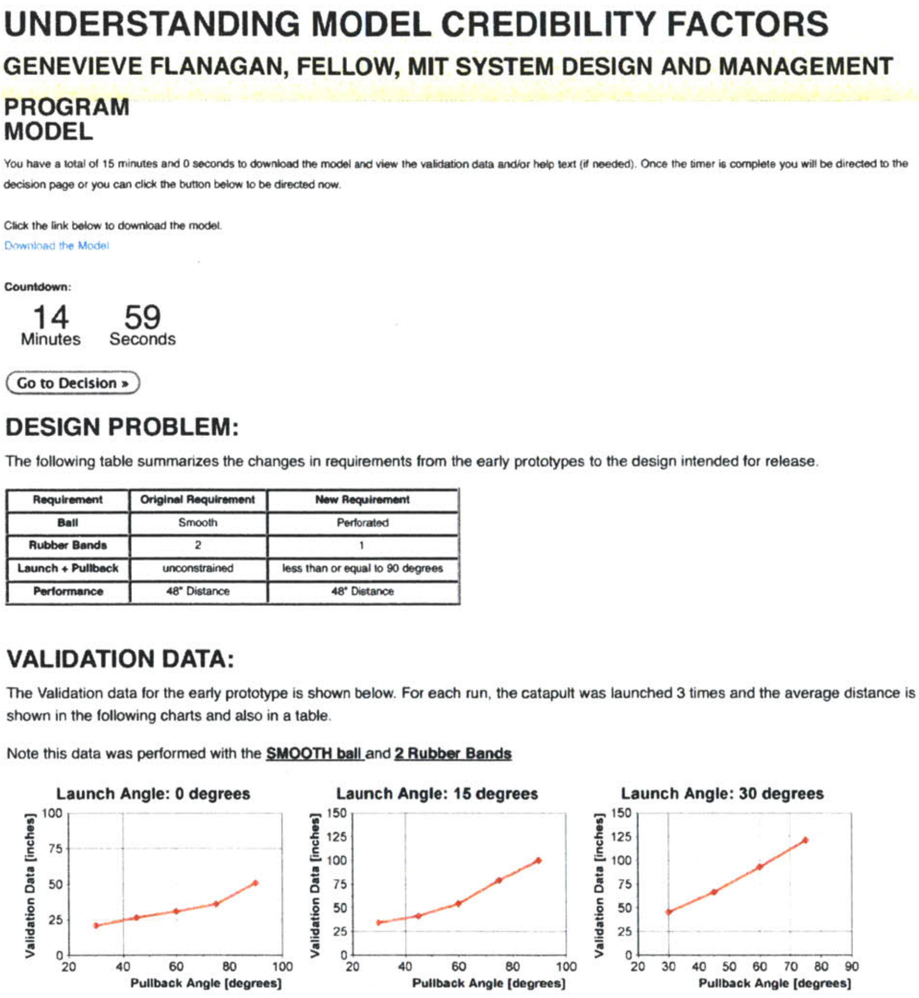

Drag back to set the Pullback Angle
Double-click here to change the ball
Double-click here to change the number of rubber bands
Drag to set the Launch Angle
This box shows the current state of all design factors
Click here to Run the Simulation (>21 deg Pullback required)
Click here to reset after each run
V
U
Run Simulation
Reset

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	0°

Go to Experiment »

8.3.2.2 Model with Validation Data Page

http://thesis.spinutech.com/model.aspx



UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

MODEL

You have a total of 15 minutes and 0 seconds to download the model and view the validation data and/or help text (if needed). Once the timer is complete you will be directed to the decision page or you can click the button below to be directed now.

Click the link below to download the model.

Download the Model

Countdown:

14 Minutes 59 Seconds

Go to Decision »

DESIGN PROBLEM:

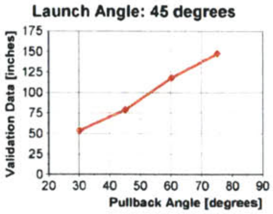
The following table summarizes the changes in requirements from the early prototypes to the design intended for release.

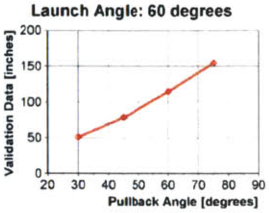
Requirement	Original Requirement	New Requirement
Ball	Smooth	Perforated
Rubber Bands	2	1
Launch + Pullback	unconstrained	less than or equal to 90 degrees
Performance	48" Distance	48" Distance

VALIDATION DATA:

The Validation data for the early prototype is shown below. For each run, the catapult was launched 3 times and the average distance is shown in the following charts and also in a table.

Note this data was performed with the **SMOOTH ball** and **2 Rubber Bands**

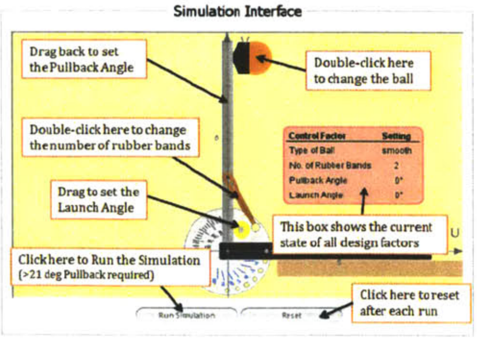




Experiment #	Launch Angle (degrees)	Pullback Angle (degrees)	Distance Travelled (inches)
1	0	30	20.7
2	0	45	26.5
3	0	60	30.8
4	0	75	36.0
5	0	90	50.8
6	15	30	34.3
7	15	45	41.3
8	15	60	54.3
9	15	75	78.7
10	15	90	99.8
11	30	30	45.5
12	30	45	66.5
13	30	60	92.8
14	30	75	121.0
15	45	30	53.8
16	45	45	79.2
17	45	60	118.0
18	60	30	51.2
19	60	45	78.7
20	60	60	114.7

QUICK MODEL USER GUIDE

Please click on the picture below to see a more detailed description if needed. It will open in a new window.


Simulation Interface

Drag back to set the Pullback Angle

Double-click here to change the ball

Double-click here to change the number of rubber bands

Control Factor	Setting
Type of Ball	smooth
No. of Rubber Bands	2
Pullback Angle	0°
Launch Angle	0°

Drag to set the Launch Angle

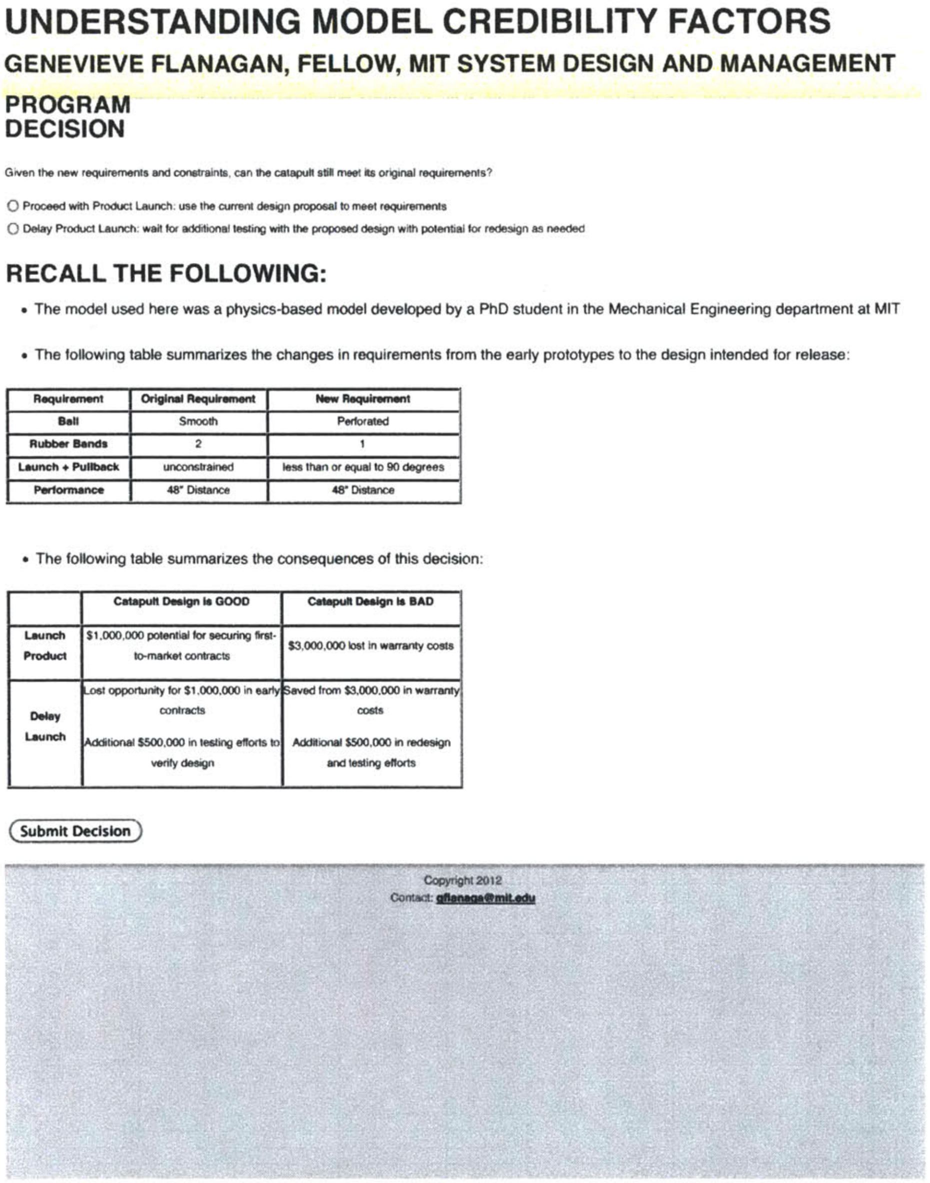
This box shows the current state of all design factors

