A Synthesized Methodology for Eliciting Expert Judgment for Addressing Uncertainty in Decision Analysis

Richard W. Monroe
Old Dominion University

A SYNTHESIZED METHODOLOGY FOR ELICITING EXPERT JUDGMENT FOR ADDRESSING UNCERTAINTY IN DECISION ANALYSIS

by

Richard W. Monroe
B.E.T, 1975, Southern Technical Institute
M.S.E.M., 1990, Western New England College

A Dissertation submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

ENGINEERING MANAGEMENT

OLD DOMINION UNIVERSITY
August 1997

Approved by:

Resit Unal (Director)

Laurence D, Richards (Member)

Han P. Bao (Member)

Abel A. Fernandez (Member)

James Schwing (Member)

ABSTRACT

A SYNTHESIZED METHODOLOGY FOR ELICITING EXPERT JUDGMENT FOR ADDRESSING UNCERTAINTY IN DECISION ANALYSIS

Richard W. Monroe
Old Dominion University, 1997
Director: Dr. Resit Unal

This dissertation describes the development, refinement, and demonstration of an expert judgment elicitation methodology. The methodology has been developed by synthesizing the literature across several social science and scientific fields. The foremost consideration in the methodology development has been to incorporate elements that are based on reasonable expectations for the human capabilities of the user, the expert in this case.

Many methodologies exist for eliciting assessments for uncertain events. These are frequently elicited in probability form. This methodology differs by incorporating a qualitative element as a beginning step for the elicitation process. The qualitative assessment is a more reasonable way to begin the task when compared to a subjective probability judgment. The procedure progresses to a quantitative evaluation of the qualitative uncertainty statement. In combination, the qualitative and quantitative assessments serve as information elicited from the expert that is in a subsequent step to develop a data set. The resulting data can be specified as probability distributions for use in a Monte Carlo simulation.

A conceptual design weight estimation problem for a simplified launch vehicle model is used as an initial test case. Additional refinements to the methodology are made

as the result of this test case and as the result of ongoing feedback from the expert. The refined methodology is demonstrated for a more complex full size launch vehicle model.

The results of the full size launch vehicle model suggest that the methodology is a practical and useful approach for addressing uncertainty in decision analysis. As presented here, the methodology is well-suited for a decision domain that encompasses the conceptual design of a complex system. The generic nature of the methodology makes it readily adaptable to other decision domains.

A follow-up evaluation is conducted utilizing multiple experts which serves as a validation of the methodology. The results of the follow-up evaluation suggest that the methodology is useful and that there is consistency and external validity in the definitions and methodology features.

Acknowledgements

My grateful appreciation to my dissertation committee - Resit Unal, Laurence Richards, Han Bao, Abel Fernandez and Jim Schwing. My appreciation is also extended to other faculty members who have been very supportive of my doctoral studies.

My sincere thanks for the endless cooperation and patience of Roger Lepsch. My gratitude to Larry Rowell for the encouragement and support of this research. And my heartfelt thanks for the aerospace/aircraft expert participants and for the helpful feedback from Jim Wilder.

Thanks to John Aquirre and Jim Schwing for programming assistance and input. A special thanks to John for assisting with all the computer problems and errors (of my own-making and otherwise).

For their friendship:

Bela, Curt, Shardul, Soundar, Murat, Aashish, Cenk, Pradeep, Murat

And especially:

Shelby, Elke, Marla, Otto and my favorite antagonist, Peggy.

I dedicate this dissertation to my best friend in the world, who has the patience of a saint, my loving wife, Christine. I also dedicate this dissertation to the loving memory of my mother, Margie.

Table of Contents

Table of Contents

List of Tables

List of Figures

Chapter I

INTRODUCTION

1.1 Conceptual Design

ab-stract, n. thought of apart from concrete realities or specific objects.
ab-strac'tion, n. an abstract idea or term. (Webster's Dictionary)

Nearly 70% of a system's life cycle cost is determined during the conceptual design stage (Fabrycky and Blanchard 1991). This makes conceptual or preliminary design a critically important developmental phase of the system's life. Conceptual design is the earliest stage of design and at this point the design domain is large and complex (Dym 1994). Uncertainty is naturally inherent in this situation due to the number of design choices and the complexity of design choices that can be conceived.

Conceptual design engineering (CDE) attempts to work from the abstract to the concrete. At this phase of design, the engineers attempt to estimate actual physical attributes of a complex system working only from somewhat abstract conceptual information. The amount of uncertain information is significant when attempting to bridge the gap from the abstract concept to a concrete physical design, especially for complex systems.

Conceptual design engineering can also be thought of as the first "theory" of a complex system. The values of design parameters that are used to develop that initial "theory" are hypotheses" about the design. Usually the only way to test those hypotheses

---

The journal, *Management Science*, has been used as a model for this document's format.

conclusively is to build the design, either as a solid model or a prototype product. For complex systems, testing the hypotheses does not immediately follow hypotheses development and complexity may actually preclude testing. Further developmental stages intercede with multiple decision mileposts along the way. The additional developmental stages and decision points result in new hypotheses and a "new theory" of the design. These additional stages and decision points are confounding factors and are an indication of the complexity and uncertainty associated with conceptual design of complex systems.

Many conceptual design problems are unique. Experts in conceptual design engineering are extremely rare and specialized. Many conceptual design environments are characterized by one-of-a-kind designs. Shipbuilding, aircraft and aerospace are prominent industries that frequently develop one-of-a-kind products. In these environments, each conceptual design engineer develops his/her own estimation models to arrive at desired estimates. One engineer may place a greater emphasis on one piece of data while a second engineer may place a greater emphasis on some other data or test result. These models range from relatively simple models to extremely complex models. Developing models for one-of-a-kind products with little or no historical data again describes a decision environment that is characteristically uncertain and challenging.

How can uncertainty be incorporated in the CDE process? Why bother? If a point estimate is provided by CDE and that estimate is used as 100% certain then the likelihood of being unrealistic is great. Uncertainty should be addressed to provide a more robust methodology for CDE. This is advocated for all CDE complex system design problems and is a very appropriate philosophy to follow in aerospace design.

Quinn and Walsh (1994) list "Project Design: NASA uses unrealistic schedule and funding" at the top of their list of 42 identifiable factors that resulted in the Hubble Space Telescope fiasco. The Hubble case and grossly underestimated resource requirements for many aerospace projects are examples from within NASA's decision domain that provide strong arguments for new methodologies and for more in-depth analyses of complex aerospace design projects.

Public policy analysis is another example of questions that frequently have little available data. When the quality and quantity of data is lacking for such questions, the associated uncertainties need to be assessed in some manner to aid the policy making process (Mullin 1986). Decisions in conceptual design engineering are analogous to policy making. Every decision becomes a policy. In conceptual design, every parameter value estimate becomes that parameter's value as a policy.

1.2 Research Summary

Decision makers are faced with uncertainty in many decision situations characterized by various sources of uncertainty. Researchers have developed several methodologies to elicit evaluations of uncertainty from decision makers in many domains. These methods have aimed at understanding uncertainties and aiding the decision maker to address the uncertainty in a systematic manner. Much of the research has attempted to understand the process by which a decision maker assesses uncertainty. This is akin to understanding the pure thought process.

Unfortunately, much of the prior research has been conducted in a laboratory setting with subjects and decision topics that raise questions as to the generalizability of the resulting research findings. Even more troubling is the fact that a majority of this laboratory research has shown poor performance for the "experts" being studied (Christensen-Szalanski and Beach 1984; Shanteau 1987, 1992; Mullin 1986).

Given the results of prior studies and other limitations, an expert judgment methodology seems like a fruitful research avenue. There is ample room for improvements and a clear need for research successes. To proceed requires careful filtering through the laboratory research to find the heuristics and other techniques that have been shown to be effective with naive and real experts alike. Heeding the warnings of other researchers is also crucial. Drawing upon the positive outcomes and the methods that led to those outcomes can allow the researcher to synthesize a useful methodology from the expert judgment literature. This is the approach taken in this research in an effort to address uncertainty in a particular decision making setting.

Launch vehicle conceptual design is the decision domain of interest. Launch vehicle conceptual design characteristically involves uncertainty due to a lack of historical data and uncertain requirements. Many estimation models in the literature typically rely upon historical data and frequently utilize a regression model. The lack of data makes regression analysis, risk analysis or any other traditional statistical techniques difficult. A supplemental technique is needed to develop a data set for analysis.

This research synthesizes an expert judgment methodology from the literature in order to elicit the expert's judgment of uncertainty in this decision domain. The uncertainty judgments are used by the expert in a multi-stage procedure administered via questionnaire to obtain the data set for analysis. A detailed description is presented for the methodology along with the rationale for each element contained in the methodology.

This research contributes to the expert judgment elicitation literature by developing a new synthesized methodology. Specifically, this methodology differs from other methodologies by incorporating a qualitative assessment as a starting point. The methodology does not elicit preferences, probabilities or utility functions. The absence of those types of elicitations is a significant difference from most of the methodologies in the literature. The documentation elements of the methodology are described in detail and serve as a model for practitioners and for future research.

In addition, this study addresses a real problem in an applied engineering setting and utilizes an actual domain expert. Addressing an applied setting problem is a contribution since the bulk of the literature has addressed experiments conducted in a "laboratory" setting.

Two cases are studied utilizing the methodology. An initial study addresses a simplified weight estimating relationship (WER) model for a launch vehicle. This provides feedback about the methodology and led to further refinement of methodology elements. A second example case is studied utilizing the refined methodology. This second case is a detailed full size WER model for a launch vehicle. The data generated for this second case is used to conduct risk analysis utilizing Monte Carlo simulation. Outcomes, outputs and potential practical uses of simulation results are presented and discussed.

Further statistical evaluation is performed for system parameters for the Monte Carlo simulation procedure. The aim is to optimize the simulation procedure parameters to assure that the simulation is efficient and effective for use as a conceptual design analysis tool. The results of the research suggest that the methodology developed is a versatile technique that can be an effective tool for addressing uncertainty related to complex system design where there is a lack of data.

A follow-up evaluation of the methodology is also conducted utilizing multiple engineering design experts to complete an abbreviated version of the questionnaire along with a set of benchmark questions. This serves as a final validation of the methodology.

In the following section, the literature for several pertinent topics are reviewed. Decision making under uncertainty is a broad field that applies in this uncertain setting. Risk analysis is a normative methodology for dealing with uncertainty and arriving at specific representations of outcomes. Expert judgment is used in a variety of situations when data is scarce and when uncertainty is present. Each of these fields contribute to the foundation of the methodology developed in this research.

Chapter II

LITERATURE REVIEW

2.1 Decision Making Under Uncertainty

A classic definition of decisions under uncertainty is offered by Scholz (1983) as those circumstances characterized by "incomplete information or [incomplete] knowledge about a situation, i.e. the possible alternatives, or the probability of their occurrence, or their outcomes, are not known by the subjects" (Scholz 1983, p. 4). March identifies three important sources of uncertainty that face decision makers - "an inherently unpredictable world, incomplete knowledge about the world, and incomplete contracting with strategic actors" (March 1994, p. 36).

Since the formalization of decision analysis, uncertainty has been an important issue that has garnered significant attention. "Decision making under uncertainty has been dominated by a single approach - the closely related theories of expected utility and subjective expected utility. As formulated and axiomatized by von Neumann and Morgenstern (1944) and Savage (1954), these theories rank among the most important in twentieth-century social science" (Einhorn and Hogarth 1986). These theories have greatly influenced the social scientists' characterization of decisions under uncertainty and serve as the "foundation for prescriptive approaches to decision making (e.g. Raiffa 1968; Keeney and Raiffa 1976)" (Einhorn and Hogarth 1986).

When faced with uncertainty, decision makers can choose to ignore uncertainty or choose to deal with uncertainty explicitly. Assuming the latter choice is made, researchers have developed several methodologies to elicit evaluations of uncertainty from decision makers in many domains. These methods have aimed at understanding uncertainties and aiding the decision maker to address the uncertainty in a systematic manner. Much of the research has attempted to understand the process by which a decision maker assesses uncertainty.

If the decision maker chooses to deal with uncertainties in some manner then some systematic approach is needed. Researchers must aid the decision maker when designing methodologies to address these uncertain decision situations.

Several approaches for dealing with uncertainty have been developed. One can ignore the existence of uncertainty (Hogarth 1975; MacCrimmon and Taylor 1976) or use one of the other approaches that have their own limitations (Hertz and Thomas 1983). Morgan and Henrion (1990) offer several arguments for addressing uncertainty rather than ignoring it. By way of analogy, natural scientists routinely report some estimate of error in their quantitative measures (Morgan and Henrion 1990). Typical uncertainties in quantitative policy analysis are larger than errors or uncertainties in natural science fields (Morgan and Henrion 1990). Based on this difference in magnitude, "policy analysts should report their uncertainties too" (Morgan and Henrion 1990). Additional substantive arguments are that:

- Explicit treatment of uncertainty forces additional and careful thought about the "important factors" in an analysis and "sources of disagreement in a problem."
- Increased reliance on experts to assist decision making may leave the decision maker confused about what experts say. Asking experts to document "the uncertainty of their judgments" will clarify their recommendations, tell us the basis for their recommendations and tell us if experts disagree with each other.
- Documentation of uncertainties for one problem may be useful as information and/or serve as a methodology template for addressing future similar problems. Careful documentation of uncertainties will give us "greater confidence that we are using the earlier work in an appropriate way" (Morgan and Henrion 1990, p. 3).

These are convincing arguments for including uncertainty rather than ignoring it. The documentation and methodological arguments are particular appealing when addressing uncertainty in an applied setting. Documentation and development of a methodology will serve practitioners well when addressing future similar problems.

Among the variety of methods that have evolved for addressing uncertainty are multiattribute utility theory (MAUT) (Keeney 1977), expected utility (von Neumann and Morgenstern 1944) and subjective expected utility (SEU) (Savage 1954). One primary drawback of these approaches is the formulation of problems in strictly economic terms. These methods also require the choice of one alternative versus another alternative. Another difficulty with these approaches is the need for complementary outcomes. Some problems do not lend themselves to direct economic utility measurement and some do not lend themselves to the expression of complementary outcomes. In general, some decisions are not made with an objective of maximizing some utility function.

One method that begins to incorporate risk in the decision making process is the specification of the estimates at high, medium and low values. Typically these are specified at pessimistic, most likely and optimistic levels for the factor and the outcomes are simple summations of the variables at the three levels (Hertz and Thomas 1983). Hertz and Thomas (1983) believe that this is a step in the right direction but that it still does not provide a thorough method for comparing alternatives. They advocate a method that is used explicitly to address uncertainty in a variety of decision domains - risk analysis (Hertz and Thomas 1983).

2.2 Risk Analysis

Risk is defined as "both uncertainty and the results of uncertainty" (Hertz and Thomas 1983). In other words, "risk refers to a lack of predictability about structure, outcomes or consequences in a decision or planning situation" (Hertz and Thomas 1983). March's sources of uncertainty - "an inherently unpredictable world, incomplete knowledge about the world, and incomplete contracting with strategic actors" (March 1994, p. 36)- are characteristically the sources of risk in many decision making situations.

Risk can also be viewed as either objective or subjective. Objective risk is based strictly on probabilities of events such as flipping a coin, rolling dice or similar acts involving chance. In engineering design, objective risk is rarely encountered. Subjective risk probabilities cannot be determined experimentally (Lapin 1982) since they are tied to human judgment where further information would alter the person's assessment.

Subjective risk is logically of interest in many conceptual design problems since human judgment is an integral part of design parameter specification for highly uncertain complex systems. Subjective risk is also inherent in decisions about technologies to be used in complex system design. This is more typical of engineering risk situations especially for engineering conceptual design.

Morgan and Henrion (1990) suggest that asking experts for their "best professional judgment" is sometimes the only option when faced with a situation that has limited data or is not fully understood. Fischhoff (1989) asserts that very little research has been conducted on the "judgmental processes in risk analysis" and proceeds to extrapolate from other applied settings of expert judgment. His work offers some guidance and a framework for a generic risk analysis structure utilizing the elicitation of expert knowledge but does not include an example application. Some of the previous research on expert judgment is discussed in the following section.

2.3 Expert Judgment

Expert judgment methods utilize recognized or identifiable expert(s) in a given domain to provide an informed judgment about some variable of interest or about some decision criteria. The techniques are particularly effective in decision domains that are narrow and are more effective in applied settings (Beach 1975; Ettenson, Shanteau, and Krogstad 1987) and particularly in settings where the expert is providing judgments about physical stimuli (Shanteau 1992; 1987).

Unfortunately, much of the research has been conducted in a laboratory setting utilizing naive subjects or non-experts and addressing trivial or unrealistic decisions (Mullin 1986). The setting, subjects and decision topics raise questions as to the generalizability of the resulting research findings. The fact that the laboratory research has shown poor performance for the "experts" being studied also raises concerns.

Christensen-Szalanski and Beach (1984) offer evidence that articles that describe "poor" expert performance were cited six times more frequently in their ten year study period than were articles describing "good" expert performance. This phenomenon referred to as the "citation bias" has led to the characterization that when it comes to human judgment, "people are no damn good" (Edwards 1992). Of course this is a biased interpretation of the literature and not the viewpoint of the majority of the researchers that continue to do research in expert judgment including Edwards (1992).

The frequently cited "poor" expert judgment literature suggests that it is a method laden with pitfalls. Several researchers have found that people do poorly when asked to give an expert judgment in probability form (Tversky and Kahneman 1973; Kahneman and Tversky 1972; Morgan and Henrion 1990). Others have found that people fare better when asked for upper and lower bounds around a midpoint than when asked for probabilities (Spetzler and Stael von Holstein 1975; Beach 1975). Qualitative assessments of uncertainty have also been shown as easier to elicit than are quantitative ones (Zimmer 1983; Budescu and Wallsten 1987; Wallsten, Budescu, Rapoport, Zwick and Forsyth 1986; Lichtenstein and Newman 1967) although agreement on the meaning of verbal descriptions of uncertainty may be lacking in some instances (Lichtenstein and Newman 1967).

There are also numerous biases that must be taken into consideration when seeking an expert's judgment (Spetzler and Stael von Holstein 1975). Using a heuristic that challenges the expert to support his/her reasoning has been helpful in overcoming many of these biases. In the course of eliciting an expert judgment, certain heuristics have been shown to achieve better results than others. In particular, effective heuristics include instructional materials that guide the expert to remove additional bias. Hoch (1984) found that judgments were noticeably influenced when experts were asked for a reason for their judgment. By asking for reasons, the judgment is debiased (Hoch 1984; Morgan and Henrion 1990). Mullin (1986) requests that experts describe scenarios that may lead to adjusting their judgments. Cautioning experts about anchoring and asking for alternative scenarios are simple steps to take that Mullin (1986) believes are useful no matter the direction of the bias (overconfident or underconfident).

Mental simulation is a useful heuristic but again is subject to bias (Kahneman and Tversky 1982). As a subject mentally simulates a situation, variables are changed in a downhill, uphill or horizontal fashion. The most frequent bias tends to be in a downhill direction with a low percentage of subjects selecting uphill or horizontal changes. Mental simulation is subject to large and systematic errors due to downhill bias (Kahneman and Tversky 1982).

Some of the expert judgment techniques that have been used extensively include the Delphi method (Dalkey 1969; Lock 1987), the Nominal Group Technique (NGT) (Van de Ven and Delbecq 1971; Lock 1987) and brainstorming (Lock 1987). Each of these involves elicitation of judgments from a group of experts through questionnaires and typically are accomplished from a distance (e.g. Delphi) or by bringing the group of experts together in one meeting (e.g. NGT and brainstorming).

Mullin (1986) discussed the problems that are associated with combining multiple experts or averaging a group of experts. She suggests that combining experts' judgments depends on how different their estimates are (Mullin 1986). If the same models are used and the experts produce relatively consistent results then combining the experts' assessments may be an acceptable practice (Mullin 1986). At the other end of the spectrum, if there is significant disagreement between experts then the analysis will not be well-served by combining (or averaging) the experts' judgments (Mullin 1986). Mullin (1986) submits that the Delphi method is one approach for trying to reach group consensus among a group of experts.

One telling observation about group techniques comes from Parente and Anderson-Parente (1987), who suggest that the Delphi technique was never meant to be used as a scientific technique. Delphi was developed to elicit judgments or opinions about topics that were not easily analyzed with normative scientific techniques (Parente and Anderson-Parente 1987). The primary benefit of Delphi is the collection of diverse viewpoints that is made possible by avoiding the "face-to-face format" where opinions may be withheld or dominated by a few individuals (Parente and Anderson-Parente 1987). This commentary serves as a strong warning when considering the Delphi technique as a means for dealing with multiple experts.

Expert calibration is often used when Bayesian methods are employed to combine expert opinions (Mullin 1986). The analyst usually adopts an axiomatic or modeling approach to Bayesian aggregation of probabilities (Winkler 1986; Mullin 1986). The axiomatic approach sticks to rigid rules of combination and does not account for new information that may be obtained by any one or several of the experts (Mullin 1986). The modeling approach treats the experts' probabilities as information and this information is aggregated into resulting likelihoods (Mullin 1986). This approach is classically Bayesian with a multiplicative relationship between the prior distribution and the likelihood function (Mullin 1986). The primary difficulty with these techniques is the large number of subjective judgments that are required (Clemen 1986; French 1986; Mullin 1986). These subjective shortcomings apply to the experts and to the analyst as well. The analyst must use subjective judgment in judging suitable calibration, information dependence between experts and in combining the experts' judgments (Mullin 1986).

These techniques are ideally suited (or at least useful) for decision topics where a large group of experts can be readily identified and where the group of experts is readily accessible. However, many decision topics are extremely narrow and preclude the use of these techniques since a group of experts cannot be identified, does not exist or is not easily accessed. The distribution of expertise is typically skewed with the greatest expertise residing with one or two experts within a given decision domain (Augustine 1979; Turban 1992).

In these instances, when the decision domain is extremely narrow, an expert judgment technique may be needed that utilizes the judgment of a single expert. This is often the case for the development of expert systems (Turban 1992). One useful guideline for determining expertise is that an "individual should not be considered an expert unless he or she is knowledgeable at the level of detail being elicited" (Meyer and Booker 1991, p. 85).

2.4 Expertise

Expertise is not limited to pure knowledge on a given topic, expertise encompasses additional skills that exhibit the full range of an expert's knowledge. Additional abilities include explaining results, learning new things about the domain, restructuring knowledge when warranted, knowing the exceptions to the rules and determining the appropriateness of one's own expertise (Turban 1992, p. 80). These additional characteristics separate the true expert from the non-expert. These characteristics allow the expert to demonstrate his/her expertise by applying it in an appropriate manner and by reformulating the knowledge or the problem in order to best apply his/her expertise.

Another perspective of expertise concerns the substantive and normative components of expertise. The expert's experience and knowledge about the topic constitute substantive expertise (Beach 1975; Meyer and Booker 1991). The expert's experience and knowledge about "the use of the response mode" constitute normative expertise (Meyer and Booker 1991). The response mode refers to the form in which the expert's knowledge is elicited (e.g. probabilities, preferences, utility functions, pairwise comparisons, etc). Thus, normative expertise refers to the expert's knowledge of statistical and mathematical principles that may relate directly to the form in which the judgment is given (Meyer and Booker 1991). Using individuals with strength in neither substantive nor normative is unwise and will likely not be very useful. Using individuals with strength in only one of the two is an improvement but will still result in substandard outcomes. Hogarth (1975) attributes many of the problems with expert judgment studies to precisely these two conditions - individuals with neither normative nor substantive expertise or individuals with expertise in only one of these categories.

Some techniques employ calibration (Cooke, Mendel and Thys 1988; Bhola, Cooke, Blaauw and Kok 1992) as an integral element of an elicitation methodology. This would be consistent with the above observation. That is, an expert with substantive expertise can be trained to develop the required normative expertise to make the elicited judgment more meaningful.

Shanteau (1992) reached some revealing conclusions in his review of the expert judgment literature. He concluded that where poor expert performance was observed, the situations were dynamic and generally involved human behavior. Poor performing experts included: clinical psychologists, psychiatrists, court judges, parole officers and personnel managers (Shanteau 1992). Good expert performance was generally associated with static objects or things (Shanteau 1987). Dawes (1987) contrasted the two by noting that "human behavior is inherently less predictable than physical stimuli" (Shanteau 1992). Examples of domain experts with competent performance included: astronomers, livestock judges, soil judges, test pilots, physicists, mathematicians and accountants (Shanteau 1992). Again, Hogarth (1975) would offer that these experts exhibit both substantive and normative expertise.

Shanteau has identified five factors associated with the competence of experts: "domain knowledge, psychological traits, cognitive skills, decision strategies, and task characteristics" (Shanteau 1992, p. 263). Assuming the first four factors are satisfied at an appropriate level, the task characteristics are the variable in expert judgment research that afford the researcher some degree of control. In other words, the researcher can design the tasks to be administered to the expert to best draw upon the subject's expertise. Shanteau (1992) goes on to suggest that expert performance cannot be seen as all good or all bad. The same expert may perform well in one setting but perform poorly in another setting. "Their competence depends on the task characteristics" (Shanteau 1992).