Click here to Run the Simulation (>21 deg Pullback required)

Click here to reset after each run

8.3.2.3 Decision Page

http://thesis.spinutech.com/decision.aspx



UNDERSTANDING MODEL CREDIBILITY FACTORS

GENEVIEVE FLANAGAN, FELLOW, MIT SYSTEM DESIGN AND MANAGEMENT PROGRAM

DECISION

Given the new requirements and constraints, can the catapult still meet its original requirements?

○ Proceed with Product Launch: use the current design proposal to meet requirements

○ Delay Product Launch: wait for additional testing with the proposed design with potential for redesign as needed

RECALL THE FOLLOWING:

- The model used here was a physics-based model developed by a PhD student in the Mechanical Engineering department at MIT
- The following table summarizes the changes in requirements from the early prototypes to the design intended for release:

Requirement	Original Requirement	New Requirement
Ball	Smooth	Perforated
Rubber Bands	2	1
Launch + Pullback	unconstrained	less than or equal to 90 degrees
Performance	48" Distance	48" Distance

- The following table summarizes the consequences of this decision:

	Catapult Design is GOOD	Catapult Design is BAD
Launch Product	$1,000,000 potential for securing first-to-market contracts	$3,000,000 lost in warranty costs
Delay Launch	Lost opportunity for $1,000,000 in early contracts Additional $500,000 in testing efforts to verify design	Saved from $3,000,000 in warranty costs Additional $500,000 in redesign and testing efforts

Copyright 2012

Contact: gflanaga@mit.edu

8.4 APPENDIX D: DATA FROM EXPERIMENT

ID	Model Quality	Factor Tested	Decision	Model Confidence	Model Confidence (Numeric)	Impact of Validation Data	Capability of Model Predicting Distance	Capability of Model Predicting Distance (binned)
1	Good	Purpose	Wait	Low	3	No	4 - Good	Good
2	Good	Purpose	Wait	Low	1	No	1 - Poor: +/- 36 inches or more	Bad
3	Good	Time	Yes	High	4	Yes	4 - Good	Good
4	Bad	Uncertainty	Wait	Low	1	Yes	3 - Neutral	Bad
5	Bad	Purpose	Wait	Low	3	Yes	3 - Neutral	Bad
6	Bad	Source	Wait	Low	3	Yes	3 - Neutral	Bad
7	Bad	Uncertainty	Wait	Low	3	No	0 - Not acceptable	Bad
8	Bad	Consequences	Yes	Low	3	No	3 - Neutral	Bad
9	Bad	Time	Wait	Low	2	Yes	2 - Not Good	Bad
10	Good	Time	Wait		*	No	3 - Neutral	Bad
11	Good	Consequences	Yes	High	4	Yes	4 - Good	Good
12	Good	Purpose	Yes	High	4	No	4 - Good	Good
13	Good	Purpose	Wait	High	4	Yes	4 - Good	Good
14	Bad	Consequences	Yes	High	4	Yes	4 - Good	Good
15	Bad	Purpose	Yes	High	4	No	4 - Good	Good
16	Good	Reference	Wait		*			
17	Good	Uncertainty	Wait	High	4	No	0 - Not acceptable	Bad
18	Good	Time	Yes		*			
19	Good	Time	Wait	High	4	No	4 - Good	Good
20	Bad	Consequences	Wait	High	5	Yes	4 - Good	Good
21	Good	Time	Wait	Low	3	No	3 - Neutral	Bad
22	Good	Source	Wait	High	4	Yes	2 - Not Good	Bad
23	Good	Uncertainty	Wait	High	4	No	3 - Neutral	Bad
24	Good	Consequences	Wait	Low	1	No	1 - Poor: +/- 36 inches or more	Bad
25	Bad	Consequences	Wait		*	No	0 - Not acceptable	Bad
26	Good	Uncertainty	Wait	Low	1	No	3 - Neutral	Bad
27	Good	Purpose	Wait	Low	3	No	3 - Neutral	Bad
28	Bad	Source	Yes	High	5	Yes	4 - Good	Good
29	Bad	Uncertainty	Yes		*			
30	Bad	Consequences	Yes	High	4	No	4 - Good	Good
31	Bad	Time	Wait	Low	1	Yes	0 - Not acceptable	Bad
32	Bad	Purpose	Wait	Low	1	Yes	4 - Good	Good
33	Good	Purpose	Yes	Low	1	Yes	2 - Not Good	Bad
34	Bad	Uncertainty	Yes		*	Yes	3 - Neutral	Bad
35	Good	Source	Wait	Low	1	Yes	2 - Not Good	Bad
36	Bad	Uncertainty	Wait		*			
37	Bad	Source	Yes		*			
38	Bad	Source	Yes	High	4	Yes	4 - Good	Good
39	Bad	Reference	Yes	High	5	Yes	5 - Excellent: +/- 2 inches or less	Good
40	Bad	Consequences	Yes	High	5	No	4 - Good	Good
41	Good	Time	Wait	Low	3	Yes	3 - Neutral	Bad
42	Good	Time	Yes	Low	3	Yes	2 - Not Good	Bad
43	Good	Consequences	Yes	High	5	Yes	4 - Good	Good
44	Good	Reference	Yes	High	4	No	4 - Good	Good
45	Bad	Time	Yes	High	5	Yes	3 - Neutral	Bad
46	Good	Reference	Yes		*			
47	Good	Consequences	Wait	Low	2	Yes	3 - Neutral	Bad
48	Bad	Reference	Yes	High	5	Yes	4 - Good	Good
49	Good	Purpose	Wait	Low	1	No	3 - Neutral	Bad
50	Good	Source	Wait	Low	1	No	3 - Neutral	Bad
51	Good	Uncertainty	Wait		*			
52	Good	Source	Wait	High	4	Yes	3 - Neutral	Bad
53	Bad	Source	Wait	Low	1	Yes	3 - Neutral	Bad
54	Good	Time	Wait	Low	2	No	3 - Neutral	Bad
55	Bad	Reference	Wait	High	4	No	5 - Excellent: +/- 2 inches or less	Good
56	Bad	Reference	Yes	High	4	Yes	3 - Neutral	Bad
57	Bad	Time	Wait	Low	3	No	3 - Neutral	Bad
58	Bad	Time	Yes	Low	3	Yes	2 - Not Good	Bad
59	Bad	Purpose	Yes		*			
60	Bad	Consequences	Yes		*			
61	Good	Reference	Wait	Low	3	Yes	3 - Neutral	Bad
62	Good	Purpose	Wait	High	4	No	3 - Neutral	Bad
63	Good	Uncertainty	Yes		*			
64	Good	Consequences	Wait	Low	1	Yes	0 - Not acceptable	Bad
65	Good	Reference	Yes	Low	3	Yes	3 - Neutral	Bad
66	Good	Source	Wait	High	5	No	4 - Good	Good
67	Bad	Reference	Wait	Low	1	Yes	2 - Not Good	Bad
68	Bad	Consequences	Wait		*			
69	Good	Consequences	Yes		*			
70	Good	Time	Yes	Low	3	No	3 - Neutral	Bad
71	Bad	Time	Wait	Low	3	No	4 - Good	Good
72	Good	Source	Yes		*			
73	Bad	Consequences	Wait	High	4	No	4 - Good	Good
74	Bad	Purpose	Wait	High	5	No	3 - Neutral	Bad
75	Bad	Uncertainty	Wait	Low	1	Yes	4 - Good	Good
76	Good	Uncertainty	Wait	Low	1	Yes	0 - Not acceptable	Bad
77	Good	Reference	Wait	High	4	Yes	4 - Good	Good
78	Good	Reference	Yes	High	5	Yes	4 - Good	Good
79	Good	Time	Yes		*	Yes	4 - Good	Good
80	Bad	Time	Wait		*			
81	Bad	Time	Wait	High	4	No	4 - Good	Good
82	Bad	Uncertainty	Wait	High	4	Yes	4 - Good	Good
83	Bad	Consequences	Yes	Low	3	Yes	3 - Neutral	Bad
84	Good	Reference	Yes	High	4	Yes	4 - Good	Good
85	Good	Reference	Yes	High	4	No	4 - Good	Good
86	Good	Source	Wait	Low	1	No	3 - Neutral	Bad
87	Good	Source	Yes	High	4	Yes	2 - Not Good	Bad
88	Good	Source	Wait	High	4	Yes	4 - Good	Good
89	Good	Uncertainty	Yes	High	4	Yes	3 - Neutral	Bad
90	Good	Source	Wait	Low	1	Yes		

ID	Model Quality	Factor Tested	Decision	Model Confidence	Model Confidence (Numeric)	Impact of Validation Data	Capability of Model Predicting Distance	Capability of Model Predicting Distance (binned)
91	Good	Reference	Yes	High	5	Yes	5 - Excellent: +/- 2 inches or less	Good
92	Bad	Consequences	Wait	Low	1	Yes	3 - Neutral	Bad
93	Bad	Time	Yes	High	4	Yes	4 - Good	Good
94	Good	Time	Yes	Low	3	Yes	3 - Neutral	Bad
95	Bad	Time	Wait	High	4	Yes	3 - Neutral	Bad
96	Bad	Time	Wait	Low	3	Yes	2 - Not Good	Bad
97	Good	Reference	Yes	High	5	No	4 - Good	Good
98	Good	Consequences	Yes	Low	3	Yes	3 - Neutral	Bad
99	Bad	Reference	Yes		*			
100	Bad	Uncertainty	Yes	Low	3	Yes	3 - Neutral	Bad
101	Bad	Purpose	Yes	Low	1	Yes	4 - Good	Good
102	Bad	Purpose	Yes	High	5	No	4 - Good	Good
103	Good	Reference	Yes	High	5	No	3 - Neutral	Bad
104	Good	Uncertainty	Yes		*			
105	Bad	Purpose	Wait	Low	3	No	3 - Neutral	Bad
106	Good	Time	Wait		*			
107	Bad	Purpose	Yes	Low	3	No	3 - Neutral	Bad
108	Bad	Consequences	Wait	Low	3	Yes	3 - Neutral	Bad
109	Bad	Time	Yes		*			
110	Good	Consequences	Wait	High	4	No	3 - Neutral	Bad
111	Bad	Source	Yes	High	4	No	3 - Neutral	Bad
112	Good	Source	Wait	Low	1	No	3 - Neutral	Bad
113	Good	Consequences	Wait	High	5	Yes	4 - Good	Good
114	Bad	Time	Yes		*			
115	Bad	Purpose	Wait	High	4	Yes	4 - Good	Good
116	Good	Time	Wait		*			
117	Bad	Time	Yes	Low	3	Yes	3 - Neutral	Bad
118	Good	Consequences	Yes	High	4	No	4 - Good	Good
119	Bad	Source	Yes	Low	3	No	3 - Neutral	Bad
120	Good	Time	Wait	Low	2	Yes	3 - Neutral	Bad
121	Bad	Uncertainty	Wait	High	4	No	3 - Neutral	Bad
122	Good	Reference	Yes	High	4	Yes	3 - Neutral	Bad
123	Bad	Reference	Wait	Low	1	Yes	3 - Neutral	Bad
124	Good	Consequences	Wait	Low	3	Yes	3 - Neutral	Bad
125	Bad	Reference	Yes	High	5	Yes	4 - Good	Good
126	Good	Consequences	Wait	Low	2	No	3 - Neutral	Bad
127	Good	Reference	Wait	Low	2	Yes	2 - Not Good	Bad
128	Good	Time	Wait	Low	1	Yes	3 - Neutral	Bad
129	Bad	Reference	Yes	High	4	Yes	3 - Neutral	Bad
130	Bad	Purpose	Wait	Low	3	No	5 - Excellent: +/- 2 inches or less	Good
131	Bad	Source	Wait	Low	1	Yes	1 - Poor: +/- 36 inches or more	Bad
132	Good	Consequences	Yes	High	4	Yes	4 - Good	Good
133	Good	Uncertainty	Yes	Low	3	No	3 - Neutral	Bad
134	Good	Uncertainty	Wait	High	4	Yes	4 - Good	Good
135	Bad	Uncertainty	Yes	High	4	Yes		
136	Bad	Purpose	Yes	Low	2	Yes	2 - Not Good	Bad
137	Bad	Source	Yes		*			
138	Good	Reference	Yes	High	5	Yes		
139	Bad	Time	Yes		*			
140	Bad	Purpose	Wait	Low	2	No		
141	Bad	Uncertainty	Yes	High	5	Yes	4 - Good	Good
142	Bad	Purpose	Wait	Low	2	Yes	3 - Neutral	Bad
143	Good	Source	Wait	Low	1	Yes	2 - Not Good	Bad
144	Good	Source	Yes		*	Yes	4 - Good	Good
145	Good	Time	Yes	High	4	Yes	3 - Neutral	Bad
146	Good	Time	Yes	Low	3	Yes	2 - Not Good	Bad
147	Bad	Reference	Yes	High	4	Yes	4 - Good	Good
148	Good	Source	Wait	High	4	Yes	3 - Neutral	Bad
149	Bad	Consequences	Wait	High	4	Yes	3 - Neutral	Bad
150	Bad	Purpose	Yes	Low	1	Yes	3 - Neutral	Bad
151	Good	Uncertainty	Wait		*			
152	Bad	Reference	Yes	High	5	Yes	5 - Excellent: +/- 2 inches or less	Good
153	Good	Purpose	Wait	Low	3	Yes	2 - Not Good	Bad
154	Good	Consequences	Wait	Low	3	No		
155	Good	Consequences	Wait	Low	1	Yes	2 - Not Good	Bad
156	Bad	Reference	Wait	High	5	Yes	5 - Excellent: +/- 2 inches or less	Good
157	Bad	Time	Yes	Low	3	Yes	4 - Good	Good
158	Bad	Reference	Wait	Low	2	No	4 - Good	Good
159	Bad	Uncertainty	Wait	Low	1	Yes	2 - Not Good	Bad
160	Bad	Source	Yes	High	4	Yes	4 - Good	Good
161	Bad	Uncertainty	Yes	Low	3	No	3 - Neutral	Bad
162	Good	Reference	Yes	Low	3	Yes	4 - Good	Good
163	Bad	Source	Wait	Low	2	No	3 - Neutral	Bad
164	Good	Source	Wait	Low	2	Yes	4 - Good	Good
165	Good	Purpose	Wait	High	4	Yes	4 - Good	Good
166	Bad	Source	Wait	Low	1	Yes	3 - Neutral	Bad
167	Bad	Time	Yes	High	5	No	4 - Good	Good
168	Bad	Source	Wait	Low	2	Yes	1 - Poor: +/- 36 inches or more	Bad
169	Bad	Reference	Yes	High	4	No	2 - Not Good	Bad
170	Good	Consequences	Yes	High	5	Yes	4 - Good	Good
171	Good	Time	Wait	Low	2	Yes	3 - Neutral	Bad
172	Good	Purpose	Yes	High	4	Yes	4 - Good	Good
173	Good	Uncertainty	Yes		*			
174	Bad	Consequences	Yes	High	5	No	4 - Good	Good
175	Bad	Time	Yes	High	4	Yes	4 - Good	Good
176	Good	Uncertainty	Wait	High	4	Yes	4 - Good	Good
177	Good	Uncertainty	Yes	High	5	Yes	4 - Good	Good
178	Bad	Reference	Wait	Low	1	Yes	1 - Poor: +/- 36 inches or more	Bad
179	Good	Reference	Wait	Low	3	No	4 - Good	Good
180	Good	Reference	Yes	High	4	Yes	4 - Good	Good