These observations by Shanteau (1992) and Dawes (1987) serve to steer researchers towards physical stimuli topics rather than behavioral stimuli topics. Shanteau (1992) hypothesizes that "the more a task contains [physical] characteristics, the greater the competence that should be seen in experts" (Shanteau 1992, p.261). And the more a task contains human behavioral characteristics, "the lesser the competence expected in experts" (Shanteau 1992, p.261). The subjective assessments in this research are associated with a physical object - a launch vehicle. The findings that suggest that subjective judgments of physical stimuli are more frequently competent judgments is a favorable indication that this is an appropriate topic to address with expert judgment. A suitable domain expert should be able to competently supply a meaningful subjective judgment of uncertainty about characteristics of the physical stimuli in this research study. In other words, the fact that the subject being analyzed in this study is a physical design is a favorable condition for competent subjective assessments by an expert as suggested by the work of Shanteau (1992; 1987).

Table 1 summarizes the results of the expert judgment literature study:

Table 1 Summary of Expert Judgment Literature

Author(s)	Findings	Guidance Drawn
Christensen-Szalanski and Beach, 1984	Poor performance articles cited 6 times more frequently during 10 year period studied.	Look for good performance articles and do not fall victim to the citation bias.
Hoch, 1984	Judgments influenced when experts asked for reasons.	Request reasons for judgments as an integral part of methodology.
Mullin, 1986	Scenarios may lead experts to adjust their assessments.	Request scenarios as an integral part of the methodology.
Spetzler and Stael von Holstein, 1975	Elicitation method depends on the quantity and the importance to the decision.	Lower and upper bounds are easier to elicit than probabilities.
Lichtenstein and Newman, 1967	Consistency in most quantities but small number of responses that rated ambiguous phrases or recognized phrases as complements.	Use fewer verbal phrases to describe uncertainty in research.
Budescu and Wallsten, 1987	Consistency in numerical assignments to verbal phrases.	Use fewer verbal phrases.
Wallsten, Budescu, Rapoport, Zwick and Forsyth, 1986	Good monotonic consistency among subjects when using 6 and 10 verbal phrases for ambiguous quantities.	Supports the use of fewer verbal phrases.
Bolger and Wright, 1992	Use percentages rather than odds or probabilities and encourage judges to decompose the problem in their own way.	Use percentages to quantify verbal phrases - use existing decomposition of the problem.
Shanteau, 1987; 1992	Poor performance associated with behavioral stimuli and good performance associated with physical stimuli.	More confidence can be expressed in judgments of physical stimuli assuming appropriate tasks are designed.

2.5 Previous Studies and Their Findings

One published study in an applied engineering setting used expert highway engineers to evaluate their problem solving strategies (Hammond, Hamm, Grassia, and Pearson 1987). The expert highway engineers were asked to evaluate the "roads' aesthetics ..., predict the ... accident rate ... and estimate the roads' carrying capacity ..." (Hoffman, Shadbolt, Burton, and Klein 1995). Stimuli for these judgments were "slides showing different views of roadways or a bar graph depicting a number of road variables" (Hoffman, et al. 1995). Their study sought to determine if different combinations of materials and different task characteristics invoked intuitive reasoning or analytical reasoning (e.g., slides were necessary for aesthetic judgments and invoked intuitive rather than analytical reasoning; analytical reasoning was logically triggered by bar graphs of road variables) (Hammond, et al. 1987; Hoffman, et al. 1995). An additional finding was that there was no deterioration in expert performance when comparing intuitive and analytical reasoning (Hammond, et al. 1987; Hoffman, et al. 1995). One generalization that may be drawn from this study is that experts are likely to use some combination of intuition and analytical techniques in the course of making an assessment.

Mullin (1989) also did research on knowledge elicitation from engineers. Her research utilized three groundwater engineers and three electrical engineers from the faculty of the respective departments of Civil Engineering and Electrical and Computer Engineering at Carnegie-Mellon University. One electrical field problem and one groundwater problem were given to the group of six "experts" and they were asked to provide a solution. The electrical engineers served as the expert on the electrical field problem and the civil engineers served as novices on the same problem. The roles were reversed for the ground water problem. The problems were realistic problems but the research was clearly not undertaken in a working engineering setting and did not deal with a problem that the engineers currently faced. One interesting result was that the engineers developed their own model to solve each of the problems. For the engineers that interpreted the problem correctly, the models were very similar. Erroneous assumptions resulted in models that were different from the group and were not valid solutions.

Pate-Cornell and Fischbeck (1994) performed a risk analysis for thermal protection system (TPS) tiles on the space shuttle. Primary risks during reentry were identified as debonding of tiles, loss of adjacent tiles following the first tile lost, burn-through and failure of a critical subsystem. Tiles were assessed in two phases, first the susceptibility of the tiles to damage from debris at liftoff was evaluated then the effect of the damage on shuttle performance was evaluated (Pate-Cornell and Fischbeck ,1994). Included among the assessments was the utilization of subjective probabilities (e.g. for critical subsystem failures if a burn-through occurred) that were based on expert opinion (Pate-Cornell and Fischbeck ,1994).

Pate-Cornell and Fischbeck's study (1994) focused on safety issues related to just one shuttle subsystem, the thermal protection system (TPS). The TPS consists of different design components - protective blankets in the areas of lower heat loads (primarily the top of the shuttle) and reinforced carbon-carbon tiles in the areas of highest heat loads (the nose and wing edges). Tiles are silicate blocks covered with black glazing and are approximately 8"x8"x2" in size (Pate-Cornell and Fischbeck

,1994). They are bonded to a felt strain isolation pad (SIP) which is in turn bonded to the shuttle's aluminum skin. A room-temperature vulcanized (RTV) material is used as the bonding agent (Pate-Cornell and Fischbeck ,1994). Gaps are designed into the TPS to permit system flexibility and to vent gases during liftoff and the ascent (Pate-Cornell and Fischbeck 1994). Some small gaps are left empty while larger gaps are filled with gap fillers. The surface must be relatively smooth to prevent unnecessary turbulence during reentry (Pate-Cornell and Fischbeck ,1994). Matching tiles and fillers is a tedious and very critical process that requires extensive maintenance time on a periodic schedule in between flights (Pate-Cornell and Fischbeck 1994).

The complexity and variability of the TPS subsystem design provide an excellent example of the difficult task that faces the conceptual design engineer. The safety issues that extend to loss of vehicle and loss of crew underscore the importance of design plans and design decisions for this subsystem (and many others). The Pate-Cornell and Fischbeck (1994) study determined that the TPS was highly susceptible to operating conditions (e.g. debris damage during liftoff) and to organizational issues (e.g. lower pay rates for tile technicians, high turnover rates for tile technicians). These issues also highlight the types of variation that occur during construction and operation that exacerbate the uncertainty of weight estimation and other design estimates at the conceptual design stage. An overly ambitious weight reduction plan may be thwarted by subsequent decisions or by assembly technicians that build the vehicle to their own design. Estimating the weight at a higher, more conservative value will not be accepted as realistic and will be frowned upon due to the associated increase in cost.

From a historical viewpoint, Pate-Cornell and Fischbeck (1994) note that probabilistic risk analysis (PRA) had not been used at NASA since the early 1960's when a consultant using PRA said there was a small probability of success for NASA's mission to the moon. The Challenger accident in January 1986 prompted NASA to reconsider the vulnerability of the space shuttle program (Pate-Cornell and Fischbeck 1994). Subsequently, in 1987 NASA began to again utilize PRA (Pate-Cornell and Fischbeck 1994). Ten years later there are numerous examples of PRA studies conducted at NASA (e.g. risk of UV radiation, etc.) and several request for proposals (RFPs) on the list of topics currently among NASA research agendas.

Other articles that address expert judgment in realistic settings that have influenced this research are summarized in Table 2:

Table 2 Summary of Expert Judgment in Realistic Settings

Author(s)	Research Subjects	Topic/Findings
Pate-Cornell and Fischbeck, 1994	NASA directors, engineers and technicians	Schedule for replacing critical space shuttle TPS tiles.
Hammond, Hamm, Grassia, and Pearson. 1987	Colorado highway engineers	Intuitive and analytical expert judgments were comparably accurate.
Mullin, 1989	Engineering Faculty	1 Groundwater problem and 1 electrical field problem; experts developed their own models.
Ettenson, Shanteau and Krogstad, 1987	Professional Accounting Auditors	Use of primary cues and secondary cues - information use.
Phelps and Shanteau, 1978	Livestock Judges	Judges integrate many dimensions in their judgments but intercorrelations reduce the total number.
Beach, B.H., 1975	Literature review of other studies of experts in medicine, meteorology, military and business (stock market analysts).	Use of subjective probabilities and Bayes Theorem is potentially profitable. Much more research needs to be done in realistic settings.

2.6 Summary

This chapter introduced some of the concepts related to decision making under uncertainty and risk analysis. The chapter also reviewed the broader literature on expert judgment and reviewed several published studies of expert judgment in applied or realistic settings. Most of the methodological elements in this research have been drawn from the literature presented in this chapter. Many of these elements can be seen in the two summary tables included in this chapter, Table 1 and Table 2. Of particular importance to the methodology, lower and upper bounds around the point estimate (Spetzler and Stael von Holstein 1975), reasons (Hoch 1984), scenarios (Mullin 1986) and the use of few verbal phrases (Wallsten, et al. 1986; Lichtenstein and Newman 1967) have been drawn from the literature presented.

The conclusion from the literature review was that no single method has been shown to be an overwhelming favorite when working with expert judgment. Most of the authors referenced above suggested that multiple techniques are needed to debias expert's assessments. In the literature, there was, however, a heavy reliance on probabilistic assessments.

One applied engineering setting study (Hammond, et al. 1987) has more in common with the "laboratory" studies that focus on questions using almanac type data. The analysis was performed on existing roadways. Mullin (1989) did employ engineers in her research but did not deal with a real world problem with the degree of complexity that is involved with CDE. The approach taken in this research and the problem domain being addressed appears to be unique compared to previous studies.

Chapter III

RESEARCH CONTEXT

3.1 Why Weight is Important

> *Engineers don't think about what their [designs] weigh. The U.S.S. Latey [ship] weighs 40,000 tons! How much does the Empire State Building weigh? If we don't know what it weighs, we don't know the performance we're getting for our investment. ... We have to do more with less.*
>
> *R. Buckminster Fuller*
> *(PBS, April 10, 1996)*

Fuller (1996) advocated focusing on weight reduction in all engineering design especially housing. His advocacy was a lifetime crusade (1895-1983) that touched a broad spectrum of design problems and he frequently advocated the use of technologies borrowed from the aerospace industry (e.g. the Wichita house, 1996).

Weight has received significant attention in vehicle and vessel design and has been an issue for as long as those engineering fields have existed. The concern for weight crosses the design domains of automobiles, sailing ships, watercraft, aircraft and space vehicles. The emphasis on weight in aircraft design and development is reflected in the following quotes:

> "It is an analytical fact that aircraft/rotorcraft performance is even more sensitive to weight than other important parameters such as lift-to-drag ratio and specific fuel consumption" (Scott 1992, p.2).
>
> "Weight was the most important development problem...[leading to a canceled program]" (Aviation Week and Space Technology, June 17, 1991; quoted in Scott 1992).
>
> "More airplanes have failed due to being overweight than for any other single cause" (Richard Gathers, aircraft designer for 51 years; quoted in Scott 1992).

> "The most important contributor to avoiding contractor-responsible weight growth is a realistic estimate" (Robert Anderson, USAF, WPAFB responding to questionnaire at the October 1991, SAWE Weight Growth Workshop in St. Louis; quoted in Scott 1992).

> "First and foremost, push for realistic weight estimates..." (NAVAIR response to questionnaire at the October, 1991, SAWE Weight Growth Workshop in St. Louis; quoted in Scott 1992).

The importance of weight was shown statistically by Gordon (1988) when using a regression procedure to estimate aircraft cost. His results indicated a coefficient of correlation (r) of 0.979 for "weight" as a predictor of "cost" (Gordon 1988). This proved to be a slightly stronger correlation than either "area" (r=0.952) or "volume" (r=0.927) (Gordon 1988). Weight also had the lowest percentage standard error of the three variables, 0.5 versus 3.6 and 8.2 respectively (Gordon 1988).

From these comments and studies, weight is posited as a critical factor affecting aircraft performance and, more importantly for this research, affecting the success of design and development programs. The comments indicate that there is a history of problems associated with weight estimates that have led to canceled design programs and failed designs.

Aerospace conceptual design engineers frequently perform spacecraft/launch vehicle design studies and weight optimization is used as a criteria in these studies (e.g. Bush, Unal, Rowell and Rehder 1992; Stanley, Unal and Joyner 1992; Stanley, Engelund, Lepsch, McMillian, Wurster, Powell, Guinta, and Unal 1993; Engelund, Stanley, Lepsch, McMillian and Unal 1993). Weight optimization is also a criteria in aircraft design studies (Wille 1990). From this emphasis, weight is viewed as an important factor that affects launch vehicle performance and possibly life cycle cost of the launch vehicle.

At the conceptual design stage, optimization may be a lofty goal given the amount of uncertainty involved. Because this research focuses on conceptual design, an optimization criteria has not been adopted. The amount of uncertainty dictates that a stochastic methodology is more appropriate than a specific optimization technique.

In addition to physical performance of the finished design there are other significant performance measures for design and development programs in the form of cost and schedule metrics. In launch vehicle conceptual design the latter performance measures are primary concerns along with satisfaction of mission performance requirements. The solution advocated in this small sample of quotes is to strive for more realistic weight estimates. The same suggestions (i.e. realistic estimates) that apply to aircraft CDE can apply to aerospace CDE and virtually any CDE dealing with the design of a complex system (i.e. push for realistic estimates).

3.2 Specifics of the Domain

Weight estimating is a critical task at conceptual design for a launch vehicle. Weight estimates are used to make management decisions in choosing among alternative designs (e.g. lower weight may mean increased performance and in some cases lower life cycle cost). Weight estimates are also important factors used for estimating cost. Typically, weight estimating relationships (WERs) developed and scaled from historical data of aircraft are used to estimate weight of the various subsystems of launch vehicles at the conceptual design phase. Since there is little historical data, these WERs are highly uncertain. The risk of under- or over-estimating launch vehicle weight is a primary concern associated with uncertainty inherent in the WERs. When weight is under- or over-estimated at the conceptual design phase, subsequent decisions throughout the design and development processes are essentially biased in one direction or the other. Weight uncertainty may lead to increased acquisition cost, schedule overruns, performance deterioration, and increased operating costs. These potential effects make it necessary to address weight estimating uncertainty and consider the life cycle consequences at conceptual design. This research develops a stochastic methodology to incorporate weight estimating uncertainty for a launch vehicle as a complex system at the conceptual design phase.

For conceptual design of a complex system, a primary barrier to overcome in the estimation process is the lack of data. The following section discusses the use of expert judgment data to overcome this barrier.

3.3 Expert Judgment Data

Morgan (1984) suggests that "point estimates are of little use unless they are accompanied by measures of their accuracy." Morgan's comment is directed at the output of a simulation but the same can be said for the inputs to simulation. A range of estimates provides more input to the risk analysis simulation procedure than does a point estimate (i.e. a point estimate cannot specify a probability distribution). Or as Kirkwood (1997) says, "giving a single number does not provide information about how much variation is possible in the actual number." Kirkwood's observation that "historical data are often only loosely relevant to the current situation" (Kirkwood 1997) warns us that we should expect variation from a point estimate that is based on historical data. We should never expect a point estimate to be an exact outcome for some future event. As

Black and Wilder (1980) have suggested, good data is needed to make risk analysis meaningful. Heeding these observations, some means is needed to provide more information and thus more data than merely a point estimate in order to provide a robust methodology.

Expert judgment is a common and essential element for situations similar to the one faced at the launch vehicle weight estimating task. Morgan and Henrion (1990) suggest that asking experts for their "best professional judgment" is sometimes the only option when faced with a situation that has limited data or is not fully understood. Limited data or lack of understanding preclude the use of conventional statistical methods such as a regression of historical data points.

As a result, an expert judgment methodology is used in this research as a primary means for obtaining upper and lower bounds and most likely values for subsystem weight estimating relationships (WERs). These bounds become primary inputs to the stochastic methodology developed in this research. Expert judgment comprises a major portion of the methodology for providing the inputs. The objective is to provide a range of estimates and their associated measures of accuracy. The data set developed through the expert judgment elicitation is used as inputs to a Monte Carlo simulation procedure. The output from the simulation becomes a range of estimates with associated measures of accuracy or confidence percentiles.

The methodology development is described in the following chapter. Refinements and reasons for changes are also discussed. Example cases are used to aid refinement and to demonstrate the methodology.

Chapter IV

METHODOLOGY SYNTHESIS

4.1 Methodology

The focus of this research is the development of a methodology to obtain a data set that can be used to conduct a risk analysis for weight estimates. This chapter describes the methodology, the example cases that are used to refine the methodology, and the issues related to integrating the methodology with existing methodologies at NASA. These are presented in a chronological or sequential fashion as they were encountered in the course of the research.

During the initial phase of this research, a questionnaire was developed to elicit uncertainty ratings from the expert for a set of WERs. The elements of this questionnaire are discussed along with the results from a simplified example analysis. Refinements are made to the questionnaire and to the methodology. These are discussed along with a subsequent full size launch vehicle example.

4.1.1 Initial Proposed Methodology

An initial questionnaire was developed that included nineteen (19) subsystems for a full launch vehicle design. This was later reduced to a simplified model utilizing only eight (8) subsystems. This simplified model was the first attempt by the expert to utilize the methodology and was used to evaluate the usefulness of the methodology. This also afforded an opportunity to make changes to simplify and improve the methodology. Some of the details of the questionnaire development are discussed in the following section.

4.1.1.1 Questionnaire

Based on the detailed information required to quantify WER parameter uncertainty, a questionnaire was developed as a practical and efficient approach for eliciting the expert's opinion. The questionnaire incorporated multiple techniques drawn from the literature. The elements of the questionnaire as developed initially are described in the following steps.

An initial assessment was requested of the expert for each of the subsystem weight estimating relationships (WERs). This assessment was provided as the Low, Most Likely and High value for each WER. After the initial assessment, the expert was requested to rank subsystems for uncertainty of WERs on a five-point scale with low, moderate or high uncertainty as the three major points and two intermediate points on the scale. This incorporated the findings of Wallsten, Budescu, Rapoport, Zwick, and Forsyth (1986), Zimmer (1983) and Lichtenstein and Newman (1967) that qualitative assessments are more easily obtainable than are probability assessments.

Next, the expert was asked to anchor the WER uncertainty by identifying the *most uncertain* and *least uncertain* subsystems first and second respectively. This allowed the expert to assess the remaining subsystems on a relative basis against these two anchor points. This incorporated the feature suggested by the research of Lichtenstein and Newman (1967), Budescu and Wallsten (1987) and Wallsten, et al. (1986) that fewer verbal descriptions of uncertainty should lead to better quantitative assessments.

After all subsystems were rated on the 5-point scale, the expert was asked to anchor his qualitative rating by explaining his understanding of "Low", "Moderate", and

"High" uncertainty. For the intermediate points, "2" was quantified as the average of the expert's rating of "Low" and "Moderate". The rating, "4", was quantified as the average of the expert's rating of "Moderate" and "High". This was again accomplished via questionnaire with a suitable range of uncertainty percentages placed along a 5-point scale for each of the qualitative ratings.

After the uncertainty ratings were completed, the expert was asked to review the initial WER range valuations and to consider making any adjustments. During this second assessment, the expert used the initial assessment and the uncertainty rating as inputs to the reevaluation. One final step asked the expert to describe any scenario that might change the valuations that he had applied to any subsystem. This allowed the expert to consider competing technologies, substitute materials and similar scenarios. This served as a methodology element that debiases the judgment as suggested by Mullin (1986).

Throughout the assessment, *mental simulation* was an implicit heuristic as the expert was asked to envision different parameter values and visualize different scenarios. The nature of technological change tended to alleviate any concern for the "downhill" bias that Kahneman and Tversky (1982) documented. That is, technological changes normally specify the direction of parameter changes as part of the objective to be achieved by the technology (e.g. carbon fiber composites offer high strength, light weight and high heat resistance). In addition, the multiple techniques employed here have challenged the expert's opinion as suggested by Hoch (1984), Spetzler and Stael von Holstein (1975) and Mullin (1986) to provide multiple filters for removing any potential bias. Example questionnaire elements are shown in Table 3. A more detailed example of the initial questionnaire is presented in Appendix A.

Table 3 Sample Questionnaire Features

Please provide a lower bound, a mode (most likely) and an upper bound for all subsystems. Estimates may be provided at any level within the subsystem group that you feel appropriate.
Please rate each subsystem on a scale 1 to 5 with 1 being LOW Uncertainty and 5 being HIGH uncertainty. MODERATE uncertainty would be rated 3. Which subsystem is most uncertain? Rate that subsystem now. Which subsystem is least uncertain? Rate that subsystem now. Use these two anchors to rate the other subsystems as HIGH, LOW or MODERATE uncertainty relative to your first ratings.
Your understanding of high uncertainty would be associated with what confidence level? In other words - what percent is uncertain? 20% 30% 40% 50% More
Repeat your assessment of Lower Bound, Mode and Upper Bound for each subsystem using subsystem weight uncertainty as additional information to assist your rating.
Please consider all subsystems one last time and describe any scenario that might add uncertainty that you have not considered in your previous assessment. Make any adjustments to the three point estimates that are affected by the scenario that you describe.

The simplified launch vehicle consisting of eight (8) subsystems was used as the example case. The WERs of these subsystems were the input variables. For the example, the expert judged the WER ranges and then the configuration and sizing program (CONSIZ) was executed to convert those to weight estimates. Resulting data from the questionnaire are presented in Table 4.

Table 4 Data resulting from questionnaire for simplified example

Subsystem	Pt. Est.	Description	Low	Mode	High
Wing - cwing	5.0	wing constant (lb/ft^2)	4.5	5.0	5.5
LH2 tank - c	0.364	unit wt of tank (lb/ft^3)	0.328	0.364	0.382
LO2 tank - c	0.458	unit wt of tank (lb/ft^3)	0.412	0.458	0.481
Basic structure-cbdy	2.0	unit wt of struct (lb/ft^2)	1.8	2.0	2.2
Secondary structure wtsec	12000	constant	9000	12000	13000
TPS - ctps	1.0	unit wt of tps (lb/ft^2)	0.9	1.0	1.3
Propulsion - towe	69.76	t/w engine (vac), ssme-77.5 at max power	45	50	55
Subsystems - csub	0.14	subsystems wt fraction	0.133	0.14	0.147

Next, the Monte Carlo procedure is discussed and the initial results are shown in the following section.

4.1.1.2 Monte Carlo Results for Simplified Example Case

Monte Carlo simulation uses random or pseudo-random numbers to sample from specified probability distributions. The sampling in Monte Carlo is entirely random, that is, a single sample may fall anywhere within the distribution range of the inputs. With enough iterations (repeated sampling) the input distributions can be entirely recreated. A sample of 1000 or more is usually sufficient to avoid clustering and fully sample the input.

For the simplified case the Monte Carlo simulation sampled from statistical distributions for weight rather than the statistical distributions for the WER parameters. This was necessary at this stage with no integrated simulation within CONSIZ. The simulation was executed on the PC-based software @Risk[R].

Empty Weight of the launch vehicle was the output variable of interest which was simply the sum of the eight input variables. In the example case, the output for Empty Weight was evaluated repeatedly using subsystem weight inputs sampled from appropriate statistical distributions. Each input (subsystem weight or WER) was specified as a statistical distribution (e.g. normal, beta, triangular, etc.). The results for the output variable (Empty Weight) were presented in histogram or line graph as either a probability density function (PDF) or cumulative distribution function (CDF). The essential elements of Monte Carlo simulation are highlighted briefly as follows:

1) **Inputs**: Each input variable was specified by a distribution (PDF) and the output variable was specified by an equation. A normal distribution was specified by the mean and standard deviation, a triangular distribution was specified by the minimum, most likely and maximum values, and other distributions would be specified by parameters particular to that distribution.

2) **Sampling**: A random number generator determined the point that was sampled from each of the eight subsystem PDFs. In this example, the output variable (Empty Weight) was determined by summing the eight subsystem weights that were randomly sampled for a given iteration.

3) **Simulation**: A simulation typically consisted of 1000 iterations, so the eight PDFs were randomly sampled 1000 times to arrive at 1000 estimates for Empty Weight. These 1000 points were displayed in PDF or converted by integration to a CDF representation of Empty Weight.

4) **Outputs**: Outputs were probabilistic representations of the output variable - Empty Weight. Results were presented in either histogram or line graph format and were shown in both PDF and CDF forms. PDF showed the relative frequency of different Empty Weight values based on the simulation procedure. The CDF allowed interpretation of percentiles associated with a given Empty Weight much like a confidence interval.

Two different probability distributions were assumed to make an initial comparison. The questionnaire data was assumed to fit the triangular distribution and a second heuristic assumed a normal distribution.

Questionnaire/Triangular example. The expert's three point estimates of each subsystem were used as the minimum, mode and maximum values to specify the triangular distribution parameters. The "TRIGEN" probability distribution (i.e. an option within @Risk[R]) was used for this example which is a variation on the triangular distribution. The TRIGEN distribution avoids the problem of the minimum and maximum values having essentially a zero probability of occurrence. The uncertainty percentages elicited in the expert questionnaire were used as probability percentiles for the minimum and maximum values (e.g. 10% uncertainty was used to specify the minimum as the 5% percentile and the maximum as 95% percentile for each subsystem weight).