ID	Model Quality	Factor Tested	Decision	Model Confidence	Model Confidence (Numeric)	Impact of Validation Data	Capability of Model Predicting Distance	Capability of Model Predicting Distance (binned)
181	Good	Source	Wait	High	4	No	0 - Not acceptable	Bad
182	Good	Uncertainty	Wait	High	5	Yes	4 - Good	Good
183	Good	Purpose	Wait	Low	1	Yes	1 - Poor: +/- 36 inches or more	Bad
184	Bad	Consequences	Wait	Low	1	Yes	0 - Not acceptable	Bad
185	Good	Purpose	Wait	Low	1	No	4 - Good	Good
186	Good	Consequences	Wait	Low	3	Yes	3 - Neutral	Bad
187	Bad	Purpose	Wait	Low	1	No	5 - Excellent: +/- 2 inches or less	Good
188	Bad	Consequences	Wait	High	5	No	3 - Neutral	Bad
189	Good	Purpose	Yes	High	5	No	4 - Good	Good
190	Bad	Source	Yes	Low	3	Yes	2 - Not Good	Bad
191	Bad	Consequences	Yes	Low	3	Yes	4 - Good	Good
192	Bad	Purpose	Wait	Low	3	No	4 - Good	Good
193	Bad	Consequences	Wait	Low	1	Yes	2 - Not Good	Bad
194	Good	Reference	Yes		*			
195	Bad	Consequences	Wait	Low	1	Yes	1 - Poor: +/- 36 inches or more	Bad
196	Good	Purpose	Wait	Low	1	No	2 - Not Good	Bad
197	Bad	Uncertainty	Wait	High	5	No	4 - Good	Good
198	Good	Uncertainty	Yes	High	4	Yes	3 - Neutral	Bad
199	Good	Uncertainty	Wait	Low	2	No	3 - Neutral	Bad
200	Bad	Source	Yes	High	4	No	4 - Good	Good
201	Bad	Reference	Yes	Low	3	Yes	3 - Neutral	Bad
202	Bad	Source	Yes	High	5	Yes	4 - Good	Good
203	Bad	Uncertainty	Yes	High	5	Yes	4 - Good	Good
204	Good	Time	Yes	Low	1	No	0 - Not acceptable	Bad
205	Good	Source	Yes	High	4	No	3 - Neutral	Bad
206	Bad	Consequences	Yes	High	4	No	5 - Excellent: +/- 2 inches or less	Good
207	Good	Uncertainty	Wait	Low	2	Yes	2 - Not Good	Bad
208	Bad	Reference	Wait	High	5	Yes	4 - Good	Good
209	Good	Uncertainty	Yes		*			
210	Bad	Reference	Wait		*			
211	Good	Reference	Wait	Low	3	Yes	2 - Not Good	Bad
212	Good	Source	Wait	High	5	No	4 - Good	Good
213	Bad	Reference	Wait		*			
214	Good	Source	Yes	High	4	No	4 - Good	Good
215	Good	Reference	Wait	High	5	Yes	5 - Excellent: +/- 2 inches or less	Good
216	Bad	Purpose	Yes		*			
217	Bad	Reference	Yes	High	5	Yes	4 - Good	Good
218	Good	Time	Yes	High	5	Yes	4 - Good	Good
219	Bad	Time	Wait	Low	3	No	3 - Neutral	Bad
220	Good	Time	Wait	Low	1	Yes	3 - Neutral	Bad
221	Good	Source	Wait	Low	1	No	0 - Not acceptable	Bad
222	Bad	Source	Yes		*			
223	Good	Consequences	Yes	High	5	Yes	4 - Good	Good
224	Bad	Consequences	Wait	High	4	Yes	0 - Not acceptable	Bad
225	Bad	Source	Yes	Low	3	Yes	3 - Neutral	Bad
226	Bad	Time	Wait	Low	1	Yes	2 - Not Good	Bad
227	Bad	Uncertainty	Yes	High	4	Yes	5 - Excellent: +/- 2 inches or less	Good
228	Bad	Time	Wait	High	4	No	3 - Neutral	Bad
229	Bad	Purpose	Yes	High	5	No	4 - Good	Good
230	Bad	Reference	Yes	High	5	No	3 - Neutral	Bad
231	Bad	Uncertainty	Wait		*			
232	Bad	Source	Wait	Low	1	No	4 - Good	Good
233	Bad	Consequences	Wait	Low	1	Yes	0 - Not acceptable	Bad
234	Bad	Time	Yes		*			
235	Good	Purpose	Wait	High	4	Yes	4 - Good	Good
236	Bad	Source	Yes	High	4	Yes	4 - Good	Good
237	Good	Source	Wait	Low	2	Yes	2 - Not Good	Bad
238	Bad	Consequences	Yes	High	4	Yes	4 - Good	Good
239	Bad	Uncertainty	Wait	High	5	Yes	3 - Neutral	Bad
240	Good	Consequences	Yes	High	5	No	3 - Neutral	Bad
241	Bad	Purpose	Yes	Low	3	No	3 - Neutral	Bad
242	Good	Uncertainty	Yes	High	5	No	3 - Neutral	Bad
243	Good	Purpose	Wait	High	4	Yes		
244	Bad	Source	Wait	Low	2	No	2 - Not Good	Bad
245	Good	Purpose	Wait	Low	1	Yes	4 - Good	Good
246	Good	Reference	Yes	High	5	Yes	4 - Good	Good
247	Good	Reference	Wait	Low	2	Yes	3 - Neutral	Bad
248	Good	Purpose	Wait	Low	3	No	3 - Neutral	Bad
249	Good	Source	Yes	High	4	No	4 - Good	Good
250	Bad	Purpose	Yes		*			
251	Bad	Reference	Wait	Low	3	Yes	3 - Neutral	Bad
252	Good	Purpose	Yes	High	4	Yes	4 - Good	Good