Point estimate example. A simple or naive heuristic using the point estimate and assuming a normal probability distribution was compared to the questionnaire results which utilized a triangular probability distribution. The point estimate method used the single point estimate of weight as the mean weight for each subsystem and assumed 10% of the mean as the standard deviation in order to specify the normal distribution parameters (i.e. mean and standard deviation) for simulation.

Both examples were evaluated by Monte Carlo simulation with 1000 iterations each. @RISK[R] personal computer software was utilized to conduct the simulation for these examples. Example graphical outputs are presented in Figure 1 and 2 respectively. Additional comparisons of outputs are presented in Table 5. These results were also presented in Monroe, Lepsch and Unal (1995).

Table 5 Simulation Results for Simplified Example

Measures of Empty Weight	Point Estimate with Normal Distribution	Questionnaire Data with TRIGEN Distribution
Minimum	164,129	138,068
Maximum	210,014	266,106
Mean	185,865	202,851
Std. Deviation	7,353	20,021
5% Percentile	173,444	170,452
10% Perc.	176,616	176,973
50% Perc.	185,994	202,378
90% Perc.	195,075	228,959
95% Perc.	197,868	236,885

This simplified example served as a demonstration that the methodology would in fact produce results and outputs that were expected and desired. In particular, weight estimates could be represented in PDF or CDF format with associated probabilities for the different weight estimates. No measure of accuracy or error was possible since the launch vehicle has not been built.

One interesting comparison for the simplified case was that the point estimate of weight fell at the 28th percentile of the CDF that resulted from the simulation using the triangular distribution. The comparison of the two assumed distributions found that the triangular distribution resulted in a larger variance and standard deviation than did the normal distribution. The triangular distribution resulted in extreme values that were a greater distance from the mean value than did the normal distribution. These differences were directly the result of using elicited values for the extreme values of the triangular distribution versus a naive assumed value (+ or - 10%) for the minimum and maximum values of the normal distribution.

Figure 1 Simulation Results (CDF)

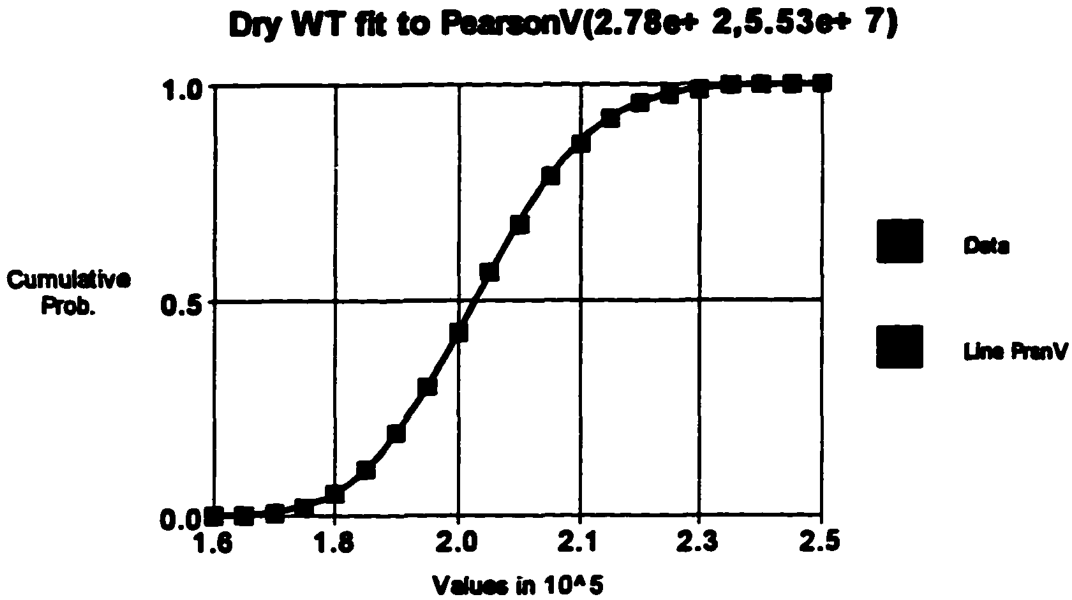
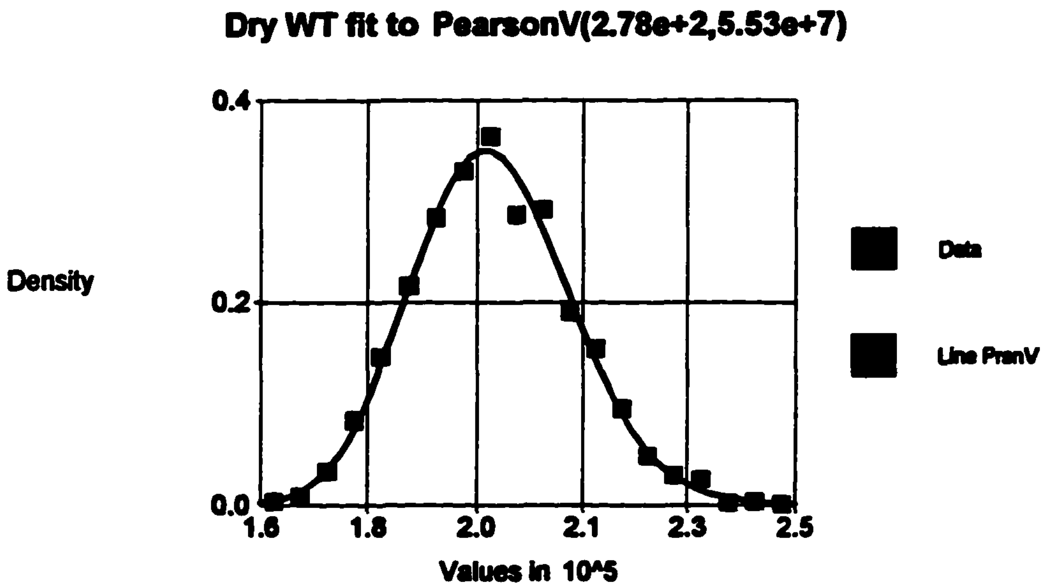


Figure 2 Simulation Results (PDF)



Although the initial case results were encouraging, some shortcomings were identified that led to refinements in the questionnaire and the methodology. The following section details some of those shortcomings.

4.1.2 Methodology Drawbacks

Based on feedback from the expert participating in this research, several drawbacks or shortcomings were recognized and addressed. The initial questionnaire was time consuming and cumbersome. The initial three point value assessment for WER parameters and intervening steps for uncertainty ratings followed by reassessment of the three point values for the WER parameters was problematic. This procedure was too long, somewhat redundant and too sequential in nature. Time was a primary metric to avoid an elicitation procedure that might take thirty to ninety minutes per quantity (Spetzler and Stael von Holstein 1975; Shephard and Kirkwood 1994). User friendliness was also a primary consideration and led to the exclusion of steps that did not satisfy this criterion.

Scenarios were a useful step for documenting alternative assessments but there was no explicit documentation for the primary uncertainty assessments. The following section discusses some of the refinements that were made to improve the methodology.

4.2 Methodology Refinement

Based on the feedback from the expert, several changes to the questionnaire were deemed appropriate. The initial assessment of parameters at three levels (Low, Most Likely and High) was dropped since this was a difficult starting point and a redundant assessment was included later in the multiple steps as originally developed. Ranking of the most uncertain subsystem and least uncertain subsystem was revised since the expert felt that this was difficult to do and provided little help for rating other subsystems on a relative basis. This was still incorporated in an intermediate questionnaire as a way of ranking the uncertainty of all subsystems. Ultimately, this was dropped altogether since the expert skipped this step in the assessment of the second example case, the full size launch vehicle model, and felt that it was not useful.

These changes resulted in the qualitative uncertainty assessment becoming the starting point. Additional discussions led to combining elements in a format so that uncertainty ratings and reasons could be documented simultaneously. Cues were added as a second prompt for the expert to document as many reasons and cues as possible that were actually influencing his ratings. The following section discusses the questionnaire elements in more detail and highlights the literature that served as a guide for the refinement/development.

Qualitative assessments. Qualitative assessments of uncertainties have been shown to be easier to elicit than are probabilities. Lichtenstein and Newman (1967) started with this premise but they devised experiments that resulted in mediocre or poor qualitative assessments. The experiments consisted of a list of 41 different verbal descriptions of some level of uncertainty which was administered to over 225 male employees at System Development Corporation. The subjects assigned numerical probabilities to the verbal phrases. The researchers found that there was a lack of consistency for some of the verbal phrases and in particular they found that phrases that they deemed as complements (summing to 1) were not quantified in that manner by the subjects.

Not discounting the value of this research, the shear magnitude of the number of phrases on the list seemed to overwhelm the subjects. Faced with a list of 41 phrases and with no instructions to develop complements, there should be no surprise that *rather unlikely* and *rather likely* were not quantified as complements by the subjects.

Even among the subjects for this experiment Lichtenstein and Newman (1967) found some level of consistency for the majority of phrases. For example, "the 124 people willing to assign a probability to this ambiguous word [i.e. rather] showed fair agreement" and "The reliability check on the duplicated entry, 'rather unlikely,' showed satisfactory stability" (Lichtenstein and Newman 1967, p. 563). The conclusion that should be drawn from the Lichtenstein and Newman (1967) experiment is that a select few verbal phrases should be utilized to describe uncertainty situations. By selecting only the vital few phrases, the quantification should be more straightforward. There should be less overlap, redundancy or duplication and there should also be no problem with overlooked complements if they exist.

Wallsten, Budescu, Rapoport, Zwick and Forsyth (1986) provide support for the vital few approach. Their experiments were executed with ten phrases and six phrases respectively. These are logically much more manageable than a list of 41 verbal phrases describing uncertain probabilities. Their experiments demonstrate good monotonic consistency among their subjects when they are asked to express vague verbal phrases over a probability interval (Wallsten, et al. 1986).

Wallsten, et al. (1986) suggest that in general, people prefer verbal expressions of uncertainty over numerical expressions. Even expert forecasters are included in this generalization (Wallsten, et al. 1986). Uncertainty assessments are really just an opinion, and since an opinion is imprecise by definition, numerical expressions may indicate precision when there is none (Wallsten, et al. 1986).

Many people also feel that they have a better understanding of words than numbers (Wallsten, et al. 1986). Probability was not formally developed until the 17th century with the work of Reverend Bayes, while language has a much longer history (Zimmer 1983; Wallsten, et al. 1986). Zimmer (1983) believes "that people generally handle uncertainty by means of verbal expressions and their associated rules of conversation, rather than by means of numbers" (Wallsten, et al. 1986).

From these observations and research findings, the elicitation procedure begins by asking for qualitative assessments of uncertainty. Qualitative verbal descriptions are limited to a very few (only five) to alleviate the overlapping or redundant categories that result in interpretation problems evidenced in other research (e.g. Lichtenstein and Newman 1967, Budescu and Wallsten 1985, Beyth-Marom 1982).

Asking an expert to evaluate a set of parameters stated in logical units (e.g. square feet of surface area, cubic feet of volume, or percent of weight reduction) is a complex undertaking. Asking an expert to apply probabilities of uncertainty directly to those logical units adds complexity unnecessarily. This process is particularly complex because each subsequent parameter is expressed in different units than the preceding one. The elicitation process has been designed to minimize adding complexity by starting with the qualitative assessments rather than starting with a quantitative assessment. As Wallsten, et al., note: " it is just when the uncertainty and the events are ill defined that non-numerical expressions are normally used" (Wallsten, et al. 1986, p. 362).

Reasons and Cues. The documentation of reasons and cues is important for three primary reasons. First, the documentation of reasons is an integral part of the methodology that forces the expert to describe the reason for the uncertainty rating. This serves as an honesty check to assure that the uncertainty rating is based on an actual reason rather than for some frivolous reason. Secondly, documentation serves as a history of the expert's thinking while providing the uncertainty ratings. Since this methodology has been developed to address uncertainty in an applied engineering setting, the documentation serves as a reference that will be used in future evaluations of this same project or for similar projects. Thirdly, the documentation serves as a history of the expert's thinking which can be evaluated as to the types of reasons and cues that are important to the expert. This evaluation may allow better understanding of the expert's assessments or may lead to further refinements to the methodology depending on the types of reasons and cues that are given. These reasons closely parallel the reasons for documentation suggested by Morgan and Henrion (1990). Hoch (1984) also suggested asking for reasons as a way of debiasing expert's judgments. This feature also serves the purposes of making the "knowledge accessible to others" and of helping "users organize their own knowledge in an effective way" (Fischhoff 1989).

The final version of the questionnaire requests that the "Reasons" will be documented simultaneously while providing the "Uncertainty" rating for each WER design parameter. This is done in order to make the documentation while the reasons and cues are current in the expert's mind. If reasons were provided at some later stage in the elicitation process, the expert would have to rely on memory and attempt to recall the thinking at the time of the uncertainty rating. By making the documentation of

"Reasons" concurrent with the "Uncertainty" rating, the need for perfect recall is eliminated and any error in the memory of the expert is eliminated.

"Cues" are requested by a separate prompt to encourage the expert to reflect on the thinking process and to provide any deep-seated cues that influence the uncertainty rating. This is done in an effort to surface any cues that reside in the depth of the expert's mind and have not been documented among the reasons thus far. A document was prepared to provide an explanation and an example to the expert to assist his understanding of cues. The document was based on an article by Ettenson, Shanteau and Krogstad (1987). This document is presented in Appendix B.

The essence of the article is that experts tend to use primary and secondary cues when making judgments. Through their experience with similar information, experts (i.e. professional auditors in the article) know which information has greater value and which information is of secondary value (Ettenson, Shanteau and Krogstad 1987).

The experiment described in the article (Ettenson, Shanteau and Krogstad 1987) is not an ideal example for an engineering problem but it does provide an example of why reasons and cues are important. That is to document primary and secondary cues that are influencing the expert's judgment. No weighting of importance is implied in this methodology for the two classes of information, reasons and cues.

The aim of requesting "cues" is very similar to one particular aim of knowledge engineering. That is the desire to draw upon "undocumented knowledge" - knowledge that resides in people's minds - and to surface "deep knowledge" - knowledge that is based on integrated human emotions, common sense, and intuition (Turban 1992, p. 120-122). According to Turban (1992), this type of knowledge is difficult to computerize. Rather than computerize this knowledge, this methodology seeks only to elicit responses from an expert that call this type of knowledge into use. Documenting this deep knowledge again provides a history that can be used as a reference in the future. Although the literature from expert systems is utilized here, the methodology that is developed is more appropriately analogous to a decision support system rather than an expert system.

Anchoring. Research indicates that there may be some uncertainty as to the quantitative value associated with verbal expressions of uncertainty (Lichtenstein and Newman 1967; Budescu and Wallsten 1985; Beyth-Marom 1982). For this reason, an anchoring step is employed to place a quantitative value on the expert's qualitative assessment of uncertainty. After all subsystem WERs are rated on the 5-point uncertainty scale, the expert is asked to anchor his qualitative rating by explaining his understanding of "Low", "Moderate", and "High" uncertainty. This is accomplished via questionnaire with a suitable range of uncertainty percentages placed along a 5-point scale (or 7-point scale) for each of the qualitative ratings. This serves as documentation of an individual expert's interpretation as to what "Low", "Moderate" and "High" uncertainty really mean on the quantitative scale. If the methodology were used for multiple experts this would serve as a check for disagreement among the group of experts. An additional step for reconciling differences might be needed in the event of using multiple experts. That discussion is beyond the scope of this research.

Anchoring and adjustment is a heuristic that is often cited in the literature and that was specifically studied by Kahneman and Tversky (1973). When experts use this heuristic, this commonly results in a bias towards the anchor (or a central tendency bias) because adjustments away from the anchor are inadequate (Morgan and Henrion 1990). This has also been used as an explanation for overconfidence when continuous probability distributions are assessed (Morgan and Henrion 1990).

Anchoring in this instance is viewed as a positive methodology element and may actually alleviate the potential of an anchor bias or central bias. The quantification of parameter values in this research follows a regular procedure that does not vary from parameter to parameter. The anchored values for the qualitative uncertainty assessments are used in combination with the qualitative assessments of individual WER parameters to arrive at the quantification of the parameter value ranges. The qualitative nature of the rating initially serves as an adjustment heuristic that will be applied according to the same rules for all parameters that received the same rating.

A strong argument for the anchoring element can be extrapolated from the following excerpt from Kirkwood (1997) when he quotes Merkhofer (1987):

> "[In a decision analysis seminar,] participants were individually asked to assign probabilities to common expressions such as "very likely to occur," "almost certain to occur," etc. The fact that different individuals assign very different probabilities to the same expression demonstrates vividly the danger of using words to communicate uncertainty. The seminar leader had just completed the demonstration when the president of the company said, "Don't remove that slide yet." He turned to one of his vice presidents and said the following: "You mean to tell me that last week when you said the Baker account was almost certain, you meant 60 percent to 80 percent probability? I thought you meant 99 percent! If I'd known it was so low, I would have done things a lot differently."

This example shows the problem associated with not having a quantification step for a verbal expression of uncertainty. The anchoring that is employed in the methodology can avoid any surprises (assuming that a higher level decision maker looks at the details). The anchoring element documents the percent of variation and becomes a permanent history of the quantitative values that are associated with the qualitative ratings provided earlier. Again, documentation makes the expert's judgment available to others.

Quantifying Parameter Value Ranges. Once the expert has placed a quantitative value on the qualitative assessment of uncertainty - in other words, quantified "Low", "Moderate" and "High" uncertainty - that quantitative value is used as the total variance from the original point estimate for each WER parameter. The expert provided feedback indicating that he interpreted the uncertainty rating to mean the full range of variance that would apply to a given parameter. *Uncertainty* here has been defined (or interpreted) as the total amount of variance for a design parameter from an initial design point estimate. In other words, given the nature of the WER parameters and what they represent, what is the potential range of a specific parameter value (assuming the variable is continuous). The expert is asked to specify the range in terms of a total percentage (i.e. total variation or total uncertainty). For example, the quantity of 20% would represent a total variation of -10% to +10% around the point estimate.

Based on the earlier individual parameter rating of Low, Moderate or High, the expert would then apply the quantitative value of the appropriate uncertainty to establish the Low and High parameter values. If a parameter was rated as having "Moderate" uncertainty and if "Moderate" were quantified as 20%, then the expert would calculate Low and High values that are -10% and +10% from the point estimate respectively.

This interpretation of the uncertainty rating is not as it was intended at the beginning of the research and in the first questionnaire. Two different frames as to how the uncertainty rating would be used actually developed. The researcher viewed the uncertainty rating as the percent of uncertain area under the tails of a probability distribution (i.e. beyond the minimum and maximum specified by the expert). The expert saw the uncertainty as the total amount of variation (much like a standard deviation) that the parameters might range across. Through discussions and revision of the questionnaire, any differences in framing were reconciled to arrive at a common frame. As described here, the expert's viewpoint was adopted as the official interpretation of the uncertainty rating. This was decided as were many methodological issues based on the desire to develop a methodology that was both meaningful and useful to the user - the expert. Forcing the expert to accept a definition that he does not find useful would hamper the desired purpose of the research and impede methodology development.

A representative section of the final version of the questionnaire is presented in Appendix C. The following summary is based on the final version of the questionnaire that was developed.

4.3 Summary of Final Questionnaire

The questionnaire has evolved through several iterations with ample feedback from the expert as to the usefulness of each element included in the questionnaire. The features of the questionnaire have also been selected to optimize the task characteristics of the elicitation process as advocated by Shanteau (1992). The multiple phases of the questionnaire consist of:

i.) Select the Parameters from WERs that will be evaluated for uncertainty.

ii.) Rate the parameter for uncertainty on a five point qualitative scale (Low, 2, Mod., 4, or High).

iii.) Document the reason(s) for the uncertainty for each parameter that is rated.

iv.) The expert is prompted to think of any additional cues that may further document the thinking process that affects the uncertainty rating.

v.) The expert is asked to anchor the three major points along the five point scale quantitatively. This documents the meaning of Low, Moderate and High uncertainty from the expert's perspective. These quantitative assessments are ultimately used as an estimate of the standard deviation for the statistical distribution.

vi.) Provide parameter values at three levels - Minimum, Most Likely and Maximum (the uncertainty rating and the quantitative anchor of uncertainty are used to aid this process).

vii.) Describe any scenario that would change a subsystem/parameter rating and also provide the changes that would result if that scenario occurred.

These questionnaire elements have also been designed with the idea of developing this into a computer-based assessment tool. Automation of the assessment process through computerization has been underway by a programmer at NASA Langley Research Center (LaRC).

The large number of parameters involved in launch vehicle design makes it very time consuming to evaluate every parameter. Selecting only those parameters that are most subject to uncertainty reduces the overall assessment task to a more manageable problem. When a given parameter is selected for evaluation, all information associated with that parameter is documented simultaneously. This allows the expert to focus on that parameter and record all the pertinent information while the reasons and cues are drawn upon to perform the uncertainty rating. The documentation of reasons and cues occurs when the assessment occurs and the expert's thinking is focused and fresh in his mind. Reasons and cues are the deep-seated knowledge in the expert's mind. Documenting these is an important means of sharing this knowledge that otherwise resides only with the expert. Cues in other research have been stimuli selected by the researcher to trigger a response by the subject. In this instance the cue is simply a stimulus that the expert acknowledges as being used in the process and that the expert documents in the process.

The procedure ultimately seeks to quantify risk and begins with a qualitative assessment by the expert for each of the subsystem WER parameters. The expert selects only those parameters within each WER that warrant an uncertainty rating. The expert is requested to rate subsystem WER parameters for uncertainty on a five-point scale with three points labeled low, moderate or high uncertainty (with two intermediate points between the three anchors). This incorporates the findings of Lichtenstein and Newman (1967), Wallsten, et al. (1986) and is empirically supported by Zimmer (1983) that qualitative assessments are more easily obtainable than are probability assessments. Next, the expert documents reasons for the uncertainty rating. The expert is then prompted to document any additional cues that may have influenced the uncertainty rating.

As a means of quantifying the qualitative ratings, the expert is asked to anchor Low, Moderate and High uncertainty as a percentage. The expert participating in this research reported that he spent a great deal of time thinking about what this meant. His inter-pretation was that "uncertainty" meant the total amount of variation that might be assoc-iated with a given parameter. Next, an assessment is given in the form of three point estimates. This incorporates the findings of Spetzler and Stael von Holstein (1975) and Beach (1975) that lower and upper bounds around a point estimate are easier to obtain than probabilities. This also heeds the advice from another source: " using a single number to represent an uncertain quantity mixes up judgments about uncertainties with assessments of the desirability of various outcomes. ... giving a single number does not provide information about how much variation is possible in the actual number" (Kirkwood 1997, p. 112).

Given the expert's interpretation of uncertainty as the total variation, the qualitative and quantitative ratings for a given parameter can be used to arrive at the Minimum and Maximum parameter values when the Most Likely value is known (from the design point estimate). One final step asks the expert to describe any scenario that might change the valuations that he has applied to any subsystem. This allows the expert to consider competing technologies, substitute materials and similar alternative scenarios.

Despite a seeming consensus that probabilities are difficult to elicit, many methodologies are based on exactly that approach (Mullin 1986; Shephard and Kirkwood 1994; Shephard 1990). Still other methods are based on eliciting five-point estimates at specified percentiles of a probability distribution (Spetzler and Stael von Holstein 1975). Both probabilities and multiple point percentiles add complexity to the elicitation process particularly for a large complex problem. The methodology elements embodied in this questionnaire have been established at the simplest possible level in order to minimize the complexity. These choices have been guided by the literature and by the input from the ultimate user - the expert. Otherwise the elicitation would become an impossible task and would result in a methodology that is not very useful (and likely would not be used).

The methodology also passes the clairvoyant test much more readily than if the methodology asked for probabilities. The clairvoyant test is a simple test to assure that questions are stated in an unambiguous manner (Morgan and Henrion 1990). As a simple example, which is more understandable?

Assuming a subject has a 5-year old son:

"What is the probability that your 5-year old son will be 5-feet tall on his 12th birthday?"

Or:

"What range of height around 5-feet do you think your son's height might vary within on his 12th birthday?"

Most will agree that the latter question is clearer, less ambiguous and easier to answer with some thought. This is comparable to the formulation of the questionnaire's approach in the methodology developed here.

For example, the two following questions could have been used in this methodology.

For the subject vehicle - single stage vehicle (ssv) dual-fuel, rd-701, 30 feet payload bay, 25 klb. payload-51.6 inc.:

"What is the probability of achieving a 0.30 (30%) weight reduction factor in the Avionics Cabling weight when compared to the Avionic Cable Weight of the space shuttle by using fiber optics?"

Or:

"What is the uncertainty associated with the point estimate of a 0.30 (30%) weight reduction factor in the the Avionics Cabling weight when compared to the Avionic Cable weight of the space shuttle by using fiber optics? In other words, how much variation due to uncertainty should be included in the estimate of this factor? What is the range for this factor (Low, Most Likely, and High)?"

The first question is matter-of-fact and offers no guidance and no additional information. The multiple set of questions guides, clarifies and informs in order to elicit the information from the expert. These types of questions are implicit within the questionnaire and are not stated explicitly.

After refining the methodology, the revised questionnaire was administered to the expert to obtain a data set for a full size launch vehicle model. Results from the full size launch vehicle model are discussed in the following section.