ID	Capability of Catapult Predicting Distance	Capability of Catapult Predicting Distance (binned)	Was there enough time?	Influenced by Consequences	Importance of Source Author	Importance of Source Author (binned)
1	3 - Neutral	Bad	Yes	Yes	3 - I considered It	Not Important
2	1 - Poor: +/- 36 inches or more	Bad	No		2 - Little importance	Not Important
3	2 - Not Good	Bad	No	Yes	3 - I considered It	Not Important
4	2 - Not Good	Bad	No	Yes	3 - I considered It	Not Important
5	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
6	3 - Neutral	Bad		Yes	4 - Somewhat important	Important
7	3 - Neutral	Bad	No	Yes	1 - Not Important	Not Important
8	3 - Neutral	Bad	Yes	No	3 - I considered It	Not Important
9	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
10	4 - Good	Good	No	Yes	3 - I considered It	Not Important
11	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
12	3 - Neutral	Bad	Yes	Yes	1 - Not Important	Not Important
13	3 - Neutral	Bad	No	Yes	2 - Little importance	Not Important
14	3 - Neutral	Bad	Yes		1 - Not Important	Not Important
15	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	4 - Somewhat important	Important
16						
17	3 - Neutral	Bad	Yes	Yes	4 - Somewhat important	Important
18						
19	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
20	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
21	3 - Neutral	Bad	No	Yes	2 - Little importance	Not Important
22	2 - Not Good	Bad	Yes	Yes	2 - Little importance	Not Important
23	3 - Neutral	Bad	Yes	No	3 - I considered It	Not Important
24	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
25	0 - Not acceptable	Bad	Yes	No	1 - Not Important	Not Important
26	3 - Neutral	Bad	Yes	Yes	5 - Very Important	Important
27	4 - Good	Good	Yes	Yes	5 - Very Important	Important
28	3 - Neutral	Bad	Yes	Yes	3 - I considered It	Not Important
29						
30	4 - Good	Good	No	Yes	4 - Somewhat important	Important
31	1 - Poor: +/- 36 inches or more	Bad	Yes		2 - Little importance	Not Important
32	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
33	5 - Excellent: +/- 2 inches or less	Good	Yes	No	3 - I considered It	Not Important
34	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
35	2 - Not Good	Bad	No		4 - Somewhat important	Important
36						
37						
38	3 - Neutral	Bad	Yes	No	3 - I considered It	Not Important
39			Yes	Yes	4 - Somewhat important	Important
40	5 - Excellent: +/- 2 inches or less	Good	Yes	No	4 - Somewhat important	Important
41	5 - Excellent: +/- 2 inches or less	Good	Yes	No	4 - Somewhat important	Important
42	4 - Good	Good	No	Yes	2 - Little importance	Not Important
43	2 - Not Good	Bad	No		4 - Somewhat important	Important
44	4 - Good	Good	Yes	Yes	5 - Very Important	Important
45	4 - Good	Good	No	Yes	1 - Not Important	Not Important
46						
47	5 - Excellent: +/- 2 inches or less	Good	No	No	3 - I considered It	Not Important
48	4 - Good	Good	Yes	Yes	5 - Very Important	Important
49	4 - Good	Good	Yes	Yes	1 - Not Important	Not Important
50	3 - Neutral	Bad		Yes	1 - Not Important	Not Important
51						
52	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
53	3 - Neutral	Bad	Yes	Yes	5 - Very Important	Important
54	4 - Good	Good	No	Yes	3 - I considered It	Not Important
55	4 - Good	Good	No	Yes	1 - Not Important	Not Important
56	4 - Good	Good	Yes	Yes	4 - Somewhat important	Important
57	4 - Good	Good	No	Yes	5 - Very Important	Important
58	2 - Not Good	Bad	No	Yes	3 - I considered It	Not Important
59						
60						
61	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	3 - I considered It	Not Important
62	4 - Good	Good		Yes	4 - Somewhat important	Important
63						
64	2 - Not Good	Bad	Yes	Yes	1 - Not Important	Not Important
65	4 - Good	Good	No	Yes	1 - Not Important	Not Important
66	4 - Good	Good	Yes	No	4 - Somewhat important	Important
67	3 - Neutral	Bad	No	No	2 - Little importance	Not Important
68						
69						
70	4 - Good	Good	No	No	4 - Somewhat important	Important
71	3 - Neutral	Bad	No	Yes	4 - Somewhat important	Important
72						
73	3 - Neutral	Bad	Yes	No	2 - Little importance	Not Important
74	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
75	4 - Good	Good	Yes	Yes	5 - Very Important	Important
76	0 - Not acceptable	Bad	Yes	Yes	3 - I considered It	Not Important
77	4 - Good	Good	Yes	Yes	5 - Very Important	Important
78	4 - Good	Good	Yes	No	2 - Little importance	Not Important
79	4 - Good	Good	No	Yes	3 - I considered It	Not Important
80						
81	4 - Good	Good	No	Yes	5 - Very Important	Important
82	3 - Neutral	Bad	Yes	Yes	4 - Somewhat important	Important
83	4 - Good	Good	Yes	Yes	1 - Not Important	Not Important
84	4 - Good	Good	Yes	Yes	5 - Very Important	Important
85	5 - Excellent: +/- 2 inches or less	Good	Yes	No	4 - Somewhat important	Important
86	4 - Good	Good	Yes	Yes	4 - Somewhat important	Important
87	5 - Excellent: +/- 2 inches or less	Good	No	Yes	3 - I considered It	Not Important
88	4 - Good	Good	Yes	Yes	4 - Somewhat important	Important
89	3 - Neutral	Bad	Yes	No	3 - I considered It	Not Important
90	3 - Neutral	Bad	Yes	Yes	5 - Very Important	Important

ID	Capability of Catapult Predicting Distance	Capability of Catapult Predicting Distance (binned)	Was there enough time?	Influenced by Consequences	Importance of Source Author	Importance of Source Author (binned)
91	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	5 - Very Important	Important
92	3 - Neutral	Bad	No	No	5 - Very Important	Important
93	4 - Good	Good	Yes	Yes	1 - Not Important	Not Important
94	2 - Not Good	Bad	No	No	2 - Little importance	Not Important
95	4 - Good	Good	No	Yes	2 - Little importance	Not Important
96	4 - Good	Good	No	Yes	3 - I considered It	Not Important
97	5 - Excellent: +/- 2 inches or less	Good	No	No	1 - Not Important	Not Important
98	2 - Not Good	Bad	No	Yes	2 - Little importance	Not Important
99						
100	4 - Good	Good	Yes	No	2 - Little importance	Not Important
101	4 - Good	Good		No	3 - I considered It	Not Important
102	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
103	2 - Not Good	Bad	No	No	4 - Somewhat important	Important
104						
105	3 - Neutral	Bad	Yes	Yes	1 - Not Important	Not Important
106						
107	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
108	3 - Neutral	Bad	Yes	No	3 - I considered It	Not Important
109						
110	3 - Neutral	Bad	No	No	1 - Not Important	Not Important
111	4 - Good	Good	No	No	4 - Somewhat important	Important
112	4 - Good	Good	Yes	Yes	5 - Very Important	Important
113	4 - Good	Good	No	No	3 - I considered It	Not Important
114						
115	5 - Excellent: +/- 2 inches or less	Good	No	Yes	4 - Somewhat important	Important
116						
117	4 - Good	Good	No	Yes	3 - I considered It	Not Important
118	5 - Excellent: +/- 2 inches or less	Good	Yes	No	3 - I considered It	Not Important
119	4 - Good	Good	Yes	No	2 - Little importance	Not Important
120	4 - Good	Good	No	Yes	2 - Little importance	Not Important
121	3 - Neutral	Bad	No	Yes	4 - Somewhat important	Important
122	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	4 - Somewhat important	Important
123	3 - Neutral	Bad	No	Yes	2 - Little importance	Not Important
124	4 - Good	Good	Yes	No	2 - Little importance	Not Important
125	3 - Neutral	Bad	Yes		5 - Very Important	Important
126	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
127	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	3 - I considered It	Not Important
128	5 - Excellent: +/- 2 inches or less	Good	No	No	1 - Not Important	Not Important
129	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	1 - Not Important	Not Important
130	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	1 - Not Important	Not Important
131	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
132	2 - Not Good	Bad	Yes	No	1 - Not Important	Not Important
133	2 - Not Good	Bad	Yes	No	3 - I considered It	Not Important
134	4 - Good	Good	Yes		4 - Somewhat important	Important
135	4 - Good	Good	No	Yes	4 - Somewhat important	Important
136	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
137						
138			Yes	Yes	4 - Somewhat important	Important
139						
140	2 - Not Good	Bad	Yes	No	3 - I considered It	Not Important
141	4 - Good	Good	Yes	Yes	5 - Very Important	Important
142	1 - Poor: +/- 36 inches or more	Bad	No	Yes	3 - I considered It	Not Important
143	0 - Not acceptable	Bad	No	No	3 - I considered It	Not Important
144	5 - Excellent: +/- 2 inches or less	Good	Yes	No	3 - I considered It	Not Important
145	2 - Not Good	Bad	Yes	Yes	5 - Very Important	Important
146	2 - Not Good	Bad	Yes	No	1 - Not Important	Not Important
147	4 - Good	Good	Yes	Yes	5 - Very Important	Important
148	4 - Good	Good	No	Yes	4 - Somewhat important	Important
149	2 - Not Good	Bad	Yes	Yes	2 - Little importance	Not Important
150	4 - Good	Good	No	No	2 - Little importance	Not Important
151						
152	4 - Good	Good	Yes	Yes	5 - Very Important	Important
153	5 - Excellent: +/- 2 inches or less	Good	No	Yes	4 - Somewhat important	Important
154	4 - Good	Good	No	No	4 - Somewhat important	Important
155	3 - Neutral	Bad	Yes	No	1 - Not Important	Not Important
156	5 - Excellent: +/- 2 inches or less	Good		Yes	3 - I considered It	Not Important
157	3 - Neutral	Bad	No	No	3 - I considered It	Not Important
158	2 - Not Good	Bad	Yes	Yes	3 - I considered It	Not Important
159	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
160	4 - Good	Good	Yes	No	4 - Somewhat important	Important
161	4 - Good	Good	Yes	No	5 - Very Important	Important
162	3 - Neutral	Bad	No	No	4 - Somewhat important	Important
163	4 - Good	Good	No	Yes	4 - Somewhat important	Important
164	4 - Good	Good	No	Yes	5 - Very Important	Important
165	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
166	4 - Good	Good	Yes	Yes	5 - Very Important	Important
167	4 - Good	Good	No	No	5 - Very Important	Important
168	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
169	4 - Good	Good	No	Yes	3 - I considered It	Not Important
170	4 - Good	Good	No	Yes	4 - Somewhat important	Important
171	4 - Good	Good	No	Yes	2 - Little importance	Not Important
172	4 - Good	Good	No	Yes	4 - Somewhat important	Important
173						
174	4 - Good	Good	Yes	No	3 - I considered It	Not Important
175	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
176	2 - Not Good	Bad	No	Yes	5 - Very Important	Important
177	4 - Good	Good	Yes		2 - Little importance	Not Important
178	4 - Good	Good	No	Yes	3 - I considered It	Not Important
179	3 - Neutral	Bad	Yes	Yes	4 - Somewhat important	Important
180	4 - Good	Good	Yes	No	2 - Little importance	Not Important