4.4 Full Size Launch Vehicle Model

A full size launch vehicle design study was conducted using the refined methodology. The launch vehicle conceptual design data consisted of 70 WERs and 399 different WER elements. For the example studied, the uncertainty analysis focuses only on the most uncertain parameters or at least the uncertainty parameters deemed worthy of evaluation by the expert.

In this case, the expert selected 100 parameters to rate for uncertainty. Of these, 7 were rated HIGH for uncertainty, 23 were rated 4 (between Moderate and High), 39 were rated MODERATE, 22 were rated 2 (between Low and Moderate), and 9 were rated LOW. The remaining 299 WER elements were not assigned an uncertainty rating. These elements will be held constant during the simulation procedure. Those parameters with an uncertainty rating will be represented by a statistical distribution during the simulation procedure. Table 6 summarizes the results of the expert's uncertainty assessments. Actual uncertainty ratings and reasons that the expert provided are shown in Appendix C.

Table 6 Summary of Questionnaire Results

Total WERs Evaluated	70
Total WER elements	399
Total WER parameters rated for uncertainty	100

Questionnaire Results. At this stage, when the expert has completed the questionnaire and calculated the three levels for parameter values, the data set for simulation has been completed. All that remains is to encode the data set in a suitably formatted UNIX file that can be accessed to perform the Monte Carlo simulation. One primary problem that had to be dealt with at this point was the nomenclature that had been used in naming variables for CONSIZ. Since CONSIZ looked at individual WERs the variable name "c" had been used repeatedly. Unique variable names were needed in the development of the all encompassing model that was needed for the Monte Carlo simulation. The minute details of file development are not presented here.

The data that was developed as a result of the questionnaire is presented in Table 7 in Appendix D. Developing this data set was a major aim of this research in order to facilitate the execution of the risk analysis.

One interesting note was that the expert recorded the following note for the omstnks_isp parameter: "extra low uncertainty, 2% (could use skewness here)" (documented in footnotes for Table 7). This type of information might not be obtained if a normal distribution was assumed and some simple algorithm was used to establish the standard deviation for the normal distribution. This also exhibits the flexibility of the methodology to allow for these types of adjustments in the rating procedure without much difficulty.

Another interesting result in the data was that the expert overruled the point estimate for all point estimates that are italicized and bold in Table 7. He then provided a range of three estimates that excluded the point estimate entirely. Proceeding through the elicitation methodology that requires some thought and documentation results in this type of information being obtained. Other naive assumptions might result in less rigorous evaluation and might fail to obtain data of this kind.

Once the full data set was obtained, the inputs for Monte Carlo were established using the triangular distribution. Results from the Monte Carlo simulation for the full size launch vehicle design are presented in Table 8.

Table 8 Simulation Results

Launch Vehicle Design Estimates	Questionnaire Data with Normal Distr.	Questionnaire Data with Triangular Distr.
Minimum Empty Weight	170,623	167,129
Maximum Empty Weight	238,017	244,769
Mean Empty Weight	199,036	199,676
Std. Dev.	9.379	11,564
Mode	197,926	198,263

4.5 Integration with computerized launch vehicle design and analysis tools

NASA Langley Research Center Vehicle Analysis Branch (VAB) utilizes a variety of computer based design analysis tools to examine Earth-to-orbit vehicle options to replace or complement the current Space Transportation System (Freeman, Wilhite, and Talay 1991; Stone and Piland 1992; Unal, Stanley, Engelund, and Lepsch 1994).

Weights and sizing analysis, is performed using the NASA-developed Configuration Sizing (CONSIZ) weights/sizing package. CONSIZ provides the capability of sizing and estimating weights for a variety of aerospace vehicles using WERs based on historical regression, finite element analysis, and technology level.

4.5.1 Weight Analysis Tool: CONSIZ

One initial objective of this research was to integrate risk analysis with the existing conceptual design evaluation programs currently in use at VAB. CONSIZ is a program that is currently used to evaluate vehicle configuration, size and weight (Lepsch, Stanley, Cruz and Morris 1991). Typical CONSIZ estimating models include all the interdependencies between subsystems so that changes that alter one subsystem are reflected by changes in other interdependent subsystems. This assures that the conceptual design satisfies all mission specifications (e.g. payload, orbit, etc.) and that the outputs represent a feasible launch vehicle. The output from CONSIZ usually is a single point weight estimate for a given launch vehicle configuration.

The risk analysis methodology and the Monte Carlo subroutine developed in this research must interface directly with CONSIZ. When Monte Carlo simulation iterations are performed, CONSIZ computes the corresponding vehicle size and weights for each iteration. The final output is a probability distribution of expected launch vehicle weight (Gross Weight and Empty Weight) determined through the CONSIZ WERs incorporating uncertainty.

4.5.2 Monte Carlo - CONSIZ Integration

For the random number generator subroutine developed in FORTRAN for the Monte Carlo-CONSIZ integration the triangular distribution was reduced to two linear equation components.

If $u \le \dfrac{b-a}{c-a}$ $then, x = a + \sqrt{u(c-a)(b-a)}$

If $u \ge \dfrac{b-a}{c-a}$ $then, x = c - \sqrt{(1-u)(c-a)(c-b)}$

Where u is the uniform random variate generated and a = minimum, b = most likely, and c = maximum values for x of f(x) for the triangular distribution. The ratio, $\dfrac{b-a}{c-a}$, maintains any skewness that has been included in the three point estimates of WERs. This was verified by plotting simulation sampling densities for each subsystem WER. Additional verification was performed using the Kolmogorov-Smirnov goodness-of-fit test for the first full set of 2000 data points generated using the random number generator. The data was converted from a UNIX format to a DOS file and all data were evaluated using BestFit® personal computer software. Examples of the goodness-of-fit analysis are presented in Appendix F. This analysis confirmed that each data set was sampled from the triangular distribution.

Note that the data was used for this confirmation rather than the random numbers since the random numbers are sampled from a uniform distribution. The FORTRAN command "RAND" was used to generate the uniform random numbers and should not require validation since prior validation of FORTRAN commands is assumed. The validation of the data sets essentially validates the random number generator indirectly.

The primary purpose of the analysis was accomplished. That is the validation of the subroutine for triangular distribution sampling that utilizes the uniform random numbers.

In order to develop an efficient methodology, the simulation parameters need to be specified. The next section discusses the Monte Carlo simulation system parameters in more detail.

4.6 Monte Carlo Simulation System Parameters

To determine the most efficient and economical simulation length, simulations were conducted for several different numbers of iterations. This analysis was conducted using the questionnaire data for the full size launch vehicle design. Based on the results of this analysis, 2000 iterations was determined to be a suitable simulation length for efficiency and effectiveness. The analysis is discussed in more detail in Appendix E.

Law and Kelton (1991) discuss several options for selecting a probability distribution in the absence of data. They suggest that the triangular distribution is appropriate for situations where a "rough model in the absence of data" (Law and Kelton 1991, p. 341) is needed. They also suggest that normal and beta distributions might be used but specifying these is obviously more difficult in the absence of data (Law and Kelton 1991).

Selection of a probability distribution was also evaluated as another simulation system parameter. Triangular and normal distributions were compared. The results of this comparison suggested that the triangular distribution did lead to a significant difference in the mean values for the simulation procedure when compared to the normal distribution. For equal sample sizes (n=15), there was a treatment effect for random number generator (RNG) when evaluating the mean value with a p-value of

0.000000327 ($F = 46.1771$ with F-critical $= 4.2252$). Details of this analysis are also presented in Appendix E.

These results were as expected and intended. One reason the triangular distribution was included in the methodology was due to its simplicity and ease of application. An additional intent was to allow for the incorporation of skewness in the assessments of parameter values. The specification of triangular distributions (potentially with skewness) would lead to significantly different results than the often assumed normal distribution. The comparison of the triangular and normal distribution results confirm these expectations.

4.7 Outputs and Potential Uses

Potential uses of the weight risk analysis methodology are threefold - as an input to other estimating analyses, as a means of WER refinement and as a comparative tool. The results would provide other analysts with a range of weight estimates at a given percentile of cumulative probability. The minimum, mean and maximum weight are also given from the Monte Carlo results. Probability distribution parameters are also available as an output. The probabilistic approach provides associated probabilities for each weight in the range of weights as depicted in the CDF. This should be a more desirable input to other estimating procedures than the single point estimate of weight (which usually forces the estimator to assume some probability distribution for weight).

A second potential use would be in the area of WER refinement, since information about model error is generated. This gives the weight engineer feedback on the estimation process and measures his confidence in the estimating model. The research has led to a better understanding of WER uncertainty and uncertainty quantification methods which will facilitate WER refinement. A subsequent follow-up evaluation with a group of experts contributed in this regard as well.

Another very promising potential use would be as a comparative tool. Competing launch vehicle designs can be evaluated through risk analysis and probabilistic weight estimates for each design will be determined. The engineer could compare the risk of the competing designs and cost estimators could use the outputs for similar comparisons of cost. For example, one possible result might be that a design with a higher mean weight may be pre-ferred due to lower risk when compared to competing designs.

These are the practical contributions of this research. The theoretical contributions are discussed in the following chapter.

Chapter V

RESEARCH FINDINGS

A number of research findings were identified throughout this research. The methodology development and some of the resulting information derived from the methodology were the primary topics for notable findings. The following sections discuss the research findings in more detail.

5.1 Methodology Development

The first finding was that careful selection of heuristics and guidelines from the existing literature was necessary in order to synthesize a workable and useful methodology. This was a fundamental observation that was recognized early in the process to avoid many of the pitfalls associated with expert judgment research. Efforts were made to identify and utilize heuristics and other techniques that had shown favorable results in previous research. The specifics of this were discussed in Chapter II.

Despite the best attempts and intentions, methodology refinement was still a necessary step in the research process. Research findings related to methodology refinement are discussed in the following section.

5.2 Methodology Refinement

The successive revisions to the questionnaire eventually led to the final version which asked for the documentation of reasons for the uncertainty ratings at the same time as the uncertainty rating was made. Reasons and cues are the deep-seated knowledge in the expert's mind. Documenting these reasons was an important means of sharing this knowledge that otherwise resides only with the expert. Cues in other research have been stimuli selected by the researcher to trigger a response by the subject.

In this instance the cue was simply a stimulus that the expert acknowledges as being used in the process and that the expert documents in the process.

The earlier version of the questionnaire had asked for uncertainty ratings and reasons in a sequential fashion. The final version recognized that the reasons for the uncertainty rating were important information that influences the rating. This made it logical to document the reasons at the time that the information was called upon to make the uncertainty rating. So the uncertainty rating and the reasons were executed simultaneously rather than sequentially. This research finding was suggested by the expert in the study and was further developed through discussion with the primary investigator and this researcher during a meeting on December 8, 1995.

This also highlighted another important research finding that crosses all dimensions of the research process. The importance of user acceptance was paramount throughout the research process. User accessibility and user feedback were essential in order to closely monitor the process for problems and to respond to the user's concerns. This finding, user acceptance, was intuitively consistent with the approach commonly adopted by software developers, knowledge base developers and decision support system developers.

Additional concerns for responsiveness to the user were evident in the framing of certain elements of the research program. Beach, et al. (1987) made observations regarding framing that support the way the methodology was developed in this study. They suggested that the way that the problem or question is framed by the researcher in many expert studies may in turn be framed differently by the subject (Beach, et al. 1987).

This leads to a measurement or interpretation that does not reflect the judgment accurately. More specifically, three types of error are possible:

- Misframing - the researcher's frame is correct but the subject responds to a different frame. Even good performance to an incorrect frame results in an inappropriate answer.

- Inadequate answer-generating process - the subject frames the problem correctly but does not know how to solve or answer the problem.

- Inadequate precision in the answer-generating process - subjects may rely upon faulty information, there may be 'noise' in the process, or the problem requires greater precision than the subject chooses to provide (Beach, et al. 1987).

They conclude by stating that "the experimenter's frame is not necessarily the only correct one and, because of this, it often is not clear upon what basis to evaluate the quality of judgment and reasoning" (Beach, et al. 1987). In order to avoid these potential sources of error, the expert's frame was considered throughout the evolution of the questionnaire and the development of the methodology. For example, there was a situation where two different frames developed as to how the uncertainty rating would be used. The researcher viewed the uncertainty as the percent of uncertain area under the tails of a probability distribution (i.e., beyond the minimum and maximum specified by the expert). The expert viewed the uncertainty as the total amount of variation (much like a standard deviation) that the parameters might range across. Through discussions and revision of the questionnaire, any differences in framing were reconciled to arrive at a common frame.

Additional framing cautions can be drawn from Lichtenstein and Newman (1967). Their research supports the use of qualitative rather than quantitative assessments of probabilistic events but their framing of the research might lead to different conclusions.

In the present research, the qualitative verbal assessments were limited to only five categories. Limiting the verbal descriptions to fewer categories and more distinct categories made for an easier assessment than would overlapping or redundant categories.

5.3 Demonstration of Methodology

Additional findings were related to the resulting outputs from the methodology and also the information used within the methodology. The probabilistic nature of the methodology required a change in mindset. This was true for the expert performing the uncertainty ratings and it was also true for administrators that are reviewing the outcomes from this methodology. A drastic change in perceptions was needed to move from a point estimate of weight to a CDF or probabilistic estimate of weight.

The expected outcome from weight estimation was the prediction of the "As built" weight at some point in the future. This expected outcome was prevalent (expressed by the expert and expressed by VAB administration) despite the lengthy timeframe between conceptual design and construction; despite the intervening design decisions; despite weight growth; and despite the uncertainty associated with the WERs themselves. This mindset did not allow for prediction of the Vehicle weight based on the design specifications at a given point in time with revised weight estimates made as new information becomes available or as new design decisions are made.

If these expectations are to be met, additional methods are needed to address all sources of uncertainty at conceptual design. This research has attempted to chip away at one segment of the uncertainty problem. Alternatively, more work is needed to encourage the shift in mindset that would gain acceptance for probabilistic estimates.

External reviewers were equally important to the research findings associated with the methodology demonstration. For example, an anonymous aircraft industry engineer served as a verification "EXPERT" when he commented on the presentation of some of these findings at the June 1996 SAWE National Conference in Atlanta, Georgia. Notably he commented that he was "not surprised that reduction factors were rated as the most uncertain WER elements" by the NASA expert. From his experience in the aircraft industry, reduction factors would likely be the most uncertain elements in virtually any aircraft/aerospace WER. This served as an indication that there was external validity in the results achieved through the methodology.

5.4 Analysis Findings

Additional findings resulted from statistical analyses that were conducted. A series of ANOVA's were conducted using different levels for number of iterations, different random number generators, and different simulation seeds.

Among factors - number of iterations, random number generator (statistical distribution) and simulation seed - only the random number generator or statistical sampling distribution resulted in a significant treatment effect for the analysis of variance. This outcome was predicted *a priori*.

While the number of iterations did result in a treatment effect when the Maximum and Minimum outputs were evaluated. No treatment effect was evident for any of the hypothesis tests for no difference in the Mean values. Only the factor, Random Number

Generator, resulted in a treatment effect when the hypothesis for no difference in the Means was tested.

One outside reviewer offered the following advice regarding the number of iterations: "Don't skimp on the number of iterations" despite what other references say (Wilder 1996). Based on this comment, additional simulations were conducted with 10,000 and 20,000 iterations. When these results were submitted to an ANOVA, no treatment effect was evident when comparing 20,000 to 2,000 and when comparing 20,000 to 5,000. From these results, the conclusion was that at 2,000 iterations, the simulation had not "skimped" on the number of iterations. Acceptable convergence had been achieved.

Another interesting finding dealt with the simulation results. Simulation outputs fit the Pearson V and Pearson VI better than the Normal, Beta, Lognormal, Triangular or any of 20 other statistical distributions evaluated. This outcome is consistent with findings reported by Law and Kelton (1991) for a number of simulations. The explanation for this tendency has not been attempted by others to date. Further work may lead to fully understanding why this is the case and what the implications are.

5.5 General Findings

The methodology can be used as a template for addressing other similar problems or entirely different problems. This is a primary research finding that applies to the methodology in the broadest sense. By wiping the slate clean and superimposing a different problem over the template, the methodology can be easily adapted to another problem. The qualitative uncertainty rating, the quantification of uncertainty, the documentation of reasons and cues would all remain as constant features of the methodology. All of these features are readily applicable to a wide range of problems.

The documentation elements of the methodology are also a "template" for future problems as suggested by Morgan and Henrion (1990). In that respect, the documentation of reasons and cues that occurs in any application of the methodology then serves as information and/or a template for future similar problems within that domain.

Chapter VI

RELIABILITY AND VALIDITY

6.1 Reliability and Validity in Research

Reliability and validity are concerns in all research and both are equally important here. Bolger and Wright (1992) suggest that more research is needed in order to maximize reliability and validity of expert judgment. This research may be seen as approaching that problem from a unique perspective given the circumstances with only one expert available.

Two means of ensuring the elicitation process has little effect on validity are:

> use percentages rather than odds or probabilities and encourage judges to decompose the problem in their own way (Bolger and Wright 1992).

Meyer and Booker (1991) argue that expert judgment is valid data and comparable to other "hard" data. "Just as the validity of hard data varies, so the validity of expert judgment varies" (Meyer and Booker 1991, p. 21). To ensure validity, they advocate careful selection of experts, vigilant monitoring and testing for bias, selection of elicitation techniques with substantial literature support, and minimization of assumptions about the expert data (Meyer and Booker 1991).

A similar viewpoint suggests that one method for ensuring validity is to utilize assessment procedures "that are based on previously developed and proven subjective assessment techniques" (Clemen and Winkler 1993). Selecting assessment techniques that have been used for similarly small samples and that have been tested for validity would be the most desirable approach. Borrowing a technique from a situation with a dissimilar sample size or dissimilar context is not advisable.

These are reasonable measures to be taken to ensure reliability and validity in subjective judgment. In developing this methodology, the researcher has attempted to adhere to as many of these suggestions as possible. Both of the latter suggestions from Bolger and Wright (1992) have been built into the elicitation process and detailed instructions that accompany the questionnaire related to my research proposal. All of these suggestions have been considered when additional refinements were made to the methodology.

Expert assessments are also improved when guided by an elicitation protocol (Shephard and Kirkwood 1994). The protocol presented here was synthesized from a variety of literature sources since no single source incorporated all the features deemed appropriate to the given situation. The development of the protocol was influenced by a wide range of research findings and numerous cautions.

The qualitative assessments are used by the expert as additional guiding information while performing his quantitative assessment. Detailed written instructions serve as guiding features throughout the questionnaire. These are additional measures aimed at ensuring that the subjective assessments are reliable and useful.

The primary methodology questionnaire was planned for completion by one NASA expert who performs weight estimation. This was necessary due to the fact that only one expert exists within NASA.

In addition to the features designed into the methodology, additional steps were taken to validate the methodology utilizing additional experts in the field of aerospace design or aircraft design. Specifically, other engineers with weight estimation expertise were the targeted subjects. This was completed during the December 1996 to March 1997 timeframe.

6.2 Follow-up Questionnaire with Multiple Experts

An abbreviated questionnaire was developed to be administered to a group of experts from within the broader domain of conceptual design engineering and weight engineering in the aircraft and aerospace industries. The International President of the Society of Allied Weight Engineers, Inc. (SAWE) was contacted and was asked to submit a list names of suitable subjects from within this domain. Additional subjects were selected from the SAWE membership roster based on their affiliation with an aircraft/aerospace agency or company. Selected subjects were contacted by e-mail to solicit their participation. Of nine subjects for which solicitation attempts were made, seven were successfully reached and six agreed to participate in the group questionnaire. The one declining stated that she had not performed any estimating tasks in several years. An alternate from this agency (NASA Lewis Research Center) was offered but further communication with the alternate led to his exclusion for inadequate relevant experience.

The selected six subjects consisted of conceptual design, preliminary design or weight engineers from Boeing, Northrup Grumman, NASA LaRC, NASA Johnson Space Center, and two individuals from Naval Air Systems Command (NAVAIR). The questionnaire was then mailed to these six individuals and they were asked to complete the questionnaire based strictly on their own knowledge (no group interaction between the six).

6.3 Instrument

The questionnaire was developed utilizing the subsystems that had been previously evaluated by the single NASA LaRC expert. A preliminary questionnaire was administered to another NASA LaRC contractor (not one of the six) and some minor changes were made prior to administering the questionnaire to the group of six experts. The questionnaire consists of two phases - conditioning and assessment.

The conditioning phase includes a brief narrative on the background of conceptual design for a launch vehicle and the problem of uncertainty at this design phase. The introductory material is followed by a set of instructions and a list of nomenclature to explain some abbreviations used in the questionnaire. Next, the group of experts are conditioned to the task by reviewing three (3) example uncertainty ratings along with reasons and cues that were completed by the original NASA LaRC expert. This parallels "calibration" that is seen frequently in the expert judgment literature but in this case the experts are conditioned to another expert's perspective of uncertainty in a given domain rather than being calibrated using almanac probability assessment tasks.

The second phase of the questionnaire starts with a set of instructions. The group of experts is then asked to perform an assessment of uncertainty and provide reasons and cues for five (5) subsystem WERs and six (6) specific parameters from those five WERs. The parameters were specified as those that were selected by the NASA LaRC expert when he performed his assessment. The subsystems were selected to include two that were specific to launch vehicle design and three that would have some commonality with aircraft subsystems. The included subsystems were Main Propulsion, Press and feed; Propellant tanks, Orbital maneuvering system (OMS) Tanks; Electric conversion and distribution, Avionic cabling; Electric conversion and distribution, Wire trays; and Main gear, Running gear. The first two subsystems are specific to a launch vehicle while the latter three should share some commonality with aircraft design.

Follow-up questions included general questions about the methodology and the interpretation of uncertainty. These took the form of the following:

Would you find the methodology useful if adapted to your own analysis problem with your own models? and Did you find the original expert's example judgments to be reasonable and understandable?

The questionnaire concludes with a set of Benchmark questions that are designed similar to a conditional probability statement. For example, "Given that a WER parameter value is based on a regression of historical data and the regression line has a good fit to the data, what is your uncertainty rating for such a parameter?" Five of these questions serve as benchmarks that require some knowledge of data sources and the estimating processes at the conceptual design phase but do not require specific model knowledge. This type of question removes the specifics of the subject launch vehicle and looks at data and sources of data in a generic manner.

The final step is to anchor the uncertainty qualitative rating for each of the group of experts. This serves as a direct comparison of the entire group of experts' quantification of the different qualitative ratings of uncertainty.

The group of experts were not asked to provide parameter values at three different levels since this is a simple application of the uncertainty qualitative rating and the quantification applied symmetrically. The multiple experts were also likely to have less experience with these specific parameters and were unlikely to place limitations (or skewness) due to theoretical limits. A complete listing of "Follow-up Questions" and "Benchmark Questions" are presented in the following pages. The full version of the follow-up questionnaire is presented in Appendix G.

6.4 Results

6.4.1 Expected Results

Given the general consensus in the literature that a group of experts may disagree by significant amounts (Mullin 1986; Lock 1987; Parente and Anderson-Parente1987), the expected outcome of this assessment by multiple experts is logically expected to be wide disagreement. Particularly for the qualitative assessment and the quantitative assessment of the qualitative rating, a wide range of interpretations is anticipated along with a wide range of qualitative ratings and quantitative ratings. These results are anticipated in keeping with the findings of Lichtenstein and Newman (1967), Budescu and Wallsten (1987), and Wallsten, et al., (1986).

"Different disciplines may have different terms for the same element or may use the same term in different ways. An inadequate modeling language may exacerbate such problems by reducing the opportunities for analysts to discover inconsistent terminology ..." (Fischhoff 1989, p. 452). "Although it can facilitate the incorporation of diverse perspectives, a risk assessment model can also inhibit the sort of unstructured interaction among analysts that helps to reveal and resolve discrepancies between their respective mental models of the system" (Fischhoff 1989, p. 453). This element of this research is susceptible to precisely these shortcomings. Although the original mathematical models have been provided for the group to evaluate, the meaning of individual model elements may be viewed differently by experts within the group.

Follow-up Questions

I. **Ease of use and/or usefulness of methodology and questionnaire.**

1. **Comment on the ease of use of the methodology.**

2. **Do you find the methodology to be useful for a weight estimation analysis?**

3. **Would you prefer to use your own models (WERs or MERs)?**

4. **Would you find the methodology useful if adapted to your own analysis problem with your own models?**

II. Uncertainty

1. **Did you find the original expert's example judgments to be reasonable and understandable?**

2. **Does this interpretation of uncertainty (as total variation) seem logical to you?**

3. **Do you have any other suggestion of how to <u>interpret</u> uncertainty?**

4. **Do you have any other method or any suggestion of how to <u>judge</u> uncertainty?**

<u>Benchmark Questions</u>

1. Given that a WER parameter value is based on a regression of historical data and the regression line has a good fit to the data:

What is your uncertainty rating for such a parameter?

Rate the degree of uncertainty that you associate with this parameter:

Low	**2**	**Moderate**	**4**	**High**

2. Given that a WER parameter value is based on someone else's analysis or experiment (for example a study at Marshall Space Flight Center or at Johnson Space Center, etc.):

What is your uncertainty rating for such a parameter? Explain your assumptions about the data source if that is an important consideration to you.

Rate the degree of uncertainty that you associate with this parameter:

Low	**2**	**Moderate**	**4**	**High**

Explanation (if required):

3. Given that a WER parameter is a reduction factor that <u>has been validated</u> using actual structures or by some other analytical techniques:

What is your uncertainty rating for such a parameter?

Rate the degree of uncertainty that you associate with this parameter:

Low	**2**	**Moderate**	**4**	**High**

4. Given that a WER parameter is based on a known design (such as the current space shuttle) and the new structure is assumed to be similar:

What is your uncertainty rating for such a parameter?

Rate the degree of uncertainty that you associate with this parameter:

Low	**2**	**Moderate**	**4**	**High**

5. Given that the subsystem structure being analyzed is not well-defined (i.e. very early in the conceptual design phase) and the WER parameter is estimated:

What is your uncertainty rating for such a parameter?

Rate the degree of uncertainty that you associate with this parameter:

Low	**2**	**Moderate**	**4**	**High**

No allowance for nor means to interact and reach a consensus understanding of models has been attempted or intended. Technical terms, language and even a limited number of uncertain verbal phrases will carry a significantly different meaning for different people including the small group of experts selected here.

The mediating factors that may produce different results (e.g. consensus agreement) are the small number of verbal qualitative descriptions of uncertainty that are used and the closely related fields from which the group of experts are drawn. These two factors may lead to greater consensus or at least greater consistency in the ratings and interpretations.

The questionnaire does require human judgment and subjective ratings. The subjective element of the methodology makes the former expected results the more likely results of this particular exercise.

This does not negate the usefulness of the methodology. Responses to the follow-up questions are anticipated to be favorable. That is, a consensus is expected for questions pertaining to the usefulness of the methodology and for the usefulness of the methodology if it incorporated the models of the expert in question. This reflects the intent of the methodology as it was developed. That is, the methodology was intended as a flexible and adaptable tool that could incorporate the models from any domain and any particular domain expert. The methodology was not intended as a consensus seeking technique for multiple experts. If used by a single expert for a specific task then the methodology is a template for ultimately developing data and for documenting the uncertainty and the reasons associated with the uncertainty ratings.

6.4.2 Actual Results

The following table shows the results from the Uncertainty Ratings for six (6) different WER parameters from five (5) different WERs.

Table 9 Subsystem WER Parameter Qualitative Uncertainty Ratings

	NASA LaRC Expert	Group expert #1	Group expert #2	Group expert #3	Group expert #4
Press and feed - cpf	Moderate	Low	2	Not rated	Moderate
OMS propellant tanks - ctnk	2	2	2	Moderate	Low
Avionic cabling - wac	4	4	4	High	Low
Wire Trays - wtrays	4	Moderate	Moderate	High	Moderate
Wire Trays - rtray	2	Not rated	4	Low	2
Main gear- Running gear cmrg	Moderate	High	*	High	Low

* The WER equation was omitted from the questionnaire. Group expert #2 developed his own model based on aircraft experience. Two versions were supplied - one based on horizontal takeoff and a second based on "needed for landing only."

The uncertainty ratings show mixed results although there is some consistency for three of the WER parameters. The OMS Propellant Tank parameter, "ctnk", was given a "2" rating by three individuals including the NASA expert. One other expert gave the parameter a "Moderate" rating, which is the next higher adjacent rating from "2".

The Avionic Cabling parameter, "wac", was given a "4" rating by the same three individuals as rated "ctnk" as a "2". The same individual that rated "ctnk" at the next higher rating chose the next higher rating for "wac" by assigning a "High" rating.

The Wire Tray parameter, "wtray", also exhibited some consistency in the ratings. The NASA expert gave this parameter a "4" rating and the first two experts gave the parameter the next lower adjacent rating of "Moderate". The third expert moved to the more extreme "High" rating which was still adjacent to the NASA expert's rating but further away from the rest of the group.

Expert #4 exhibited what might be considered extreme conservatism by providing the more "Low" uncertainty ratings than any other expert. Notably, this expert did assign the exact same rating as the NASA LaRC Expert on two out of six parameters and was adjacent to the NASA LaRC Expert's rating for a third parameter. This is particularly interesting because this agreement occurred when Expert #4 did not assign a "Low" uncertainty rating.

Obtaining results that show four individuals achieving some degree of consistency was encouraging and suggested that the "Conditioning" phase and the methodology itself serve as mediating factors. These results were more consistent than anticipated.

Group expert #3 showed a general tendency to be less conservative than others in the group and less conservative than the NASA expert. When this "non-conservative" individual was taken into account, the consistency of the responses was quite good.

The next table presents the results from the Benchmark Questions.

Table 10 Benchmark Question Replies

Question	NASA Exp.	Expert #1	Expert #2	Expert #3	Expert #4
Q1	Low	#	Low	Moderate	2
Q2	Difficult to answer.*	#	2	Moderate	Moderate
Q3	Low	#	Moderate	Low	Low
Q4	2	#	Moderate	Moderate	Low
Q5	High	#	High	High	4

Expert #1 did not respond to Benchmark Questions. They were developed after his responses.
* Note: "Without knowledge of how the analysis or experiment was performed and the experience level of the engineers, I would have to rate the uncertainty as high. With more understanding, the uncertainty level could potentially decrease, but would probably not be low." These were the additional comments from the NASA LaRC Expert.

The results from the Benchmark Questions showed some degree of consensus for Questions 4 and 5, adjacent assignments for Question 2, and somewhat different ratings for Questions 1 and 3. The disagreement on Questions 1 and 3 would indicate that the individuals place different importance on regression results using historical data (Q1) and different importance on validation by "actual structures or by some other analytical techniques" (Q3). Agreement on Questions 4 and 5 would indicate that each expert perceived the same uncertainty as the next expert for data based on "a known design" (Q4) and data related to a "structure [that is] not well-defined" (Q5). Expert #4 exhibited conservatism again by providing a "Low" uncertainty rating that diverged from the group's ratings.

The next table presents the results for the quantification of uncertainty.

Table 11 Uncertainty Quantification Replies

Qualitative Rating	NASA Expert	Expert #1	Expert #2	Expert #3	Expert #4
Low	10%	$	<5%	10%	7.5%
High	50%	$	30%	40%	40%
Moderate	30%	$	10%	20%	20%

$ Missing data. Expert #1 failed to return this portion of the questionnaire.

These results were mixed. Little consistency is evident but extreme values are not evident either. Expert #3 and Expert #4 were consistent and nearly perfectly calibrated with each other. However, if we review some of the results in combination with the earlier parameter rating an indication of consistency can be found. For the Wire Tray parameter, "wtray", the NASA expert gave a rating of "4" and Expert #3 gave a rating of "High". Now if we quantify those ratings, a "4" rating for the NASA expert can be derived by averaging his quantifications for "Moderate" and "High". This results in a percentage of 40% assigned to his "4" rating. Expert #3 assigned a quantitative value of

40% as his belief in what "High" uncertainty means. So, despite different qualitative ratings, the quantitative assessment is identical for this example. This also exhibits the importance and the value of having the qualitative rating and the quantitative anchor as elements in the methodology. This also exhibits the value of having both of these documented for further evaluation (especially when analyzing a group's ratings).

The next section presents the results from the "Follow-up Questions".

Table 12
I. Ease of use and/or usefulness of methodology and questionnaire.

1. Comment on the ease of use of the methodology.

Exp.	Response
#1	Basically well structured.
#2	Fairly easy to use - even though all of the examples were specific to rocket launch design.
#3	Fairly easy to use. My lack of reference material limited some answers.
#4	It's easy to pick a parameter. It's also easy to make assumptions. But its hard to get the assumption package "tuned" quickly because they all relate to one another.

2. Do you find the methodology to be useful for a weight estimation analysis?

Exp.	Response
#1	Yes, good supplemental information - but could weigh against effect on overall vehicle %.
#2	Yes - It is important to understand the limits of our estimating.
#3	Using expert opinion is always useful. [emphasis as originally provided by #3]
#4	Yes - It should bring focus to overall uncertainty and uncertainties in specific areas. Continual scrutiny and refinement should reduce the uncertainty or invalidate the approach.

3. Would you prefer to use your own models (WERs or MERs)?

Exp.	Response
#1	Most of WERs are my own models.
#2	Typically yes - each private entity in industry has spent years developing Parametric and Relational Data for initial estimates and Actual Products w/Analysis to support detailed estimates.
#3	Totally dependent on problem and WER documentation/reference material.
#4	Yes - always. This is a result of comfort and familiarity, also each engineering house knows their strengths and weaknesses and would naturally adjust focus.

4. Would you find the methodology useful if adapted to your own analysis problem with your own models?

Exp.	Response
#1	Of course.
#2	Maybe - estimating uncertainty has its place but [...]
#3	Yes.
#4	Yes, but I wouldn't consider that a finished approach. For each new product study the method would need a fresh review to adapt to the current design scenario and its unique sensitivities.

II. Uncertainty

1. Did you find the original expert's example judgments to be reasonable and understandable?

Exp.	Response
#1	Reasonable
#2	Yes.
#3	Mostly.
#4	Yes.

2. Does this interpretation of uncertainty (as total variation) seem logical to you?

Exp.	Response
#1	Yes, providing each subsystem is given a weighting factor.
#2	Yes.
#3	For the conceptual level.
#4	Yes.

3. Do you have any other suggestion of how to interpret uncertainty?

Exp.	Response
#1	This would be difficult to do.
#2	No.
#3	Consider other distributions for data collecting (Triangular?).
#4	Yes. Programmatic definition for key performance design issues have uncertainties of their own which impact the design. These are outside the loop of independent functional design and result in "sliding" the uncertainty scale.

4. Do you have any other method or any suggestion of how to judge uncertainty?

Exp.	Response
#1	One way might be to somehow quantify level of detail in the WERs.
#2	Not sure what you mean by Judge uncertainty. I think you mean evaluate and or Quantify. If so an approach of looking at historical trends of prediction vs. actual wt. of various systems and components could establish statistical variation over time and give plausible results - Note - structural variation very low; systems and payload variation typically High. New methods seldom as effective as advertised 50% or less.
#3	This question could be very broad. Please call and discuss it with me.
#4	Yes. "Beating on desks". Which means discuss concepts with designers to investigate whether their approach is well-known and confident or if there are significant technical issues that they are still groping with.

6.5 Summary Analysis

Mixed results were evident among the group of experts for uncertainty ratings of the example WER parameters, for the Benchmark Questions and for the Follow-up Questions. Much of the variation in responses and the non-replies might be attributed to a lack of experience with this set of WERs or to the fact that some "experts" were not expert in launch vehicle design. Although the group was well qualified in their respective fields, aircraft or aerospace, some of the specific WER parameters (i.e. for Propellant tanks) were unfamiliar to them.

Among the results there was consistency for portions of the questionnaire. Responses to Question #4 on the Methodology were particularly encouraging. These responses indicated a consensus on the usefulness of the methodology if it were adapted for the individual expert's models. This supports the assertion that the methodology can be used as a template for other problems. By replacing the current problem and current models with a different set of models (i.e., their own), the expert's are viewing the methodology as a template that they could use.

An important insight from Pitz is that a person's knowledge of and representation of "variability ... and other distributional properties such as skewness ... is less clear" (Pitz 1980, p. 88). This research takes a small step towards documenting some of the thought process and in particular the reasons that are used by people to describe variability and skewness. Documentation of reasons for uncertainty and the quantification of uncertainty ratings move in this direction. In particular, the NASA expert chose to overrule the scale and provide a rating of "Extra Low" along with a quantification and a reason.

The group of experts also provided revealing answers for the Section II Follow-up Questions on Uncertainty. The direct tie between uncertainty and variability made by this research was addressed by this set of questions. The group also provided reasons for their uncertainty ratings and provided a quantitative interpretation. The responses were a starting point for addressing the issue raised by Pitz (1980). This is another significant contribution of this research.

One of the most significant results from the follow-up was the demonstration of the usefulness of the combined qualitative rating and quantification of the qualitative rating. Despite different qualitative ratings, the NASA expert and one other expert arrived at the same quantitative rating which would then result in the same three parameter levels. This clearly demonstrated the benefit of having these two steps in combination within the methodology.

Chapter VII

DISCUSSION AND CONCLUSIONS

7.1 Discussion

The initial task was to find a method to perform a risk analysis for weight estimates of a launch vehicle. A risk analysis would provide weight estimates in probability density function (PDF) form or more appropriately in cumulative distribution function (CDF) form. While this is a graphical representation, the associated numerical values could be given as probability density function parameters (i.e., mean and standard deviation for normal distribution) or as a range of estimates or as some percentile value with an associated probability. Each of these are considered to be desirable forms that could be useful inputs to other estimating analyses.

As the research progressed, the primary hurdle to overcome was the scarcity of data. To overcome this hurdle, an expert judgment methodology was developed. The methodology borrowed many features from the fields of psychology and knowledge engineering or computer science.

For the first test, the methodology was applied to a simplified case for weight estimation of a launch vehicle. The results were satisfactory but the methodology had some rough edges. This led to refinement of the methodology to make it easier to use and to make each element more meaningful. Most of the revisions were prompted by comments from the end user, the weight estimating engineer.

Multiple techniques were included as integral features of the methodology that was developed for obtaining expert judgment. Problematic techniques identified through the literature review have been avoided. This research contributes to the expert judgment elicitation literature by presenting this synthesized methodology. This is one piece of the research puzzle that will begin to fill the perceived gap in "judgmental processes in risk analysis" (Fischhoff 1989).

Specifically, this methodology differs from other methodologies by incorporating a qualitative assessment as a starting point. The methodology does not elicit preferences, probabilities or utility functions. The absence of those types of elicitations is an additional difference from most methodologies. The documentation elements of the methodology are described in detail and serves as a model for other researchers or practitioners.

Most previous studies of expert judgment have dealt with antiseptic laboratory experiments utilizing non-experts. This study addresses a real problem in an applied engineering setting and utilizes an actual domain expert. Addressing an applied setting problem is a contribution since the bulk of the literature has addressed experiments conducted in a "laboratory" setting.

Of the previous applied setting research, neither Hammond, et al. (1987) nor Mullin (1989) dealt with the level of complexity and the degree of uncertainty that the problem in this dissertation involves. The approach taken in this dissertation and the problem domain being addressed appears to be unique when compared to the existing literature. Mullin (1986) seems to support this sentiment when she states, "an appropriate structuring of the estimation problem is crucial ... in the 'real world', [but] there is relatively little published work in this area to offer specific guidance" (Mullin 1986, p. 48). The methodology presented here describes the structuring of the problem and details all the related assessment elements required to accomplish the estimation task.

Other contributions of this research are also consistent with conceptualizations presented by Fischhoff (1989). That is, "to make ... knowledge accessible to others who hope either to exploit the ... expertise or to solve the ... problem" and " to help users organize their own knowledge in an effective way" (Fischhoff 1989). Both of these purposes will be well served through the methodology developed here.

In order to execute the risk analysis, Monte Carlo simulation was integrated with CONSIZ and demonstrated for a simple case. This included development of a random number generator for sampling from the triangular probability distribution. Data generated during the simulation procedure demonstration was submitted to a Goodness-of-Fit test. Tests were conducted for each data set to verify the most appropriate statistical distribution for the data. Matlab was also utilized to perform statistical analyses of the simulation results and to produce the basic graphical outputs (PDF and CDF).

At each phase, the aim was to make the methodology and associated procedures easy to use so that they would be used. After several refinements, the methodology was applied to a full launch vehicle weight estimation task. The final revision of the methodology incorporated all the recent suggestions including the opportunity to document the <u>reasons</u> for uncertainty ratings at the time that the rating is made. Data generation for the full vehicle design was completed in March 1996 and a Monte Carlo simulation was executed during the last week of March 1996. This effectively demonstrated the methodology for a full vehicle design, that is, every step of the methodology was executed and resulting outputs were achieved.

Subsequent activity focused on experimentation related to the execution of the Monte Carlo simulations. In order to optimize simulation parameters, more than 70 independent simulations were executed. The primary simulation parameters of number of iterations and random number generator were varied to achieve the optimal combination of these system parameters. The optimal system parameters were recommended from these results (see Chapter 5). The final outcome was a recommended simulation procedure that was designed to provide an appropriate amount of information from the simualtion results while also economizing on computer central processing unit (CPU) time.

One final task was to validate the methodology utilizing additional experts in the field of aerospace design or aircraft design. Specifically, other engineers with weight estimation expertise were the targeted subjects. This was completed during the December, 1996 to March, 1997 timeframe. All of the results from these additional evaluations were discussed in Chapter 5 and in Chapter 6 under Research Findings.

7.2 Conclusions

The methodology was developed, refined and demonstrated. Based on the expert's evaluation and on the comments from the group of experts, the methodology is a workable and useful methodology. Based on these results the methodology is expected to be a flexible risk analysis approach that can become a valuable analysis tool in the conceptual design of complex systems with uncertain design parameters. Programming is underway to implement the methodology as an analysis tool at NASA LaRC.

The methodology reduces the uncertainty rating task by focusing only on the parameters that warrant a rating, other factors are held constant. This primary feature of the methodology facilitates the development of data that can then be used as inputs to perform a risk analysis for weight estimates of a launch vehicle. The real contribution of the methodology is the development of expert judgment data in usable form. Documentation of the reasons for the uncertain parameter ranges provide a history for future evaluations. This integral feature of the methodology is a significant contribution since it is developed from a synthesis of other methodologies taken from examples in the literature. The methodology could also be adapted for other parametric analyses that need to address uncertainty and have little or no data available.

7.3 Limitations

"Often the most important judgments (requiring the skills of the most accomplished experts) concern matters that will not be resolved for years. As a result, there is little opportunity to learn about the overall quality of one's judgmental processes or how they can be improved" (Fischhoff 1989);(e.g. Fischhoff 1982; Brehmer 1980; Henrion and Fischhoff 1986). Research in realistic settings "may appear to be more 'relevant' ... than laboratory research, it may not necessarily be more generalizable or yield greater predictive accuracy, particularly because of the difficulties inherent in establishing controls in realistic settings and/or the often small number of experts used as subjects in such studies" (Beach 1975).

These observations are true of the research in this dissertation. The judgments cannot be verified conclusively until and unless the actual launch vehicle in question is built. While this does place some limitations on the research findings and the generalizability of this research, other means have been pursued to verify methodology features.

Experts have been used from related domains, aircraft and aerospace, in order to obtain some external verification and validity check for methodology features and the methodology as a whole. In addition, other critiques have been solicited for the earlier papers from this research and those criticisms have been acknowledged and incorporated into this document. These efforts lend some credence to statements of generalizability across a limited range of decision domains.

7.4 Future Extensions

Future research might include using the methodology for addressing uncertainty for the conceptual design of a different launch vehicle design. This would serve the purpose mentioned earlier of becoming a comparative tool.

The methodology could also be employed to a similar problem from a different domain such as aircraft design or shipbuilding. A more generalized test of the methodology would involve applying the methodology to a different type problem from an entirely different domain.

An analysis of the group process and group outcomes might be conducted employing the methodology and a larger targeted group of experts. This might reveal more about the consensus or disagreement among experts and might lead to an enhanced methodology for group ratings.

One of the group of experts suggested "pounding on desks". By this he meant an investigation of existing methodologies that are used by practitioners could be conducted. This type of investigation would serve to explore and document existing methods that have not previously been publicized.

The latter types of research (i.e., involving a group of experts) would afford a greater opportunity to draw generalizations and to explore multiple domains with the methodology. Other decision making research analysis techniques might be employed to assist in this effort. Metric conjoint analysis (Priem 1992) is one example of a technique that has been employed to study executive decision rules as they relate to organizational outcomes (Priem 1992). By adapting the metric conjoint analysis technique (or some other technique) to the group analysis, a more rigorous statistical analysis could be conducted. This would be particularly useful for problems where final outcomes can be analyzed as part of the research.

On the whole, the future research opportunities are abundant. Based on the results in this dissertation, the topics and the methodology are worthy of additional attention and investigation. Any of the future extensions of this research may serve to demonstrate the methodology's use as a template and add to the generalizability of this research.

REFERENCES

Augustine, N.R. 1979. Distribution of expertise. *Defense Systems Management*, Spring, 1979.

Beach, B.H. 1975. Expert judgment about uncertainty: Bayesian decision making in realistic settings. *Organizational Behavior and Human Performance*, 14, pp. 10-59.

Beach, L. R., J.J.J. Christensen-Szalanski and V. Barnes. 1987. Assessing human judgment: Has it been done, can it be done, should it be done? In *Judgmental forecasting*. Edited by G. Wright and P. Ayton. pp. 49-62. John Wiley and Sons: Chichester.

Beyth-Marom, R. 1982. How probable is probable? Numerical translation of verbal probability expressions. *Journal of Forecasting*, Vol. 1, pp. 257-269.

Bhola, B., R.M. Cooke, H.G. Blaauw, M. Kok. 1992. Expert opinion in project management. *European Journal of Operational Research*, Vol. 57, pp. 24-31.

Black, R. and J. Wilder. 1980. Cost risk methodology: A pragmatic approach. *Proceedings of 2nd Annual International Society of Parametric Analysts Conference*, Virginia Beach, VA, April, 1980, pp. 6-19.

Bolger, F. and G. Wright. 1992. Reliability and validity in expert judgment. In *Expertise and decision support*, edited by G. Wright and F. Bolger. pp. 47-74. New York: Plenum Press.

Bolger, F. and G. Wright [Editors]. 1992. *Expertise and decision support*. Plenum Press: New York.

Bolger, F. and Wright, G. 1993. Coherence and calibration in expert probability judgment. *OMEGA International Journal of Management Science*, Vol. 21, No. 6, pp. 629-644.

Brehmer, B. 1980. In a word: Not from experience. *Acta Psychology*, vol. 45, pp. 223-241.

Budescu, D. V. and T. S. Wallsten. 1985. Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, Vol. 36, pp. 391-405.

Budescu, D. V. and T. S. Wallsten. 1987. Subjective estimation of precise and vague uncertainties.. In *Judgmental forecasting*, G. Wright and P. Ayton [eds.], John Wiley and Sons: Chichester, pp. 63-82.

Bush, L., R. Unal., L. Rowell,. and J. Rehder. 1992. Weight optimization of an aerobrake structural concept for a lunar transfer vehicle. *NASA Technical Paper 3262*. NASA, Office of Management, Scientific and Technical Information Program.

Christensen-Szalanski, J.J.J. and L.R. Beach. 1984. The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist*, Volume 39, pp. 75-78.

Clemen, R.T. 1986. Calibration and the aggregation of probabilities. *Management Science*. Volume 32, No. 3, pp. 312-314.

Clemen, R.T. and R.L. Winkler. 1993. Aggregating point estimates: A flexible modeling approach. *Management Science*. Volume 39, No. 4, pp. 501-515.

Cooke, R.., M. Mendel, and W. Thys. 1988. Calibration and information in expert resolution: A classical approach. *Automatica* 24, 1, pp. 87-94.

Dalkey, N.C. 1969. *The Delphi method: An experimental study of group opinion*, Santa Monica, CA.: RAND Corporation.

Dawes, R.M. 1987. Personal communication to J. Shanteau (referenced in Shanteau, 1992).

Dawes, R.M. 1988. *Rational choice in an uncertain world*. Harcourt Brace Jovanovich: San Diego, CA.

Dym, C.L. 1994. *Engineering design: A synthesis of views*. New York: Cambridge University Press.

Edwards, W. 1954. The theory of decision making. *Psychological Bulletin*, 51, pp. 380-417.

Edwards, W. 1961. Behavioral decision theory. *Annual Review of Psychology*, 12, pp. 473-498.

Edwards, W. 1992. Discussion: Of human skills. *Organizational Behavior and Human Decision Processes*, Volume 53, pp. 267-277.

Einhorn, H.J. and R.M. Hogarth. 1986. Decision making under abiguity. *Journal of Business*, Vol. 59, No. 4, part 2, pp. S225-S250.

Engelund, W., D. Stanley, R. Lepsch, M. McMillian, and R. Unal. 1993. *AIAA 93-3967*, Aerodynamic configuration design using response surface methodology analysis. AIAA, Aircraft Design, Systems and Operations Meeting, August 11-13, 1993, Monterey, California.

Ettenson, R., J. Shanteau, and J. Krogstad. 1987. Expert judgment: Is more information better? *Psychological Reports*, Volume 60, pp. 227-238.

Fabrycky, W.J. and B.S. Blanchard. 1991. *Life-cycle cost and economic analysis*. Prentice Hall: Englewood Cliffs, New Jersey.

Fairbairn, R.E. and L.T. Twigg. 1991. Obtaining probability distributions of cost with approximations of general cost model functions. *Proceedings of International Society of Parametric Analysts*, pp. RA21-RA37

Fischhoff, B. 1982. Debiasing. In *Judgment under uncertainty: Heuristics and biases*, D. Kahneman, P. Slovic and A. Tversky, Eds. Cambridge University Press: New York.

Fischhoff, B. 1989. Eliciting knowledge for analytical representation. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 3, May/June, pp. 448-461.

Fishman, G.S. 1973. *Concepts and methods in discrete event digital simulation*. John Wiley and Sons: New York.

Forster, A. 1989. A risk cost analysis procedure as applied to advanced space programs. *Proceedings of International Society of Parametric Analysts*, 1989, pp. 249-259.

Freeman, D.C., A.W. Wilhite , and T.A. Taley. 1991. Advanced manned launch system study status. *IAF Paper 91-193*, October, 1991.