ID	Capability of Catapult Predicting Distance	Capability of Catapult Predicting Distance (binned)	Was there enough time?	Influenced by Consequences	Importance of Source Author	Importance of Source Author (binned)
181	5 - Excellent: +/- 2 inches or less	Good	No	Yes	5 - Very Important	Important
182	2 - Not Good	Bad	Yes	Yes	2 - Little importance	Not Important
183	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
184	4 - Good	Good	No	No	2 - Little importance	Not Important
185	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	3 - I considered It	Not Important
186	4 - Good	Good	Yes	No	2 - Little importance	Not Important
187	5 - Excellent: +/- 2 inches or less	Good	No	Yes	1 - Not Important	Not Important
188	4 - Good	Good	No	Yes	5 - Very Important	Important
189	3 - Neutral	Bad	No	Yes	2 - Little importance	Not Important
190	3 - Neutral	Bad	Yes	No	4 - Somewhat important	Important
191	4 - Good	Good	Yes	Yes	5 - Very Important	Important
192	3 - Neutral	Bad	No	Yes	3 - I considered It	Not Important
193	3 - Neutral	Bad	Yes	Yes	1 - Not Important	Not Important
194						
195	4 - Good	Good	No	Yes	3 - I considered It	Not Important
196	3 - Neutral	Bad	Yes	Yes	1 - Not Important	Not Important
197	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
198	3 - Neutral	Bad	Yes	Yes	3 - I considered It	Not Important
199	4 - Good	Good	No	Yes	2 - Little importance	Not Important
200	4 - Good	Good	No	Yes	3 - I considered It	Not Important
201	4 - Good	Good	Yes	Yes	4 - Somewhat important	Important
202	4 - Good	Good	Yes	No	2 - Little importance	Not Important
203	4 - Good	Good	Yes	Yes	4 - Somewhat important	Important
204	1 - Poor: +/- 36 inches or more	Bad	No	Yes	1 - Not Important	Not Important
205	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
206	3 - Neutral	Bad	Yes	No	3 - I considered It	Not Important
207	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
208	4 - Good	Good	Yes	No	5 - Very Important	Important
209						
210						
211	3 - Neutral	Bad	Yes	Yes	2 - Little importance	Not Important
212	4 - Good	Good	No	No	2 - Little importance	Not Important
213						
214	4 - Good	Good	Yes	Yes	2 - Little importance	Not Important
215	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	5 - Very Important	Important
216						
217	4 - Good	Good	No	Yes	3 - I considered It	Not Important
218	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
219	4 - Good	Good	No	Yes	3 - I considered It	Not Important
220	3 - Neutral	Bad	Yes	Yes	3 - I considered It	Not Important
221	3 - Neutral	Bad	No	No	5 - Very Important	Important
222						
223	4 - Good	Good	No	No	5 - Very Important	Important
224	5 - Excellent: +/- 2 inches or less	Good	No	Yes	3 - I considered It	Not Important
225	4 - Good	Good	No	Yes	5 - Very Important	Important
226	3 - Neutral	Bad	No	No	2 - Little importance	Not Important
227	4 - Good	Good	Yes	Yes	1 - Not Important	Not Important
228	4 - Good	Good	Yes	Yes	4 - Somewhat important	Important
229	4 - Good	Good	Yes	Yes	3 - I considered It	Not Important
230	3 - Neutral	Bad	Yes	Yes	1 - Not Important	Not Important
231						
232	3 - Neutral	Bad	Yes	No	2 - Little importance	Not Important
233	5 - Excellent: +/- 2 inches or less	Good	Yes	Yes	4 - Somewhat important	Important
234						
235	4 - Good	Good	Yes	Yes	5 - Very Important	Important
236	4 - Good	Good	No	Yes	3 - I considered It	Not Important
237	4 - Good	Good	Yes	Yes	5 - Very Important	Important
238	2 - Not Good	Bad	Yes	No	1 - Not Important	Not Important
239	3 - Neutral	Bad	No	No	4 - Somewhat important	Important
240	1 - Poor: +/- 36 inches or more	Bad	Yes	No	5 - Very Important	Important
241	3 - Neutral	Bad	Yes	Yes	4 - Somewhat important	Important
242	3 - Neutral	Bad	Yes	Yes	4 - Somewhat important	Important
243	3 - Neutral	Bad	No	No	2 - Little importance	Not Important
244	2 - Not Good	Bad	Yes	Yes	5 - Very Important	Important
245	4 - Good	Good	Yes	Yes	1 - Not Important	Not Important
246	4 - Good	Good	No	Yes	2 - Little importance	Not Important
247	4 - Good	Good	Yes	Yes	1 - Not Important	Not Important
248	3 - Neutral	Bad	No	No	1 - Not Important	Not Important
249	4 - Good	Good	Yes	Yes	5 - Very Important	Important
250						
251	4 - Good	Good	No	Yes	3 - I considered It	Not Important
252	2 - Not Good	Bad	No	Yes	3 - I considered It	Not Important