Freeman, D.C., D.O. Stanley, C.J. Camarda, R.A. Lepsch, and S.A. Cook. 1994. Single-stage-to-orbit - A step closer. *45th Congress of the International Astronautical Federation.* IAF 94-V3.534, Oct. 9-14, 1994, Jerusalem, Israel.

French, S. 1986. Calibration and the expert problem. *Management Science.* Volume 32, No. 3, pp. 315-321.

Fuller, R. Buckminster. 1996. *Buckminster Fuller: Thinking Out Loud.* A film by Karen Goodman and Kirk Simon. Broadcast on PBS-TV, April 10, 1996. Copyright 1996 S and G/BFI.

Gordon, L. 1988. The importance of weight in a changing cost estimating environment. *SAWE Paper No. 1855.* Presented at the 47th Annual Conference of The Society of Allied Weight Engineering, Inc., Plymouth, Michigan. May 23-24, 1988.

Hammond, K.R., R.M. Hamm, J. Grassia, and T. Pearson. 1987. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-17, no. 5, September/October, pp. 753-770.

Hertz, D.B. and H. Thomas. 1983. *Risk analysis and its applications*. John Wiley and Sons: Chichester.

Hillier, F.S. and G.J. Lieberman. 1986. *Introduction to operations research.* Holden-Day, Inc.: Oakland, California.

Hoch, S.J. 1984. Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 10, p. 4.

Hoffman, R., N. Shadbolt, A.M. Burton and G. Klein. 1995. Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes*, Vol. 62, No. 2, May, 1995, pp. 129-158.

Hogarth, R.M. 1975. Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, 70, pp. 271-291.

Hull, K. and R. Walters. 1990. Risk analysis techniques in UK defence procurement. *Proceedings of International Society of Parametric Analysts*, 1990, pp. 660-682.

Kahneman, D. and A. Tversky. 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, pp. 430-454.

Kahneman, D. and A. Tversky. 1973. On the psychology of prediction. *Psychological Review*, 80, pp. 237-251.

Kahneman, D. and A. Tversky. 1982. The simulation heuristic. In *Judgment under uncertainty: Heuristics and biases*. Edited by Kahneman, Slovic and Tversky, Cambridge University Press, Cambridge.

Keeney, R.L. 1972. Utility functions for multiattributed consequences. *Management Science*, Vol. 18, No. 5, pp. 276-287.

Keeney, R. L. and H.A. Raiffa. 1976. *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Keeney, R.L. 1977. The art of assessing multiattribute utility functions. *Organizational Behavior and Human Performance*, 19, pp. 267-310.

Keeney, R.L. and D. von Winterfeldt. 1991. Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management*, Vol. 38, No. 3, pp. 191-201.

Kelly, R.V., Jr. 1991. *Practical knowledge engineering*. Digital Press: Bedford, MA.

Kirkwood, C.W. 1997. *Strategic decision making: Multiobjective decision analysis with spreadsheets*. Duxbury Press at Wadsworth Publishing: Belmont, California.

Law, A.M. and W.D. Kelton. 1991. *Simulation modeling & analysis*. 2nd Edition. McGraw Hill: New York.

Lapin, L.L. 1981. *Statistics for modern business decisions*. Harcourt Brace Jovanich: San Diego, California.

Lepsch, R.A., Jr., D.O. Stanley, C.I. Cruz and S.J. Morris, Jr. 1991. Utilizing Air-turborocket and rocket propulsion for single-stage-to-orbit vehicle. *Journal of Spacecraft and Rockets*, Vol. 28, No. 5, pp. 560-566.

Lichtenstein, S. and Newman, J.R. 1967. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, Vol. 9, pp. 563-564.

Lock, A. 1987. Integrating group judgments in subjective forecasts. In *Judgmental forecasting*, G. Wright and P. Ayton [eds.]. John Wiley and Sons: Chichester.

MacCrimmon, K.R. and R.N. Taylor. 1976. Decision-making and problem-solving. In M.D. Dunnette (ed.), *Handbook of Industrial and Organisational Psychology*, Rand-McNally: Chicago.

March, J.G. 1994. *A primer on decision making*. The Free Press, Macmillan: New York.

Marse, K. and S.D. Roberts. 1983. Implementing a portable FORTRAN Uniform (0,1) Generator. *Simulation*, Volume 41, pp. 135-139.

Merkhofer, M.W. 1987. Quantifying judgmental uncertainty: Methodology, experiences and insights. *IEEE Transactions on Systems, Man, and Cybernetics*, Volume SMC-17, pp. 741-752.

Meyer, M. and J. Booker. 1991. *Eliciting and analyzing expert judgment: A practical guide*. Academic Press: San Diego.

Monroe, R.W., R.A.Lepsch, Jr. and R. Unal. Risk analysis of weight estimates for a launch vehicle. *SAWE Paper No. 2323*. Paper presented at the 55th Annual International Conference of the Society of Allied Weight Engineers, Inc., June, 1996, Atlanta, GA.

Monroe, R.W., R.A. Lepsch, Jr. and R.Unal. Weight estimation risk analysis: Incorporating uncertainty at conceptual design for a launch vehicle. In *Proceedings of 1995 American Society for Engineering Management National Conference*. pp. 25-32.

Morgan, B.J.T. 1984. *Elements of simulation*. Chapman and Hall: New York.

Morgan, M.G. and M. Henrion. 1990. *Uncertainty: A guide to dealing with uncertainty in qualitative risk and policy analysis*. Cambridge University Press, Cambridge.

Mullin, T.M. 1986. *Understanding and supporting the process of probabilistic estimation*, Ph. D. dissertation, Carnegie Mellon University, Pittsburgh, U.M.I.

Mullin, T.M. 1989. Experts' estimation of uncertain quantities and its implications for knowledge acquisition. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 3, May/June, pp. 616-625.

Parente, F.J. and J.K. Anderson-Parente. 1987. Delphi inquiry systems. In *Judgmental forecasting*, G. Wright and P. Ayton [eds.], John Wiley and Sons: Chichester, pp. 129-156.

Pate-Cornell, M.E. and P.S. Fischbeck. 1994. Risk management for the tiles of the space shuttle. *Interfaces*, 24: 1, January-February, 1994, pp. 64-86.

Phelps, R.H., and J. Shanteau. 1978. Livestock judges: how much information can an expert use? *Organizational Behavior and Human Performance*, 21, pp. 209-219.

Pitz, G.F. 1980. The very guide of life: The use of probabilistic information for making decisions. In *Cognitive processes in choice and decision behavior*, T. S. Wallsten [ed.], Lawrence Erlbaum Associates, Publishers: Hillsdale, New Jersey.

Priem, R.L. 1992. An application of metric conjoint analysis for the evaluation of top managers' individual strategic decision making processes: A research note. *Strategic Management Journal*, Volume 13, pp. 143-151.

Quinn, R.E. and J.P. Walsh. 1994. Understanding organizational tragedies: The case of the Hubble space telescope. *Academy of Management Executive*. Volume 8, No. 1, pp.62-67.

Raiffa, H. A. 1968. *Decision analysis*. Reading, Mass.: Addison-Wesley.

@Risk User's Guide, 1994, Palisade Corporation, Newfield, NY.

Savage, L. J. 1954. *The foundations of statistics*. New York: Wiley.

Scholz, R.W. 1983. *Decision making under uncertainty*. Elsevier: North-Holland, Amsterdam.

Scott, P.W. 1992. Developing highly accurate empirical weight estimating relationships: Obstacles and tactics. *SAWE Paper No 2091*. Presented at the 51st Annual International Conference of the society of Allied Weight Engineers, Inc. Hartford, Connecticut, May 18-20, 1992.

Shanteau, J. 1987. What about experts? Paper presented at the meeting of *Judgment/Decision Making Society*, Seattle, WA.

Shanteau, J. 1992. Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, Volume 53, pp. 252-266.

Shephard, G.G. 1990. *Guided self-elicitation of decision analyst expertise*, Ph.D. Dissertation. Arizona State University, Tempe, AZ: U.M.I.

Shephard, G.G. and C.W. Kirkwood. 1994. Managing the judgmental probability elicitation process: A case study of analyst/ manager interaction. *IEEE Transactions on Engineering Management*, Vol. 41, No. 4, pp. 414-425.

Spetzler, C.S. and C.-A. S. Stael von Holstein. 1975. Probability encoding in decision analysis. *Management Science*, Vol. 22, No. 3, pp. 340-358.

Stanley, D., R. Unal, and R. Joyner. 1992. Application of Taguchi methods to propulsion system optimization for SSTO vehicles. *Journal of Spacecraft and Rockets*. Vol. 29, No. 4, July-August, 1992, pp. 453-459.

Stanley, D., W. Engelund, R. Lepsch, M. McMillian, K. Wurster, R. Powell, A. Guinta, and R. Unal. 1993. *AIAA-93-1053*, Rocket-powered single-stage vehicle configuration selection and design. AIAA/AHS/ASEE Aerospace Design Conference, February 16-19, 1993, Irvine, California.

Stone, H.W. and W.M. Piland. 1992. An advanced manned launch system concept. *IAF-92-0870*, August, 1992.

Turban, E.. 1992. *Expert systems and applied artificial intelligence*. New York: Macmillan.

Tversky, A. and D. Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin*, Vol. 76, No. 2, pp. 105-110.

Tversky, A. and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, Vol. 4, pp. 207-232.

Tversky, A. 1977. On the elicitation of preferences: Descriptive and prescriptive considerations. In *Conflicting objectives in decisions*. David E. Bell, Ralph L. Keeney, and Howard Raiffa, editors. New York: Wiley and Sons.

Unal, R., D.O. Stanley, W. Engelund, and R. Lepsch. 1994. Design for quality using response surface methods: An alternative to Taguchi's orthogonal arrays. *Engineering Management Journal*, Vol. 6, No. 3.

Van de Ven, A.H. and A.L. Delbecq. 1971. Nominal versus interacting group processes for committee decision making effectiveness. *Academy of Management Journal*, Vol. 14, pp. 203-213.

von Neumann, J. and O. Morgenstern. 1944. *The theory of games and economic behavior*. Princeton, N.J.: Princeton University Press.

Wallsten, T. S., D. Budescu, A. Rapoport, R. Zwick and B. Forsyth. 1986. Measuring the vague meanings of probability terms. *Journal of Experimental Psychology*, Vol. 115, No. 4, pp. 348-365.

Wilder, J. 1996. Personal communication by phone with Richard Monroe, October 9, 1996.

Wille, R.H. 1990. Landing gear weight optimization using Taguchi analysis. *SAWE Paper No.1966*, Category No. 11. Paper presented at the 49th Annual International Conference of Society of Allied Weight Engineers, Inc., May 14-16, 1990, Chandler, Arizona.

Williams, R. 1989. Weight and cost forecasting for advanced manned space vehicles. Final report: NASA/ASEE Summer Faculty Fellowship Program. Contract: NGT 44-001-800. Conducted at NASA Johnson Space Center, Houston, Texas.

Williams, K. and F. Freiman. 1989. Probabilistic risk analysis in conjunction with Fast-E parametric cost estimating. *Proceedings of International Society of Parametric Analysts*. pp. 550-565.

Winkler, R.L. 1986. Expert resolution. *Management Science*. Volume 32, No. 3, pp.298-306.

Wolfson, L.J. 1996. Eliciting prior distributions the predictive way. Presented at the conference on *International Business Logistics, INFORMS Atlanta, Fall, 1996.*

Zimmer, A.C. 1983. Verbal vs. numerical processing of subjective probabilities. In *Decision making under uncertainty*, Roland W. Scholz [editor], pp. 159-182. Elsevier: North-Holland, Amsterdam.

Appendix A

Preliminary Questionnaire for Simplified Case

RATING SHEET FOR SUBSYSTEM WEIGHT RANGES

Vehicle: unmanned ssv dual-fuel, sd-701, horz. 30 ft p/l bay,
25 klb p/l - 51.6 inc.

PLEASE Provide a midpoint, a lower range value and an upper range value for each of the following subsystems.
Use the rating sheet for uncertainty as a guide to determine how broad the range should be for a given subsystem.

Subsystems

		Low	Mid	High
1.0	Wing			
2.0	Tail			
3.0	Body			

RATING SHEET FOR SUBSYSTEM WEIGHT UNCERTAINTY

Vehicle: unmanned ssv dual-fuel, sd-701, horz. 30 ft p/l bay,
25 klb p/l - 51.6 inc.

FOR WEIGHT ESTIMATING:
PLEASE RATE EACH OF THE FOLLOWING SUBSYSTEMS ON THE SCALE 1 TO 5 WITH 1 BEING LOW UNCERTAINTY AND 5 BEING HIGH UNCERTAINTY. MODERATE UNCERTAINTY WOULD BE RATED 3. THIS EXERCISE IS INTENDED TO PROVIDE A RELATIVE MEASURE OF UNCERTAINTY AGAINST OTHER SUBSYSTEMS.
Circle any lower level subsystems that are the primary source of uncertainty in a given subsystem.

Based on your judgment:
Of all the subsystems which one is the most uncertain with regard to weight?
__________ Rate that subsystem now.
Which subsystem has the least uncertainty with regard to weight? __________ Rate that subsystem now. Use these two anchors to rate the other subsystems as high, low or moderate.

Subsystems

1.0	Wing		1	2	3	4	5
		Exposed wing surface (expwing)					
		Carry-through (cthru)					
2.0	Tail		1	2	3	4	5
3.0	Body		1	2	3	4	5
		LH2 tank					
		Kerosene tank					
		LO2 tank					
		Basic structure					
		Secondary structure					

UNCERTAINTY VALUATION

Your understanding of high uncertainty would be associated with what confidence level? In other words - what percent is uncertain?

20% ***30%*** ***40%*** ***50%*** ***More***

Your understanding of low uncertainty would be associated with what confidence level? Again, express this in the percent that is uncertain.

1% ***5%*** ***10%*** ***15%*** ***More***

Your understanding of moderate uncertainty would be expressed as what percent uncertain?

10% ***15%*** ***20%*** ***25%*** ***More***

Appendix B

Rationale for Reasons and Cues Documentation

The following document was sent to the expert at NASA LaRC on February 28, 1996:

2/28/96
Roger:
The following should further explain why I have added "CUES" as another element of the questionnaire.

First from an article by:
Ettenson, R., Shanteau, J. and Krogstad, J. 1987. Expert judgment: Is more information better? *Psychological Reports*, 60, 227-238.

The abstract reads as follows:
"Two groups of professional auditors (expert ns = 10 and 11) and one group of 11 accounting students (novices) made judgments for 32 hypothetical auditing cases which were based on 8 dimensions of accounting-related information. Analyses indicated that the experts die not differ significantly from the novices in the number of significant dimensions: both the professionals and the students had roughly three significant factors. When evaluating the information, however, the experts' judgments primarily reflected one source of information, with other cues having secondary impact. In comparison, no single cue was dominant for the students' judgments. These results were interpreted to indicate that the nonuse of information by experts does not necessarily indicate a cognitive limitation. Instead, experts have better abilities to focus on relevant information. The professional auditors also exhibited greater consistency and consensus than did the students. In contrast to much previous work, the experts here are viewed as being skilled and competent judges."

The SCENARIO:
"Normal audit procedures lead you to believe that the year-end 'Allowance for doubtful accounts' should be increased."

The 8 dimensions (or 8 cues) are:
1. Company is nondiversified in declining industry with sales declining at 15% annually.
2. Co. is closely held corp. with creditors as primary users of financial statements.
3. Co.'s management is less than completely cooperative and open with you during audit.
4. Co.'s management has conservative accounting policies and reported earnings are high quality.
5. Your review disclosed no material weaknesses in accounting practices.
6. Proposed adjustment reduces current ratio from 2.1 - 1 to 1.7 to 1 (industry is 2 - 1 typically).
7. Proposed adjustment will decrease current income after taxes by 2.7%.
8. The after-tax impact will reverse an otherwise upward earnings per share trend that has prevailed for the preceding three years.

While the accounting scenario was set up for experimental purposes, this should give you an idea of the types of things that might cue a decision in a particular direction. In essence, the request for "cues" may be redundant with reasons but the intent is to determine how much information you are using to make your uncertainty rating. This may also take the form of PRIMARY and SECONDARY information (reasons and/or cues). The intent is also to prompt you to think of additional information that you are actually thinking about that influences your uncertainty rating and document those as either reasons or cues. This serves to document your knowledge that might otherwise be lost in the process.

Appendix C
Part 1

Instructions, Uncertainty Rating Questionnaire and Parameter Three Point Levels Questionnaire

INSTRUCTIONS FOR QUESTIONNAIRE
(prior to final revision)

1. Rate subsystems for MOST UNCERTAIN and LEAST UNCERTAIN to prioritize the subsequent parameter evaluations.

2. Rate WER uncertainty QUALITATIVELY from Low, Moderate to High uncertainty. Focus only on those WER parameters that you feel should be evaluated in this manner.

3. Anchor your QUALITATIVE measure of uncertainty to a QUANTITATIVE measure on the 5-point scale provided.

4. Provide 3 point estimates [Low, Mode or Most Likely, and High] for each of the MOST UNCERTAIN WER parameters identified in the preceding steps.

5. Describe the reason for the uncertainty and the reasoning behind the parameter value ranges for the MOST UNCERTAIN WERs.

6. Describe any scenarios that may change WER PARAMETER values. Provide the alternative WER PARAMETER values that in your judgment would be appropriate for the scenario.

FINAL REVISED
INSTRUCTIONS FOR QUESTIONNAIRE

1. Rate WER parameter uncertainty QUALITATIVELY from Low, Moderate to High uncertainty (and the 2 intermediate ratings for a total of 5 possible ratings). Focus only on those WER parameters that you feel should be evaluated in this manner.

2. Describe the reason for the uncertainty and the reasoning behind the parameter value ranges for the UNCERTAIN WERs that you rated. Do this simultaneously while rating each WER parameter to document your thinking.

3. Think of any other cue (or reason that you have not documented) and record that information at this time.

4. After rating all WER parameters, next anchor your QUALITATIVE measure of uncertainty to a QUANTITATIVE measure on the 5-point scale provided.

5. Provide 3 point estimates [Low, Mode or Most Likely, and High] for each of the MOST UNCERTAIN WER parameters identified in the preceding steps.

6. Describe any scenarios that may change WER PARAMETER values. Provide the alternative WER PARAMETER values that in your judgment would be appropriate for the scenario.

Qualitative WER Uncertainty

Focus on the UNCERTAIN Subsystem WERs and rate each WER for the amount of uncertainty.

The rating choices are LOW, 2, MODERATE, 4, HIGH and None.

Choose Low, Moderate or High based on the level of Uncertainty that you feel applies to that particular subsystem WER.

Choose 2 if Uncertainty is more than Low but less than Moderate.

Choose 4 if Uncertainty is more than Moderate but less than High.

Choose NONE if the WER is constant or 100% certain.

Provide a Quantitative explanation of your understanding of Low, Moderate and High uncertainty.

The amount of uncertainty or variation that I associate with Low Uncertainty is:

Less5% 7.5% 10% 15% 20% More

The amount of uncertainty or variation that I associate with High Uncertainty is:

Less15% 20% 30% 40% 50% More

The amount of uncertainty or variation that I associate with Moderate Uncertainty is:

Less 10% 15% 20% 25% 30% More

For ratings of 2 or 4 on the Qualitative rating sheet:

the midpoint between Low and Moderate will be used for a 2 rating

the midpoint between Moderate and High will be used for a 4 rating

3 Point Estimates
for WER parameters

Provide 3 point estimates for each of the WERs for the MOST UNCERTAIN Subsystems.

The 3 points should be the MINIMUM, MODE (MOST LIKELY) and MAXIMUM values for the WER parameter.

The nominal case is listed on the questionnaire as the MODE. If this is not a correct assumption, make the necessary adjustment by crossing out the number and providing 3 point estimates for that WER parameter.

Review your Qualitative rating of the WER parameter when assigning the 3 point values. The percent of Uncertainty can be considered as the percent of potential variation in the parameter values.

Describe the reasons for Uncertainty on a form in the Questionnaire section immediately following this section.

WING subsystem
Select the WER parameters from the following list that you want to evaluate for uncertainty.

(expwing) parameters

'c'	'1'	1.0	constant
'c'	'c1'	.82954	equation coefficient
'c'	'c2'	.001	divide load by 1000
'c'	'usf'	1.75	ultimate safety factor
'c'	'lf'	2.0	load factor
'v'	'wland'		landed wt
'a'	'exp'	3360.	exposed wing area
'c'	'ar'	1.48	aspect ratio based on exposed area
'c'	'tr'	.34	taper ratio ct/cr
'c'	'toc'	.10	thickness to chord ratio
'c'	'e1'	.48	exponent
'c'	'e2'	.67	exponent
'c'	'e3'	.64	exponent
'c'	'e4'	.40	exponent
'c'	'rew'	.40	reduction factor (lo2-lh2 ssv, ezedesit, Gr/Ep)

From the WING (expwing) WER parameters you have selected:

c	eq. coef.	0.82954

Rate the degree of uncertainty that you associate with this parameter:

Low	2	Moderate	4	High

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

Questionnaire (by WER for parameter value ranges)
for ssv dual-fuel, rd-701, horz. 30 ft p/l bay, 25klb p/l - 51.6 inc.

Wing		Low	Mode	High	
c	usf		1.75		ultimate safety factor
c	nf		2.0		load factor
c	ar		1.48		aspect ratio
c	tr		0.34		taper ratio
c	toc		0.10		thickness to chord ratio
c	eq. coef.	______	0.82954	______	
c	rew		0.40		reduction factor
a	exp		3360		exposed wing area
s	wland				

cthru		Low	Mode	High	
c	usf		1.75		ultimate safety factor
c	nf		2.0		load factor
c	ar		1.48		aspect ratio
c	tr		0.34		taper ratio
c	toc		0.10		thickness to chord ratio
c	eq. coef.	______	319.29	______	
c	rct		0.40		reduction factor
s	wland				

tail		Low	Mode	High	
c	eq. coef.	______	5.00	______	
c	rtf		0.10		reduction factor
a	exp		3200		exposed wing area

Body (LH2 tank)

lh2tstr		Low	Mode	High	
c	c		0.364		unit wt of tank
c	d_lh2		4.43		LH2 density
s	wlh2				lh2 prop weight
c	ull		0.0425		tank ullage fraction
c	rlh2tnk		0.		reduction tank

lh2ins		Low	Mode	High	
c	c		0.286		unit wt of insulation
c	c0		4.3169		k factor const. term
c	c1		0.50194		k factor linear term
c	d_lh2		4.43		LH2 density
s	wlh2				lh2 prop weight
c	ull		0.0425		tank ullage fraction
l	wb		0.		body width (ft)
c	rlh2ins		0.0		reduction factor

SCENARIOS

Describe any scenarios that may change WER parameter values.
Provide the alternative WER parameter value ranges that apply to the scenario.

SCENARIO

ALTERNATIVE WER parameter values

Appendix C
Part 2

Uncertainty Ratings, Reasons and Cues

[Bold and italicized statements are ratings, reasons and cues provided by expert]

WING subsystem
Select the WER parameters from the following list that you want to evaluate for uncertainty.

(expwing) parameters

'c'	'1'	1.0	constant
'c'	'c1'	.82954	equation coefficient
'c'	'c2'	.001	divide load by 1000
'c'	'usf'	1.75	ultimate safety factor
'c'	'nf'	2.0	load factor
'v'	'wland'		landed wt
'a'	'exp'	3360.	exposed wing area
'c'	'ar'	1.48	aspect ratio based on exposed area
'c'	'tr'	.34	taper ratio ct/cr
'c'	'toc'	.10	thickness to chord ratio
'c'	'e1'	.48	exponent
'c'	'e2'	.67	exponent
'c'	'e3'	.64	exponent
'c'	'e4'	.40	exponent
'c'	'rew'	.40	reduction factor (lo2-lh2 ssv, ezedesit, Gr/Ep)

From the WING (expwing) WER parameters you have selected:

c	eq. coef.	0.82954

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 High

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

"For conceptual design, WERs for wings are typically more accurate than for other components."

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

"1. WER is based on a regression of historical data points."
"2. Fit to data is good."
"3. Data points are applicable to vehicle type."

"Size of applicable data set.
Basis of weight (actual, calculated, estimated)."

From the WING (expwing) WER parameters you have selected:

c	rew	0.40	reduction factor

Rate the degree of uncertainty that you associate with this parameter:

Low	2	Moderate	4	High

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

"Reduction factor is for the use of composites. Little historical data exists for composite structure usage."

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

"1. Reduction factor has not been validated with actual structures.
2. Factor represents changes in construction type as well as material."

Wing cthru subsystem
Select the WER parameters from the following list that you want to evaluate for uncertainty.

(cthru) parameters

'c'	'1'	1.0	constant
'c'	'.5'	.5	constant
'c'	'2'	2.0	constant
'c'	'c1'	319.29	equation coefficient
'c'	'c2'	.001	divide load by 1000
'c'	'usf'	.175	ultimate safety factor
'c'	'nf'	2.0	load factor
's'	'wland'		landed wt
'I'	'wc'	36	carry-through width
'I'	'bs'	87.94	structural span
'c'	'ar'	1.48	aspect ratio
'c'	'tr'	.34	taper ratio
'I'	'wspan'	70	exposed wing span
'c'	'toc'	.10	thickness to chord ratio
'I'	'rootc'	80.0	root chord (exp wing)
c'	'mc'	1.66e-5	rho/sigma material const. aluminum
'c'	'gp'	1.14	geometric parameter
'c'	'el'	.50	exponent
'c'	'rct'	.40	reduction factor (lo2-lh2 sav, ezedesit, Gr/Ep)

From the cthru WER parameters you have selected:

c	eq. coef.	319.29

Rate the degree of uncertainty that you associate with this parameter:

Low *2* Moderate 4 High

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

"WER formulated specifically for this vehicle type using semi-analytical approach."

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

"1. Regression of historical data points.
2. Excellent fit to data.
3. Low number of data points.
4. Data points are estimates, not actual weights."

Body (LH2 tank)
lh2tstr subsystem
Select the WER parameters from the following list that you want to evaluate for uncertainty.

'lh2tstr' parameters

'c'	'1'	1.0	const
'c'	'c'	.364	unit wt of tank (lb/ft³) (lo2-lh2 sav, cxedesit, Al-Li)
'c'	'd_lh2'	4.43	lh2 density (lb/ft³)
's'	'wlh2'		lh2 prop weight
'c'	'ull'	.0425	tank ullage fraction
'c'	'rlh2tnk'	0.	reduction factor

From the lh2tstr WER parameters you have selected:

c	c	0.364	unit wt of tank

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 High

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

"An over-simplified WER and the use of a new material (Al-Li) lead to relatively high uncertainty."

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

"1. Weight derived from FEM analysis with non-optimum factor applied.
2. Scaled by volume only, no other geometry parameters considered.
3. Al-Li material."

Body (Kerosene tank)
hctstr subsystem

Select the WER parameters from the following list that you want to evaluate for uncertainty.

'hctstr' parameters

'c'	'1'	1.0	const
'c'	'c'	.656	unit wt of tank (lb/ft^2)
'c'	'd_hc'	50.5	hydrocarbon density (lb/ft^3)
's'	'whc'		hydrocarbon prop weight
'c'	'ull'	.0425	tank ullage fraction
'c'	'rhctnk'0.		reduction factor

'hcins' parameters

'c'	'1'	1.0	const'
'c'	'c'	0.	unit wt of insulation (lb/ft^2)
'a'	'shctnk'	7760.	hydrocarbon tank area (ft^2)
'c'	'rhcins'	0.0	reduction factor

From the hctstr WER parameters you have selected:

c	c	0.656	unit wt of tank

Rate the degree of uncertainty that you associate with this parameter:

Low 2 *Moderate* 4 High

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

"An inappropriate, but conservative WER and the use of a new material (Al-Li) lead to moderate uncertainty. Uncertainty is reduced with the assumption of a minimum gage structure."

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

"1. Largely minimum gage structure.
2. Scaled by vloume only, no other geometry parameters considered.
3. Al-Li material.
4. Conservative weight calculated when tnak size grows."

Body (Secondary structure)
plstr subsystem
Select the WER parameters from the following list that you want to evaluate for uncertainty.

'plstr' parameters

'c'	'1'	1.0	const
'c'	'c'	6500.	p/l bay/ker. tank support and nose gear bay str. (Robinson est.)
'c'	'rshrd' 0.		reduction factor

From the plstr WER parameters you have selected:

c	c	6500.	p/l bay/ker. tank support and nose gear bay str.

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 ***High***

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

"Lack of definition in structure design and use of new material (Gr-Ep) results in high uncertainty."

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

"1. Provided by another weight analyst.
2. Rough estimate.
3. Highly conservative."

Body (Secondary structure)

hsdstr subsystem

Select the WER parameters from the following list that you want to evaluate for uncertainty.

'hsdstr' parameters

'c'	'1'	1.0	const
'c'	'2'	2.0	const
'c'	'c'	2.50	unit wt of heat shield str (lb/ft^2)
'c'	'pi'	3.1416	
'I'	'wb'		body width
'a'	'sbase'		base area (ft^2)
'c'	'rhtsd' .35		reduction factor

From the hsdstr WER parameters you have selected:

c	c	2.50	unit wt of ht shield str

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 High

Now that you have rated the uncertainty for this WER parameter, please provide a reason or reasons for your rating.

"Use of shuttle data and assumption of similarity results in relatively low uncertainty."

To further document your thinking, please provide any cues (or triggers) that influence your thinking about this parameter.

"1. Shuttle derived.
2. Aluminum structure.
3. Area scaling.
4. Approximation (cut-outs not considered)."

Appendix D

Data Developed from Expert Questionnaire

Table 7

Subsystem WER	Parameter	Pt. Est.	Uncertainty rating	Low value	Mode	High value
Wing	eq. coef.	0.82954	Low	0.7881	0.82954	0.8710
Wing	rew (reduction factor)	0.40	4 (Md-High)	0.32	0.40	0.48
cthru	eq. coef.	319.29	2 (Low-Md)	287.36	319.29	351.22
cthru	rct (reduction factor)	0.40	4 (M-H)	0.32	0.40	0.48
tail	eq. coef.	5.0	2 (L-M)	4.5	5.0	5.5
tail	rtf (reduction factor)	0.10	2 (L-M)	0.08	0.10	0.12
Body - LH2 Tank lh2tstr	c (unit wt of tank)	0.364	4 (M-H)	0.291	0.364	0.437
Body - lh2ins	c (unit wt of insulation)	0.286	Moderate	0.243	0.286	0.329
Body - Kerosene Tank hctstr	c (unit wt of tank)	0.656	Mod.	0.558	0.656	0.754
Body - hcins	none rated					
Body - LOX Tank loxtstr	c (unit wt of tank)	0.458	4 (M-H)	0.366	0.458	0.550
Body - loxins	c (unit wt of insul)	0.232	Mod.	0.197	0.232	0.267
Body - Basic Structure - nose	c (unit wt of structure)	1.11	4 (M-H)	0.888	1.11	1.33
Body - Basic Structure - inter	c (unit wt of structure)	1.64	4 (M-H)	1.31	1.64	1.97
Body - Basic Str. - aftbdy	c (unit wt of struct)	4.0	4 (M-H)	3.2	4.0	4.8
Body - Basic Str. - thrst	c (constant (lb/lb))	0.0021	4 (M-H)	0.0017	0.0021	0.0025
Body - Basic Str. - engbay	c (unit wt of struct)	1.31	4 (M-H)	1.05	1.31	1.57
Body - Secondary Str. - crcab	none rated					
Body - Sec Str - doors	c (3 lb/ft^2 doors, 30 ft length)	2100	Low	1995	2100	2205
Body - Sec Str - plstr	c (p/l bay/ker. tank support and nosegear bay str.)	6500	High	3575	6500	6825
Body -Sec Str - shrd	c (1.0 lb/ ft^2)	1800	4 (M-H)	1440	1800	2160
Body-Sec Str-hadstr	c (unit wt of ht shield str)	2.50	2 (L-M)	2.25	2.50	2.75
Body-Sec Str-hadstr	rhtsd (reduction factor)	0.35	4 (M-H)	not provided		
Body-Sec Str-bflap	c (unit wt of body flap)	3.58	Mod	3.04	3.58	4.12
Induced Environment protection - TPS						
Fuselage - fusetps	c (unit wt of tps (lb/ft^2))	1.152	Mod	0.979	1.152	1.325
Fuselage - fusetps	rmecu (reduction factor)	0.268	4 (M-H)	0.214	0.268	0.322
Wing - wingtps	c (unit wt of tps (lb/ft^2))	1.287	Mod	1.030	1.287	1.480
Wing - wingtps	rwl (reduction factor)	0.268	4 (M-H)	not provided		
Internal insulation Nose ninsul	cnsins (insulation unit wt. - shuttle)	0.75	2 (L-M)	0.675	0.75	0.825

Table 7

Subsystem WER	Parameter	Pt. Est.	Uncertainty rating	Low value	Mode	High value
Payload bay doors - plinsul	cplins (insulation unit wt. - shuttle)	0.23	2 (L-M)	0.207	0.23	0.253
Equipment bays - eqinsul	wequins (equip. bay insul. wt - shuttle)	650.0	Mod	553.0	650.0	748.0
Purge, vent, drn, and hazard gas det - pvd	cbmid (pvd const - mid body, lb/ft)	12.3	2 (L-M)	11.1	12.3	13.5
pvd	cbaft (pvd const - aft body, lb/ft)	18.3	2 (L-M)	16.5	18.3	20.1
pvd	cwng (pvd const - wing, lb/ft)	1.3	2 (L-M)	1.2	1.3	1.4
Undercarriage and aux. systems - Nose gear - Running gear	nsgrg cnrg (running gear const.- nose)	18.9	Mod	16.1	18.9	21.7
nsgstr	cnstr (gear structure const. - nose)	9.48	Mod	8.06	9.48	10.9
nsgstr	rlg (reduction factor)	0.15	2 (L-M)	0.135	0.15	0.165
nsgcntrl	cncntrl (controls constant - nose)	0.08	Mod	0.068	0.08	0.092
Main gear Running gear	cmrg	173.0	Mod	147.0	173.0	199.0
Structure mngstr	cmstr	35.2	Mod	29.9	35.2	40.5
Str. mngstr	rlg	0.15	2 (L-M)	0.135	0.15	0.165
Controls mngcntrl	cmcntrl	0.06	Mod	0.051	0.06	0.069
Propulsion, main Engines eng	towe	*81.1*	2 (L-M)	61.0	70.2	76.4
press	cpf	44.4	Mod	37.7	44.4	51.1
Helium pneumatic and purge system he	chesys	5.92e-4	Low	5.62e-4	5.92e-4	6.22e-4
he	chetnk	15.9	Low	15.1	15.9	16.7
Propulsion,reaction control (RCS)						
Thrusters andsupports Fwd	nthrstr	9	Low	8	9	10
Fwd	wthstr	5.3	4 (M-H)	4.2	5.3	6.4
Aft	nthstrv	12	Low	11	12	13
Aft	nthstrp	18	Low	17	18	19
Aft	wthstrv	5.3	4 (M-H)	4.2	5.3	6.4
Aft	wthstrp	22.0	4 (M-H)	17.6	22.0	26.4
Propellant tanks rcstanks	ctnk	0.34	2 (L-M)	0.31	0.34	0.37
Distribution and recir-culation distr	cdistr	1304.0	4 (M-H)	1043.0	1304.0	1565.0
distr	crecirc	5	4 (M-H)	4	5	6
Valves	cvalves	569.0	4 (M-H)	455.0	569.0	683.0

Table 7

Subsystem WER	Parameter	Pt. Est.	Uncertainty rating	Low value	Mode	High value
Propulsion, orbital maneuver (OMS)						
Engines omseng	ceng	181.8	Mod	154.5	181.8	209.1
Propellant tanks omstanks	ctnk	0.037	2 (L-M)	0.033	0.037	0.041
omstanks	isp [*see note]	462.2	Low	453.0	462.2	471.4
Pressurization omspras	chetnk	1.12	Mod	0.95	1.12	1.29
Prime power Fuel cell system fcell	c	3.70	Mod	3.15	3.70	4.26
fcell	pfc	240.0	Mod	204.0	240.0	276.0
Reactant dewars dewar	c	0.99	2 (L-M)	0.89	0.99	1.09
dewar	rrdw (reduction factor)	0.10	Low	0.095	0.10	0.105
Batteries	none rated					
Electric conversion and distr. powcon	pfcnom	22.0	Mod	12.8	15.0	17.3
powcon	cpc	81.2	Mod	69.0	81.2	93.4
powcon	cinst	1.75	Mod	1.49	1.75	2.01
powcon	rpc (reduction factor)	0.20	High	0.15	0.20	0.25
Circuitry Elect. pwr dist and cntrl epdc	cepdc	81.6	Mod	69.4	81.6	93.8
epdc	pfcnom	22.0	Mod	12.8	15.0	17.3
epdc	cinst	1.23	Mod	1.05	1.23	1.41
epdc	repdc	0.10	High	0.075	0.10	0.125
Avionic cabling avcable	wac (avionic cable wt-shuttle)	2565.0	4 (M-H)	2052.0	2565.0	3072.0
avcable	winst (supports and installation wt. - shuttle)	564.0	4 (M-H)	451.0	564.0	677.0
avcable	rcab (reduction factor -fiber optics)	0.30	2 (L-M)	0.27	0.30	0.33
avcable	rinst (reduction factor supports and installation)	0.20	2 (L-M)	0.18	0.20	0.22
RCS cabling rcscab	wrcscab (rcs cabling wt-shuttle)	89.0	4 (M-H)	71.0	89.0	107.0
rcscab	rcab (reductn fact. fiber optics)	0.30	2 (L-M)	0.27	0.30	0.33
OMS cabling omscab	womscab	276.0	4 (M-H)	221.0	276.0	331.0
omscab	rcab (reductn fact. fiber optics)	0.30	2 (L-M)	0.27	0.30	0.33
Connector plates conplt	wconn	207.0	Mod	176.0	207.0	238.0
Wire trays tray	wtrays (wire trays wt -shuttle)	592.0	4 (M-H)	474.0	592.0	710.0
tray	rtray (reduction factor)	0.20	2 (L-M)	0.18	0.20	0.22

Table 7

Subsystem WER	Parameter	Pt. Est.	Uncertainty rating	Low value	Mode	High value
emcable	ktvc (pwr/wt for gimbal actuators, peak)	0.105	Mod	0.089	0.105	0.121
emcable	kcs (pwr/wt for control surface actuators, peak)	0.035	Mod	0.030	0.035	0.040
EMA control units emcon	cem (actuator and motor contrl unit const.)	8.0	4 (M-H)	6.4	8.0	9.6
emcon	ktvc (pwr/wt for gimbal actuators)	0.105	Mod	0.089	0.105	0.121
emcon	kcs (pwr/wt for control surface actuators, peak)	0.035	Mod	0.030	0.035	0.040
emcon	remcu (reduction factor)	0.10	High	0.075	0.10	0.125
Hydraulic conversion and distr.	[no weight allowance]					
Control surface actuation Elevons el act	kel (elevon const., ema)	0.0043	Mod	0.0037	0.0043	0.0049
el act	relact (reduction factor)	0.10	High	0.075	0.10	0.125
Tip fins tf act	ktf (rudder const. ema)	0.0036	Mod	0.0031	0.0036	0.0041
tf act	rtfact (reduction factor)	0.10	High	0.075	0.10	0.125
Body flap bf act	kbf (body flap const., ema)	0.0040	Mod	0.0034	0.0040	0.0046
bf act	rbfact (reduction factor)	0.10	High	0.075	0.10	0.125
Avionics guid., nav., and cntrl. gnc	rgnc (reduction factor)	0.73	Mod	0.62	0.73	0.84
Comm. and tracking comtrk	rcts (reduction factor)	0.75	Mod	0.64	0.75	0.86
Displays and contrl.	none rated					
Instrum. system instr	ris (reduction factor)	0.456	Mod	0.388	0.456	0.524
Data processing dproc	rdps (reduction factor)	0.751	Mod	0.638	0.751	0.864
Environmental control Personnel system perr	wi (invariant wt (lb/man)	81.0	2 (L-M)	72.9	81.0	89.1
perr	c (constant lb/man-hr)	0.295	2 (L-M)	0.266	0.295	0.325
perr	renv (reduction factor)	0.10	High	0.075	0.10	0.125
Equipment cooling eqcool	cec (cabin environ constant - lb/kw)	41.4	Mod	35.2	41.4	47.6
eqcool	pfcnom (nominal fuel cell power req - kw)	*22.0*	Mod	12.8	15.0	17.3
eqcool	renv (reduction factor)	0.10	High	0.075	0.10	0.125
Heat transport loop loop	cht (freon loop const)	0.386	2 (L-M)	0.347	0.386	0.425
loop	pfcnom (fuel cell nom pwr kw)	*22.0*	Mod	12.8	15.0	17.3
loop	renv (reduction factor)	0.10	High	0.075	0.10	0.125
Heat rejection system Radiators rad	crad (radiator const lb/ft^2)	0.805	Low	0.765	0.805	0.845

Table 7

Subsystem WER	Parameter	Pt. Est.	Uncertainty rating	Low value	Mode	High value
rad	renv (reduction factor)	0.10	High	0.075	0.10	0.125
Flash evaporator system evap	cfe (flash evaporator constant lb/kw)	7.36	Low	6.99	7.36	7.73
evap	pfcnom (fuel cell nom pwr kw)	***22.0***	Mod	12.8	15.0	17.3
evap	ctnk (water tank const.)	0.048	Mod	0.041	0.048	0.055
evap	cinst (supprts andinstall. factor)	1.10	2 (L-M)	0.99	1.10	1.21
evap	renv (reduction factor)	0.10	High	0.075	0.10	0.125
Personnel provisions	none rated					
seats	none rated					
Payload provisions	none rated					
Margin	none rated					

*Note: The expert recorded this note for the omstnks isp parameter:
"extra low uncertainty, 2% (could use skewness here)"

Note: The expert over ruled the point estimate for all point estimates that are italicized and bold. He then provided a range of three estimates that excluded the point estimate entirely.

Appendix E

Statistical Analysis and
Summary Tables for Analysis of Variance

Appendix E

Statistical Analyses

Analysis of variance (ANOVA) was conducted to ascertain the optimum number of iterations for the Monte Carlo simulation procedure. When the factor was specified at three levels (e.g. 500, 1000, and 2000) ANOVA tests the hypothesis:

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: not all μ_i are equal

Rejecting the null hypothesis indicates that there is a treatment effect for the factor being analyzed. Failure to reject the null hypothesis means there is no treatment effect for the factor analyzed.

The anticipated results for the analysis of the simulation system parameters were that only the statistical distribution used for sampling distribution of the random number generator would show a significant treatment effect. There would be no treatment effect for number of iterations and no treatment effect for the random number seed. These *a priori* expectations were based on the intuitive difference between the triangular and normal distributions. Prior results of Monte Carlo simulations suggested that convergence would occur early, between 500 and 1000 iterations, so no treatment effect was anticipated for number of iterations.

E.1 Analysis of Variance for Simulation System Parameters

E.1.1 Number of Iterations

Analysis of variance (Anova) was performed utilizing simulation outputs as the data being analyzed. The analysis was conducted in order to optimize the system parameters for the Monte Carlo simulation. The initial Anova's were performed to determine the optimal number of iterations to be executed. The first Anova evaluated the mean with iterations varied at three levels - 500, 1000, and 2000 iterations while holding the random number generator constant using the triangular distribution and the seed values were matched for each iteration level so there was no variation attibutable to different seeds.. For the factor, "number of iterations", there was no treatment effect (p-value = 0.769533 and F = 0.264529) at the $\alpha = 0.05$ significance level.

The random number seeds were presented in Law and Kelton (1991, p.450) as suggested by Marse and Roberts (1983). The first ten seeds from the string of seed was utilized to perform the series of ten simulations.

The nature of simulation tends to reaffirm the mean value rather than to promote differences in mean values when the model is relatively stable. In view of this fact, other statistics were evaluated to determine the effect of the number of iterations on those statistics. In particular, the standard deviation, the mode, the maximum and the minimum values from the simulation results were analyzed. These statistics were also analyzed with the factor, "number of iterations", varied at 500, 1000, and 2000 iterations while again holding the random number generator constant using the triangular distribution and the seed values were again matched so there was no variation attibutable to different seeds. No factor effect was found for the standard deviation (p-value = 0.994194), for the mode (p-value = 0.356416), and for the maximum (p-value = 0.109997) at the $\alpha = 0.05$ significance level. For the minimum value, a factor effect was evident for the number of iterations (p-value = 0.0346 and F = 3.819702 with F-critical = 2.51061) at the $\alpha = 0.05$ significance level.

After determining that there was a treatment effect for "number of iterations" when evaluated at three factor levels, additional Anova's were conducted with only two levels of the factor. First the number of iterations was evaluated at 1000 and 2000 iterations with the random number generator held constant using the triangular distribution. There was no treatment effect when the minimum value was evaluated at these two levels (1000 and 2000) (p-value = 0.612385) at the $\alpha = 0.05$ significance level.

The next evaluation was conducted for 500 and 1000 iterations. There was no treatment effect when the minimum value was evaluated at these two levels (500 and 1000) (p-value = 0.067) at the $\alpha = 0.05$ significance level.

Next the extremes of the three iteration levels were analyzed. There was a treatment effect for the "number of iterations" when the minimum value was evaluated for 500 and 2000 iterations (p-value = 0.023259 and F = 6.149859 with F-critical = 4.413863).

Since a factor effect was found when comparing 500 and 2000 iterations for the minimum value, additional evaluations were conducted for the mean and the maximum at these same levels. No treatment effect for "number of iterations" was found for either the mean (p-value = 0.481905) or the maximum value (p-value = 0.057335) using an $\alpha = 0.05$ significance level. One reason for conducting the risk analysis simulation is to arrive at a range of probable values. The fact that there was a treatment effect for "number of iterations" when the minimum value was analyzed suggests that 2000 iterations is preferred over 500 or 1000 iterations. The existence of a factor effect is due to the simulation results which consist of an average minimum value of 170,242.6 pounds for 500 iterations (n=10) and an average minimum value of 166,976 pounds for 2000 iterations (n=10). The maximum value analysis did not result in a treatment effect for "number of iterations" but the Anova showed only a slight difference in favor of the null hypothesis, no treatment effect, at the $\alpha = 0.05$ significance level with a p-value of 0.057335. The average maximum value of 240,425.5 pounds for 500 iterations (n=10) and the average maximum value of 244,091.2 pounds for 2000 iterations (n=10) demonstrates the magnitude of variance at the two levels of "number of iterations". Performing 2000 iterations will provide a broader range of values than will simulations conducted with only 500 or 1000 iterations.

The next set of simulations were conducted using 5000 iterations. The results from these simulations were analyzed against the earlier simulations at 2000 iterations to determine if these two levels of "number of iterations" resulted in a factor effect. The random number generator was again held constant using the triangular distribution and the seed values were again matched so there was no variation attibutable to different seeds. The Anova's evaluating "number of iterations" at these two levels were conducted using the minimum, the maximum, the mean, the standard deviation and the mode. The only treatment effect was found for "number of iterations" when evaluating the maximum value (p-value = 0.037238 and F = 5.059931 with F-critical = 4.413863) using an $\alpha = 0.05$ significance level.

This treatment effect was significant at the $\alpha = 0.05$ significance level despite a lower magnitude of difference than was seen for the maximum values at 500 and 2000 iterations.

The average maximum value of 240,425.5 pounds for 500 iterations (n=10) and the average maximum value of 244,091.2 pounds for 2000 iterations (n=10) did not exhibit a significant factor effect. The average maximum value of 244,090.5 pounds for 2000 iterations (n=10) and the average maximum value of 246,751.1 pounds for 5000 iterations (n=10) did exhibit a treatment effect for "number of iterations". The percent change for the 500 versus 2000 iterations analysis is 1.52% while the percent change for the 2000 versus 5000 iterations analysis is 1.09% for the maximum values. Similarly, the percent change for the minimum values is 1.92% for the 500 versus 2000 iterations analysis while the percent change is 0.94% for the 2000 versus 5000 iterations analysis. Taking several factors into consideration - the lower magnitude of change in the maximum value, the minimum value being far from a significant factor effect, no other statistic resulting in a factor effect between 2000 and 5000 iterations, and the economy of computer time - 2000 iterations is recommended over 5000 iterations for the Monte Carlo simulations. This decision is consistent with convergence thresholds that are used in Monte Carlo programs such as @Risk[R]. Typically, the convergence threshold is set to monitor the statistic and to check for a 1% change in the statistic at regular intervals throughout a simulation. The minimum and maximum values are the statistics of interest in these simulations. In the comparison of 500 and 2000 iterations, both statistics exhibit a change greater than 1% (1.92% for the minimum and 1.52% for the maximum). Therefore, the higher number of iterations is warranted since the 1% threshold (or difference) is not satisfied until the higher number is reached and the simulation has converged to a stable state. In the latter comparison of 2000 and 5000 iterations, one statistic, the minimum, exhibits a change of less than 1% (0.94%) and the other, the maximum, exhibits a change of only slightly more than 1% (1.09%). Based on this slight difference from the threshold, and the fact that one statistic is below the threshold, the lower number of iterations is a suitable choice. The 2000 iteration simulations show acceptable convergence when compared to the 5000 iteration simulations.

E.1.2 Alternate Method

Morgan and Henrion (1990) offer an alternative method for determining sample size or number of iterations for simulation procedures. They suggest running a short simulation to determine the sample variance. With this variance, a given confidence level (i.e. the corresponding standardized Z value), and a given number of class intervals, the sample size can be calculated by:

$$m>(2cs/w)^2$$

This technique was applied to the simulation data which was analyzed using ANOVA above. For example, the mean empty weight for 500 iterations, the sample variance was 230,699.4 and this calculation resulted in 8862 as the appropriate number of interations. The following table, E-1, summarizes the results of this technique using 95% confidence (Z=1.96) and 20 class intervals:

Table E-1 Summary of Alternate Method Calculations

# of iterations	sample variance	sample std. dev.	$m > (2cs/w)^2$
500	230,699.4	480.31	8862
1000	146,920.4	383.302	5644
2000	60,594.59	246.16	2327
5000	45,352.73	212.96	1742

Based on these results, this technique did not seem well-suited for selecting an economical and efficient simulation length based on a very limited simulation (e.g. the variance for 500 iterations). Longer simulations with reduced variance did appear to result in a more reasonable number of iterations. In particular, the m > 2327 and the m >

1742 appeared to be consistent with the ANOVA results that suggest that 2000 iterations was an appropriate number of iterations. The m > 1742 for the 5000 iteration sample variance appeared to confirm that 5000 iterations were unnecessary. Based on the m > 2327, the number of iterations might be set at 2500 rather than the 2000 that were suggested based on the ANOVA results.