ID	Familiarity with Catapults	Familiarity with Catapults (binned)	Familiarity with Models	Familiarity with Models (binned)	Are Models available at Work?
1	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
2	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
3	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	Yes
4	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
5	1 - Minimal	Not Familiar	1 - No experience	Not Familiar	Yes
6	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
7	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
8	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
9	2 - I am aware of the concepts, but not practiced recently	Not Familiar	1 - No experience	Not Familiar	No - and we don't use them
10	4 - I am familiar with the concepts and I use them often	Familiar	5 - I do it every day	Familiar	Yes
11	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
12	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
13	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
14	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
15	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
16					
17	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
18					
19	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
20	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
21	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
22	3 - I am aware of the concepts	Familiar	5 - I do it every day	Familiar	No - so we have to build them as needed
23	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	No - so we have to build them as needed
24	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
25	1 - Minimal	Not Familiar	1 - No experience	Not Familiar	No - and we don't use them
26	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
27	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
28	4 - I am familiar with the concepts and I use them often	Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
29					
30	1 - Minimal	Not Familiar	4 - Often	Familiar	No - so we have to build them as needed
31	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	No - and we don't use them
32	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
33	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
34	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
35	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	Yes
36					
37					
38	4 - I am familiar with the concepts and I use them often	Familiar	3 - Occasional	Not Familiar	Yes
39	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	No - and we don't use them
40	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
41	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	No - so we have to build them as needed
42	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
43	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	No - and we don't use them
44	4 - I am familiar with the concepts and I use them often	Familiar	5 - I do it every day	Familiar	No - so we have to build them as needed
45	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
46					
47	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
48	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	No - so we have to build them as needed
49	1 - Minimal	Not Familiar	4 - Often	Familiar	Yes
50	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
51					
52	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
53	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
54	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
55	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	Yes
56	3 - I am aware of the concepts	Familiar	1 - No experience	Not Familiar	No - and we don't use them
57	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
58	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
59					
60					
61	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
62	3 - I am aware of the concepts	Familiar	1 - No experience	Not Familiar	No - and we don't use them
63					
64	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
65	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
66	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
67	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
68					
69					
70	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
71	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
72					
73	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
74	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
75	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
76	3 - I am aware of the concepts	Familiar	1 - No experience	Not Familiar	No - and we don't use them
77	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
78	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	Yes
79	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
80					
81	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	No - so we have to build them as needed
82	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
83	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
84	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
85	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	Yes
86	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
87	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
88	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
89	1 - Minimal	Not Familiar	1 - No experience	Not Familiar	Yes
90	2 - I am aware of the concepts, but not practiced recently	Not Familiar	5 - I do it every day	Familiar	Yes

ID	Familiarity with Catapults	Familiarity with Catapults (binned)	Familiarity with Models	Familiarity with Models (binned)	Are Models available at Work?
91	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	No - so we have to build them as needed
92	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	Yes
93	4 - I am familiar with the concepts and I use them often	Familiar	2 - Minimal	Not Familiar	Yes
94	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	Yes
95	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
96	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
97	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
98	4 - I am familiar with the concepts and I use them often	Familiar	5 - I do it every day	Familiar	Yes
99					
100	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
101	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
102	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
103	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
104					
105	2 - I am aware of the concepts, but not practiced recently	Not Familiar	1 - No experience	Not Familiar	No - so we have to build them as needed
106					
107	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
108	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	No - so we have to build them as needed
109					
110	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
111	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
112	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
113	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	No - so we have to build them as needed
114					
115	4 - I am familiar with the concepts and I use them often	Familiar	3 - Occasional	Not Familiar	Yes
116					
117	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
118	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
119	4 - I am familiar with the concepts and I use them often	Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
120	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
121	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
122	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
123	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
124	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	No - so we have to build them as needed
125	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	No - so we have to build them as needed
126	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
127	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - and we don't use them
128	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
129	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
130	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - and we don't use them
131	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	Yes
132	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
133	3 - I am aware of the concepts	Familiar	5 - I do it every day	Familiar	Yes
134	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
135	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
136	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
137					
138	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	No - so we have to build them as needed
139					
140	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
141	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
142	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
143	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
144	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
145	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
146	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
147	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - and we don't use them
148	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
149	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
150	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
151					
152	4 - I am familiar with the concepts and I use them often	Familiar	2 - Minimal	Not Familiar	Yes
153	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
154	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
155	2 - I am aware of the concepts, but not practiced recently	Not Familiar	1 - No experience	Not Familiar	No - so we have to build them as needed
156	4 - I am familiar with the concepts and I use them often	Familiar	2 - Minimal	Not Familiar	No - and we don't use them
157	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
158	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	No - and we don't use them
159	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	No - so we have to build them as needed
160	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	Yes
161	4 - I am familiar with the concepts and I use them often	Familiar	5 - I do it every day	Familiar	No - so we have to build them as needed
162	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
163	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	No - and we don't use them
164	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
165	4 - I am familiar with the concepts and I use them often	Familiar	2 - Minimal	Not Familiar	Yes
166	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
167	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
168	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
169	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
170	4 - I am familiar with the concepts and I use them often	Familiar	5 - I do it every day	Familiar	Yes
171	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	No - so we have to build them as needed
172	4 - I am familiar with the concepts and I use them often	Familiar	5 - I do it every day	Familiar	Yes
173					
174	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
175	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
176	3 - I am aware of the concepts	Familiar	5 - I do it every day	Familiar	No - so we have to build them as needed
177	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	Yes
178	2 - I am aware of the concepts, but not practiced recently	Not Familiar	5 - I do it every day	Familiar	Yes
179	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
180	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - and we don't use them

ID	Familiarity with Catapults	Familiarity with Catapults (binned)	Familiarity with Models	Familiarity with Models (binned)	Are Models available at Work?
181	4 - I am familiar with the concepts and I use them often	Familiar	3 - Occasional	Not Familiar	Yes
182	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
183	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	No - so we have to build them as needed
184	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
185	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	Yes
186	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - and we don't use them
187	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	Yes
188	3 - I am aware of the concepts	Familiar	5 - I do it every day	Familiar	Yes
189	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - and we don't use them
190	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	No - and we don't use them
191	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	No - so we have to build them as needed
192	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
193	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
194					
195	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
196	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - and we don't use them
197	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
198	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
199	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
200	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
201	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
202	4 - I am familiar with the concepts and I use them often	Familiar	5 - I do it every day	Familiar	Yes
203	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
204	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
205	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
206	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
207	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
208	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
209					
210					
211	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	Yes
212	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	Yes
213					
214	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - and we don't use them
215	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
216					
217	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
218	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	Yes
219	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
220	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
221	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
222					
223	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	No - and we don't use them
224	4 - I am familiar with the concepts and I use them often	Familiar	3 - Occasional	Not Familiar	Yes
225	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
226	2 - I am aware of the concepts, but not practiced recently	Not Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
227	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
228	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	No - so we have to build them as needed
229	1 - Minimal	Not Familiar	2 - Minimal	Not Familiar	Yes
230	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
231					
232	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
233	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	Yes
234					
235	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
236	3 - I am aware of the concepts	Familiar	2 - Minimal	Not Familiar	No - and we don't use them
237	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	No - so we have to build them as needed
238	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	Yes
239	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	Yes
240	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
241	2 - I am aware of the concepts, but not practiced recently	Not Familiar	3 - Occasional	Not Familiar	Yes
242	3 - I am aware of the concepts	Familiar	3 - Occasional	Not Familiar	No - so we have to build them as needed
243	3 - I am aware of the concepts	Familiar	5 - I do it every day	Familiar	Yes
244	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	Yes
245	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
246	2 - I am aware of the concepts, but not practiced recently	Not Familiar	4 - Often	Familiar	No - so we have to build them as needed
247	1 - Minimal	Not Familiar	3 - Occasional	Not Familiar	Yes
248	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
249	4 - I am familiar with the concepts and I use them often	Familiar	4 - Often	Familiar	Yes
250					
251	3 - I am aware of the concepts	Familiar	4 - Often	Familiar	Yes
252	1 - Minimal	Not Familiar	4 - Often	Familiar	Yes

ID	Are the models suited for the problem?	What are the requirements like in your organization?	Does your organization do more physical testing or modeling?
1	Yes	4 - Many requirements with some level of validation	3 - About even
2	No	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
3	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
4	No	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
5	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
6	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
7	Yes	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
8	Yes	2 - Some requirements that are not well validated	3 - About even
9	No	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
10	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	4 - More modeling, but some testing is done
11	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
12	Yes	4 - Many requirements with some level of validation	5 - Decisions are made from models only
13	No	2 - Some requirements that are not well validated	3 - About even
14	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
15	Yes	4 - Many requirements with some level of validation	3 - About even
16			
17	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
18			
19	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	4 - More modeling, but some testing is done
20	No	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
21	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
22		3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
23	No	2 - Some requirements that are not well validated	1 - Decisions are made with testing only
24	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
25		1 - Few to no requirements for systems and / or not well validated	1 - Decisions are made with testing only
26	No	4 - Many requirements with some level of validation	3 - About even
27	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
28	Yes	1 - Few to no requirements for systems and / or not well validated	4 - More modeling, but some testing is done
29			
30		5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
31	No		
32	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
33	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
34	No	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
35	No	4 - Many requirements with some level of validation	1 - Decisions are made with testing only
36			
37			
38	No	2 - Some requirements that are not well validated	1 - Decisions are made with testing only
39	No	2 - Some requirements that are not well validated	5 - Decisions are made from models only
40	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	4 - More modeling, but some testing is done
41	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
42	No	3 - Some requirements with some level of validation	3 - About even
43		4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
44	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
45	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
46			
47	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
48	No	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
49	Yes	3 - Some requirements with some level of validation	1 - Decisions are made with testing only
50	Yes	2 - Some requirements that are not well validated	5 - Decisions are made from models only
51			
52	Yes	1 - Few to no requirements for systems and / or not well validated	2 - More testing, but some modeling is done
53	Yes	4 - Many requirements with some level of validation	3 - About even
54	No	3 - Some requirements with some level of validation	4 - More modeling, but some testing is done
55	Yes	4 - Many requirements with some level of validation	3 - About even
56	No	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
57	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
58	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
59			
60			
61	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
62	No	2 - Some requirements that are not well validated	1 - Decisions are made with testing only
63			
64	Yes	1 - Few to no requirements for systems and / or not well validated	2 - More testing, but some modeling is done
65	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
66	Yes	3 - Some requirements with some level of validation	3 - About even
67	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
68			
69			
70	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	1 - Decisions are made with testing only
71	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
72			
73		4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
74	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
75	Yes	4 - Many requirements with some level of validation	3 - About even
76	No	5 - Complete set of requirements for our systems and systematic plans to validate those	1 - Decisions are made with testing only
77	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
78	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
79	Yes	4 - Many requirements with some level of validation	3 - About even
80			
81	No	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
82	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
83	Yes	3 - Some requirements with some level of validation	3 - About even
84	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
85	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
86	Yes	4 - Many requirements with some level of validation	3 - About even
87	Yes	4 - Many requirements with some level of validation	3 - About even
88	Yes	4 - Many requirements with some level of validation	5 - Decisions are made from models only
89	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
90	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done

ID	Are the models suited for the problem?	What are the requirements like in your organization?	Does your organization do more physical testing or modeling?
91	Yes	4 - Many requirements with some level of validation	3 - About even
92	Yes	3 - Some requirements with some level of validation	3 - About even
93	Yes	4 - Many requirements with some level of validation	3 - About even
94	Yes	4 - Many requirements with some level of validation	3 - About even
95	No	5 - Complete set of requirements for our systems and systematic plans to validate those	1 - Decisions are made with testing only
96	No		3 - About even
97	No	4 - Many requirements with some level of validation	3 - About even
98	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
99			
100	No	3 - Some requirements with some level of validation	3 - About even
101	Yes	4 - Many requirements with some level of validation	3 - About even
102		4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
103	Yes	3 - Some requirements with some level of validation	3 - About even
104			
105	No		2 - More testing, but some modeling is done
106			
107	Yes	4 - Many requirements with some level of validation	5 - Decisions are made from models only
108	No	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
109			
110	Yes	2 - Some requirements that are not well validated	3 - About even
111	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
112	Yes	4 - Many requirements with some level of validation	3 - About even
113	Yes	4 - Many requirements with some level of validation	3 - About even
114			
115	Yes	4 - Many requirements with some level of validation	3 - About even
116			
117	Yes	4 - Many requirements with some level of validation	3 - About even
118	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
119	No	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
120	Yes	2 - Some requirements that are not well validated	3 - About even
121	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
122	No	3 - Some requirements with some level of validation	3 - About even
123	Yes	3 - Some requirements with some level of validation	3 - About even
124	Yes	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
125	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
126	Yes	3 - Some requirements with some level of validation	3 - About even
127		5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
128	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
129	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
130	No	1 - Few to no requirements for systems and / or not well validated	2 - More testing, but some modeling is done
131	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
132	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
133	No	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
134	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
135	Yes	4 - Many requirements with some level of validation	3 - About even
136	Yes	4 - Many requirements with some level of validation	3 - About even
137			
138	Yes	3 - Some requirements with some level of validation	3 - About even
139			
140	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
141	Yes	2 - Some requirements that are not well validated	1 - Decisions are made with testing only
142	No	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
143	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
144	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
145	Yes	4 - Many requirements with some level of validation	1 - Decisions are made with testing only
146	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
147	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
148	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
149	Yes	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
150	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
151			
152	No	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
153	Yes	3 - Some requirements with some level of validation	5 - Decisions are made from models only
154	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
155	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
156	No	3 - Some requirements with some level of validation	1 - Decisions are made with testing only
157	Yes		3 - About even
158		2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
159	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
160	Yes	4 - Many requirements with some level of validation	3 - About even
161	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
162	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
163		5 - Complete set of requirements for our systems and systematic plans to validate those	1 - Decisions are made with testing only
164	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
165	Yes	3 - Some requirements with some level of validation	4 - More modeling, but some testing is done
166		4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
167	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
168	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
169	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
170	Yes	3 - Some requirements with some level of validation	4 - More modeling, but some testing is done
171	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
172	Yes	4 - Many requirements with some level of validation	3 - About even
173			
174	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
175	No	4 - Many requirements with some level of validation	1 - Decisions are made with testing only
176	Yes	2 - Some requirements that are not well validated	4 - More modeling, but some testing is done
177	Yes	3 - Some requirements with some level of validation	3 - About even
178	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
179	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
180		2 - Some requirements that are not well validated	

ID	Are the models suited for the problem?	What are the requirements like in your organization?	Does your organization do more physical testing or modeling?
181	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
182	No	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
183	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
184	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
185	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
186	No	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
187	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
188	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
189			
190	No	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
191	Yes	3 - Some requirements with some level of validation	3 - About even
192	No	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
193	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
194			
195	No	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
196		4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
197	Yes	3 - Some requirements with some level of validation	1 - Decisions are made with testing only
198		2 - Some requirements that are not well validated	3 - About even
199	Yes	4 - Many requirements with some level of validation	3 - About even
200	Yes	2 - Some requirements that are not well validated	4 - More modeling, but some testing is done
201	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
202	Yes	4 - Many requirements with some level of validation	3 - About even
203	Yes	3 - Some requirements with some level of validation	3 - About even
204	No	4 - Many requirements with some level of validation	
205	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
206	Yes	4 - Many requirements with some level of validation	3 - About even
207	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
208		5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
209			
210			
211	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
212	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
213			
214			
215	Yes	4 - Many requirements with some level of validation	3 - About even
216			
217	Yes	3 - Some requirements with some level of validation	3 - About even
218	Yes	3 - Some requirements with some level of validation	4 - More modeling, but some testing is done
219	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
220	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
221	No	3 - Some requirements with some level of validation	3 - About even
222			
223	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
224	Yes	4 - Many requirements with some level of validation	3 - About even
225	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	5 - Decisions are made from models only
226	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
227	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
228	No	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
229	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
230	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
231			
232	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	3 - About even
233	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
234			
235	No	3 - Some requirements with some level of validation	4 - More modeling, but some testing is done
236	No	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
237	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
238	Yes	2 - Some requirements that are not well validated	2 - More testing, but some modeling is done
239		4 - Many requirements with some level of validation	3 - About even
240	No	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
241	No	3 - Some requirements with some level of validation	4 - More modeling, but some testing is done
242	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
243	Yes	3 - Some requirements with some level of validation	3 - About even
244	Yes	5 - Complete set of requirements for our systems and systematic plans to validate those	2 - More testing, but some modeling is done
245	Yes	4 - Many requirements with some level of validation	3 - About even
246	Yes	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
247	Yes	4 - Many requirements with some level of validation	2 - More testing, but some modeling is done
248	No	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
249	Yes	4 - Many requirements with some level of validation	4 - More modeling, but some testing is done
250			
251	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done
252	Yes	3 - Some requirements with some level of validation	2 - More testing, but some modeling is done