E.1.3 Random Number Generator

Following the conclusion of the simulations and the Anova's utilizing the triangular distribution random number generator (RNG), another series of simulations were conducted utilizing a gaussian (or Normal distribution) RNG. The mean value was evaluated to determine if there was a factor effect for the "RNG" factor. Both RNGs, triangular and gaussian, were used to execute simulations for 2000 iterations using identical seeds again to control for variation due to seed values. For equal sample sizes (n=15), there was a treatment effect for RNG when evaluating the mean value with a p-value of 0.000000327 (F = 46.1771 with F-critical = 4.2252).

This treatment effect was as hypothesized (i.e. there will be a treatment effect for RNG) and was as expected since the triangular distribution incorporates skewness rather than symmetry. The triangular distribution was incorporated in the methodology to allow for skewness and to avoid the assumption of normality that is so often invoked. This Anova result served as a statistically significant argument against assuming a normal distribution for the simulations described in this research and provided a warning for other simulation problems as well.

E.1.4 Simulation Follow-up

After 2000 iterations had been selected as the optimal simulation length, a discussion with an outside reviewer led to the advice "Don't skimp on the number of iterations in your simulation" (Wilder 1996). Heeding this advice, a series of follow-up simulations were conducted using 20,000 iterations. Several Anova's were executed to analyze the results of these longer simulations.

The first Anova was a repeat of the preceding analysis. The RNG factor was evaluated using the mean value. Both RNGs, triangular and gaussian, were used to execute simulations for 20,000 iterations using identical seeds again to control for variation due to seed values. For equal sample sizes (n=6), there was a factor effect for RNG when evaluating the mean value with a p-value=0.00000351 (F = 126.4738 with F-critical = 5.317645). This effect is similar to the treatment effect that was determined for 2000 iterations. There is some erosion in the difference between the means as evidenced by the change in the p-value (by a factor of 10) but there is ample evidence against the null hypothesis (H_0: $\mu_1 = \mu_2$) at the $\alpha = 0.05$ significance level.

Next, the two RNGs were evaluated separately to check for a treatment effect for number of iterations between 2000 and 20,000. For the gaussian RNG, no treatment effect was detected for "number of iterations" when evaluating the mean value (p-value = 0.216087, F = 1.743753 and F-critical = 4.964591) at the $\alpha = 0.05$ significance level. For the triangular RNG, no treatment effect was detected for "number of iterations" when evaluating the mean value (p-value = 0.06566, F = 4.270713 and F-critical = 4.964591) at the $\alpha = 0.05$ significance level.

From these analyses, the choice of including the triangular distribution in the methodology was reaffirmed. The follow-up analyses for the "number of iterations" exhibited no treatment effect between the different levels of 2,000 and 20,000. This confirmed the choice and recommendation of 2000 as an acceptable number of iterations for this simulation procedure. Summary tables of the ANOVA analyses are presented in the following tables - E-2 and E-3.

Table E-2 Evaluation of # of iterations for Triangular RNG Simulation

Filename	Parameter	Factor levels analyzed	p-value	at α =.05
nasaaov1	Mean	500, 1000, 2000 iterations	0.769533	not significant
nasaaov2	Std. Dev.	500, 1000, 2000 iterations.	0.994194	not significant
nasaaov3	Mode	500, 1000, 2000 iterations	0.356416	not signifcant
nasaaov4	Maximum	500, 1000, 2000 iterations	0.109997	not significant
nasaaov5	Minimum	500, 1000, 2000 iterations	0.0346	* significant
nasaaov6	Minimum	1000 and 2000 iterations	0.612385	not significant
nasaaov7	Minimum	500 and 1000 iterations	0.067	not significant
nasaaov8	Minimum	500 and 2000 iterations	0.023259	* significant
nasaaov9	Mean	500 and 2000 iterations	0.481905	not significant
nasaov2a	Variance	500, 1000, 2000 iterations	0.988852	not significant
nasaaovv	Minimum	500 and 2000	0.023	* significant
nasaaovx	Maximum	500 and 2000	0.057	not significant
nasaovmx	Maximum	2000 and 5000	0.037	* significant
nasaovmn	Minimum	2000 and 5000	0.1623	not significant
nasaovme	Mean	2000 and 5000	0.553588	not significant
nasaovsd	Std. Dev.	2000 and 5000	0.899463	not significant
nasaovmd	Mode	2000 and 5000	0.825861	not significant

Table E-3 Evaluation of Triangular vs. Gaussian Random Number Generator and Evaluation of 2000 vs. 20,000 iterations

Parameter analyzed	Factor levels analyzed	p-value	at α =.05
Mean at 2000 iterations	Triang. vs. Gaussian RNG	0.000000327	significant *
Mean at 20,000 iterations	Triang. vs. Gaussian RNG	0.00000351	significant **
Mean for Gaussian RNG	2000 vs. 20,000 iterations	0.216087	not significant ***
Mean for Triangular RNG	2000 vs. 20,000 iterations	0.06566	not significant ****

* F statistic = 46.1771 and F-critical = 4.2252 (n=15).
** F statistic = 126.4738 and F-critical = 5.317645 (n=6).
*** F statistic = 1.743753 and F-critical = 4.964591 (n=6).
**** F statistic = 4.270713 and F-critical = 4.964591 (n=6).

E.3 Data Interpretation

Once data was obtained through the expert judgment methodology, there remained another element of subjectivity as to how that data was used. Law and Kelton (1991) discussed several options for selecting a probability distribution in the absence of data (p. 403). They suggested one approach was to "obtain subjective estimates of *a* and *b*," by asking 'experts' for "their most optimistic and pessimistic estimates" (Law and Kelton 1991, p. 403). They also discussed the "triangular approach" which required a "subjective estimate of the most likely" value in question (Law and Kelton 1991, p. 403). An alternative approach was to fit a beta distribution between the subjectively assessed minimum and maximum, *a* and *b*. This allowed the specification of a wide variety of distribution shapes but specifying the parameters for the beta again are subjective. One simplistic approach was to specify the parameters as alpha1 = alpha2 = 1, which converts the beta to a uniform distribution (ie. assumes that X is equally likely to take on any value between *a* and *b*) (Law and Kelton 1991, p. 404). Other shape parameter values were used to specify skewness and Keefer and Bodily (1983) offer alternate methods for "specifying the parameters of a beta distribution" (Law and Kelton 1991, p. 404).

E.3.1 Triangular Distribution

Law and Kelton (1991) demonstrated that choosing the wrong distribution can significantly affect the accuracy of a model's results (p. 326). They also suggested that the triangular distribution was appropriate for situations where a "rough model in the absence of data" (p. 341) was needed. They asserted that a theoretical distribution was preferred over an empirical distribution since extreme values are unlikely to be sampled from an empirical distribution (i.e. only what has occurred historically will be sampled with a high frequency) (Law and Kelton 1991, p.327). Another drawback of an empirical distribution function was that there may be "certain 'irregularities', particularly if only a small number of data values is available" (Law and Kelton 1991, p.327). The triangular distribution was incorporated in the proposed methodology due to the lack of data and because a rough model that was simple to utilize and apply was required. The triangular distribution also afforded the ability to include skewness and avoided a potential central tendency bias that assuming normality might have introduced.

E.3.2 Simulation Inputs

The literature search revealed examples of researchers using different probability distributions based correctly or incorrectly on certain assumptions. Black and Wilder (1980) used the Beta distribution which required the specification of the four moments of the Beta distribution - the mean, the standard deviation, the skewness and the kurtosis. This was particularly important to specify skewness (either left or right) and to specify the degree of "peakedness" (kurtosis). Despite their use of the Beta distribution, Black and Wilder (1980) admitted that similar results were obtained using the Triangular distribution for their data.

Of particular interest, the elicitation procedure was designed to avoid the typical elicitation of probabilities, choice preferences or utility functions. At the recent annual INFORMS conference in Atlanta, November 1996, a presenter (Wolfson 1996) stated that a decision analyst should never attempt to elicit anything more than the first two moments of a probability distribution (i.e. the mean and standard deviation). From the audience, Ward Edwards (see Edwards 1954; Edwards 1961; Edwards 1992) voiced his wholehearted agreement. Note that the moments are not a probability but statistics that estimate population parameters of a probability distribution.

E.3.3 Statistical Distributions Goodness-of-Fit

Since inputs to the Monte Carlo simulation were specified as probability distributions, a statistical goodness-of-fit test was used to verify that proper distributions were utilized. A suitable goodness-of-fit test was the Kolmogorov-Smirnov test.

The Kolmogorov-Smirnov Goodness-of-Fit test evaluates the hypothesis that "sample data was drawn from a specified continuous distribution *F*. The test is nonparametric and exact for all sample sizes"(Fishman 1973) unlike the Chi-square test which is not robust for small sample sizes and assumes normality. The test compares the cumulative frequency distribution (usually the observed CDF but the simulated data CDF in this case) for the sample to that expected for the population specified by the null hypothesis (Lapin 1982). That is, the null hypothesis proposes the CDF that is expected to fit the data. The Kolmogorov-Smirnov test statistic is the maximum deviation between the observed and the expected distributions (Lapin 1982). The results of the Kolmogorov-Smirnov goodness-of-fit tests for input data will indicate that input data fit a particular statistical distribution. Seeking a theoretical probability distribution that best fits the data is recommended for all situations by Hillier and Lieberman (1986) to avoid "reproducing the idiosyncrasies of a certain period in the past" if historical data is used.

Kolmogorov-Smirnov tests were conducted for input data (2000 iterations or 2000 data points) utilizing BestFit® personal computer software. The results of the Goodness-of-Fit tests for input data indicated that all input data fit the triangular distribution better than any of twenty-four other statistical distributions evaluated (Normal and Beta were typically second and third best).

The triangular distribution is specified by either the PDF or CDF. These are expressed as follows. The density or PDF is:

$$\begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & \text{if } a \le x \le b \\[2ex] \dfrac{2(c-x)}{(c-a)(c-b)} & \text{if } b < x \le c \end{cases}$$

where a = minimum, b= most likely, and c=maximum.

Appendix F

Goodness-of-Fit Example Results

Goodness-of-Fit test for Basic structure "cbdy" parameter.

Minimum=	1.8
Maximum=	2.2
Mode=	2.0
Mean=	2.12
Std Deviation=	0.110755
Variance=	0.012267
Skewness=	-0.989258
Kurtosis=	2.954631
Input Settings:	
Type of Fit:	Full Optimization
Tests Run:	Chi-Square	K-S Test

Best Fit Results				
Function	Chi-Square	Rank	K-S Test	Rank
Weibull(26.91,2.16)	0.974884	1.0	0.150305	2.0
Normal(2.12,0.11)	2.211942	2.0	0.208376	3.0
Lognorm(2.12,0.11)	2.798333	3.0	0.211234	4.0
Beta(1.35,0.69) + 1.80	233.366524	4.0	0.784877	5.0
Triang(1.80,2.00,2.20)	1.0e+34	5.0	0.025	1.0

Goodness-of-Fit test for Wing - tail equation coefficient parameter.

Minimum=	4.5	
Maximum=	5.5	
Mode=	5.0	
Mean=	5.316667	
Std Deviation=	0.25766	
Variance=	0.066389	
Skewness=	-0.915328	
Kurtosis=	2.609145	
Input Settings:		
Type of Fit:	Full Optimization	
Tests Run:	Chi-Square	K-S Test

Best Fit Results				
Function	Chi-Square	Rank	K-S Test	Rank
Weibull(28.78,5.40)	0.397563	1.0	0.156864	3.0
Normal(5.32,0.26)	0.722805	2.0	0.209251	4.0
Lognorm(5.32,0.26)	0.918183	3.0	0.211788	5.0
Rayleigh(3.76)	29.785073	4.0	0.510664	7.0
Chisq(6.00)	33.659136	5.0	0.475774	6.0
Beta(1.21,0.76) + 4.50	221.080789	6.0	0.094089	2.0
Triang(4.50,5.00,5.50)	1.0e+34	7.0	0.0125	1.0

Goodness-of-Fit test for Empty Weight Simulation Results.

Minimum=	**1.648926e+5**	
Maximum=	**2.455585e+5**	
Mode=	**2.011923e+5**	
Mean=	**1.999114e+5**	
Std Deviation=	**1.205242e+4**	
Variance=	**1.452607e+8**	
Skewness=	**0.254714**	
Kurtosis=	**2.924987**	
Input Settings:		
Type of Fit:	**Full Optimization**	
Tests Run:	**Chi-Square**	**K-S Test**

Best Fit Results				
Function	Chi-Square	Rank	K-S Test	Rank
PearsonV(2.78e+2,5.53e+7)	17.86695	1.0	0.011031	1.0
PearsonVI(7.70e+4,2.79e+2,7.22e+2)	17.971984	2.0	0.011185	2.0
Lognorm(2.00e+5,1.20e+4)	20.498594	3.0	0.011603	3.0
Normal(2.00e+5,1.21e+4)	40.520243	4.0	0.023952	5.0
Triang(1.65e+5,1.96e+5,2.46e+5)	352.521746	5.0	0.121918	6.0
Beta(4.34,5.66) * 8.07e+4 + 1.65e+5	646.304251	6.0	0.017197	4.0

Appendix G

Follow-up Questionnaire for Multiple Experts

Expert Data Methodology for Risk Analysis of Weight Estimates for a Launch Vehicle

Introduction
Uncertainty is significant at conceptual design. Conceptual design engineering attempts to work from the abstract to the concrete. The amount of uncertain information is significant when attempting to bridge the gap from the abstract concept to a concrete physical design, especially for complex systems.

Purpose of this study
Weight estimating is a major concern in conceptual design. Weight estimating is used to make management decisions in choosing among alternative designs (e.g. lower weight may mean lower life cycle cost). Weight estimates are also important factors used for estimating cost. Typically, weight estimating relationships (WERs) developed and scaled from historical data of aircraft (or previous launch vehicles) are used to estimate weight of the various subsystems of a launch vehicle at the conceptual design phase. Since there is little historical data, the WERs are highly uncertain. Weight uncertainty may lead to increased acquisition cost, schedule overruns, performance deterioration, and increased operating costs. These potential effects make it necessary to address uncertainty and consider the life cycle consequences at conceptual design.

Expert Questionnaire
This study develops a methodology to obtain expert judgment data for quantifying WER parameter ranges including uncertainty. Based on the detailed information required to quantify WER parameter ranges including uncertainty, a questionnaire was developed as a practical and efficient approach for eliciting the expert's opinion. The questionnaire has evolved through several iterations with ample feedback from one NASA expert as to the usefulness of each element included in the questionnaire. The latest iteration of the questionnaire consists of:

i.) Select the Parameters from WERs that will be evaluated for uncertainty.
ii.) Rate the parameter for uncertainty on a five point qualitative scale (Low, 2, Mod., 4, or High).
iii.) Document the reason(s) for the uncertainty for each parameter that is rated.
iv.) The expert is prompted to think of any additional cues that may further document the thinking process that affects the uncertainty rating.
v.) The expert is asked to anchor the three major points along the five point scale quantitatively. This documents the meaning of Low, Moderate and High uncertainty from the expert's perspective. These quantitative assessments are ultimately used as an estimate of the standard deviation for the statistical distribution.
vi.) Provide parameter values at three levels - Minimum, Most Likely and Maximum (the uncertainty rating and the quantitative anchor of uncertainty are used to aid this process).
vii.) Describe any scenario that would change a subsystem/parameter rating and also provide the changes that would result if that scenario occurred.

The following pages contain Weight Estimating Relationships (WERs) for various subsystems for a launch vehicle. The vehicle in question is one of the proposed designs for replacing the current space shuttle [specifically - single-stage vehicle, rd-701, horz. 30 ft. p/l bay, 25 klb p/l - 51.6 inc.]. This vehicle is being evaluated at the conceptual design phase.

<u>Instructions for Group Questionnaire:</u>
Based on your membership in SAWE, you have been selected as a knowledgable individual in this subject matter. You are asked to assume the role of conceptual design engineer. Your task is to evaluate the uncertainty of the design parameters that are used in the Weight Estimating Relationships (WERs). Your participation will serve to verify and validate the first two steps (or more) in the expert elicitation procedure.

The questionnaire consists of 2 phases.

Phase I
You will be provided with the parameters, the uncertainty ratings and the reasons for those ratings as identified by the NASA LaRC conceptual design engineer. You are given this information to familiarize yourself with the methodology and the types of ratings and reasons that identify the level of uncertainty.

Phase II
You will be asked to provide uncertainty ratings. More instructions will be given at that point in the questionnaire.

The primary purpose of this questionnaire is to provide a validation technique to satisfy research requirements associated with the completion of my doctoral dissertation. Your participation will be greatly appreciated.

I ask your permission to include some information about you and your qualifications in the appendix or body of my dissertation. Your name, your employer and all other information will be protected and will not be published in any other journal or conference paper without your permission.

Please call me or send an E-mail if you have questions about the questionnaire at any time. Thank you for your participation.

Richard Monroe
Home phone: ***(757) 622-6240*** ***Work:*** ***(757) 683-4161***
e-mail: **rwm400z@mail.odu.edu**

Nomenclature

'a'	geometry parameter
'c'	constant parameter
'l'	geometry parameter
's'	calculated parameter
const	constant
eq.coef.	equation coefficient
ssv	single stage vehicle
reduction factor	weight reduction % from reference data point (shuttle, etc.)
(shuttle)	current shuttle subsystem is used as reference data point
(marshall study)	source of data or reason for data is listed in parentheses, e.g. marshall, shuttle, composites, etc.
Gr/Ep	Graphite/Epoxy
Al-Li	Aluminum lithium
*	multiply variables (as listed in WER statements)
**	the term following is an exponent (in WER statements)

Follow other mathematical operations as normally executed.

PHASE I
You are provided with the WER parameters, the uncertainty ratings and the reasons that were assessed by the NASA LaRC engineer. The following 3 pages are examples.

READ THE FOLLOWING EXAMPLE WER ASSESSMENTS TO GAIN AN UNDERSTANDING OF HOW THE METHODOLOGY IS USED and to see TYPICAL EXAMPLE RESULTS.

EXAMPLE

WING subsystem
Select the WER parameters from the following list that you want to evaluate for uncertainty.

(expwing) parameters

'c'	'1'	1.0	constant
'c'	'c1'	.82954	equation coefficient
'c'	'c2'	.001	divide load by 1000
'c'	'usf'	1.75	ultimate safety factor
'c'	'nf'	2.0	load factor
's'	'wland'		landed wt
'a'	'exp'	3360.	exposed wing area
'c'	'ar'	1.48	aspect ratio based on exposed area
'c'	'tr'	.34	taper ratio ct/cr
'c'	'toc'	.10	thickness to chord ratio
'c'	'e1'	.48	exponent
'c'	'e2'	.67	exponent
'c'	'e3'	.64	exponent
'c'	'e4'	.40	exponent
'c'	'rew'	.40	reduction factor (lo2-lh2 ssv, ezedesit, Gr/Ep)

Given WER:

c1*(c2*usf*nf*wland)**e1*(exp)**e2*ar**e3*((1+tr)/toc)**e4*(1-rew)

1. Choose to select parameters from the above WER and rate them for uncertainty.

From the WING (expwing) WER parameters you have selected:

c	eq. coef.	0.82954

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 High

The NASA LaRC engineer provided the following reason for the uncertainty rating.

"For conceptual design, WERs for wings are typically more accurate than for other components."

The following cues were also listed by the NASA LaRC engineer:

"1. WER is based on a regression of historical data points.
2. Fit to data is good.
3. Data points are applicable to vehicle type."

"Size of applicable data set. Basis of weight (actual, calculated, estimated)."

INSTRUCTIONS FOR QUESTIONNAIRE [Phase II]

1. Rate WER parameter uncertainty QUALITATIVELY from Low, Moderate to High uncertainty. Focus only on those WER parameters that you feel should be evaluated in this manner.

2. Simultaneous with your UNCERTAINTY rating provide a REASON for your rating.

3. Anchor your QUALITATIVE description of uncertainty to a QUANTITATIVE measure on the 5-point scale provided.

Steps 4 and 5 are not required in this evaluation.

4. Provide 3 point estimates [Low, Mode or Most Likely, and High] for each of the MOST UNCERTAIN WER parameters identified in the preceding steps. ***{not shown here}***

5. Describe any scenarios that may change WER PARAMETER values. Provide the alternative WER PARAMETER values that in your judgment would be appropriate for the scenario. ***{not shown here}***

Uncertainty **here has been defined (or interpreted) as the total amount of variance for a design parameter from an initial design point estimate. In other words, given the nature of the WER parameters and what they represent, what is the potential range of a specific parameter value (assuming the variable is continuous). Specify the range in terms of a total percentage (i.e. total variation or total uncertainty). For example, the quantity of 20% would represent a total variation of -10% to +10% around the point estimate.**

Keep this definition in mind as you attempt to rate each of the WER parameters. Ultimately, your rating would be used to calculate an upper bound and a lower bound around the point estimate or most likely value.

The rating choices are LOW, 2, MODERATE, 4, HIGH and None.
Choose Low, Moderate or High based on the level of Uncertainty that you feel applies to that particular subsystem WER.

Choose 2 if Uncertainty is more than Low but less than Moderate.

Choose 4 if Uncertainty is more than Moderate but less than High.

Choose NONE if the WER is constant or 100% certain.

One of three possible actions are requested of you for each WER for the listed subsystems:

1. **Select appropriate parameters to rate for uncertainty and perform the rating.**
2. **Reply that you do not feel comfortable making an uncertainty rating because you do not have sufficient information to make a judgment.**
3. **Develop your own WER model using parameters that you think are appropriate for a given subsystem then select parameters and rate their uncertainty.**

On the following pages parameters are provided that were selected by the NASA LaRC engineer. Selection does not automatically assume HIGH UNCERTAINTY. Any uncertainty rating can be applied to the selected parameters for a given WER.

Propulsion, main
Press and feed
Select the WER parameter from the following list that you want to evaluate for uncertainty.

Given WER:

cpf*tox*(gorss+adpay)/(ispsl*pwr)/dbulk*(1-reng)

1. Choose parameters and rate for uncertainty.

'press' parameters

'c'	'1'	1.0	const
'c'	'cpf'	44.4	pres. and feed const, based on vol. flow rate (Marshall study)
'c'	'dbulk'	62.6	propellant bulk density, o/f=
'c'	'ispsl'	452.2	sea level isp (sec)
'c'	'pwr'	1.0	power level
'c'	'tvac'	2054000.	vacuum thrust (lb)
'c'	'tow'	1.3	lift-off t/w
'c'	'adpay'	0.	additional down p/d capability
's'	'gross'		gross wgt
'c'	'reng'	0.0	reduction factor

From the press WER parameters you have selected:

c	cpf	44.4	pres. and feed const, based on vol. flow rate

Rate the degree of uncertainty that you associate with this parameter:

Low	2	Moderate	4	High

Optional: provide a reason for your uncertainty rating.

2. I choose not to rate this WER due to lack of information.

3. I choose to develop my own WER and select and rate parameters from that new WER.

Propellant tanks

Given WER:

ctnk*((1 - e**(c*delv/isp/ge)) *inserts + omsres) * (1 - romstnk)

1. Choose parameters and rate for uncertainty.

'omstnks' parameters

'c'	'1'	1.0	const
'c'	'c'	-1.0	
'c'	'ctnk'	.037	low pressure tank const (lb/lb), o/f=6
'c'	'e'	2.71828	value of e
'c'	'isp'	462.2	vac. specific impulse (sec)
'c'	'delv'	1350.	delta v req. 1350 ft/sec, due east req.
'c'	'ge'	32.174	gravity const (ft/sec^2)
's'	'inserts'		insertion wt
's'	'omsres'		oms reserve propellant
'c'	'romstnk'	0.	reduction factor

From the omstnks WER parameters you have selected:

c	ctnk	.037	low pressure tank const (lb/lb), o/f=6

Rate the degree of uncertainty that you associate with this parameter:

Low	2	Moderate	4	High

Optional: provide a reason for your uncertainty rating.

2. I choose not to rate this WER due to lack of information.

3. I choose to develop my own WER and select and rate parameters from that new WER.

Electric conversion and distr.
Avionic cabling

Given WER:

wac*(1-rcab) + winst*rinst

1. Choose parameters and rate for uncertainty.

'avcable' parameters

'c'	'1'	1.0	const
'c'	'wac'	2565.	avionic cable wt. (shuttle)
'c'	'winst'	564.	supports and installition wt. (shuttle)
'c'	'rcab'	.30	reduction factor (fiber optics)
'c'	'rinst'	.20	reduction factor

From the avcable WER parameters you have selected:

c	wac	2565.	avionic cable wt. (shuttle)

Rate the degree of uncertainty that you associate with this parameter:

Low	2	Moderate	4	High

Optional: provide a reason for your uncertainty rating.

2. I choose not to rate this WER due to lack of information.

3. I choose to develop my own WER and select and rate parameters from that new WER.

Electric conversion and distribution
Wire trays

Given WER:

wtrays*(1-rtray)

1. Choose parameters and rate for uncertainty.

'tray' parameters

'c'	**'1'**	**1.0**	**const**
'c'	**'wtrays'**	**592.**	**wire trays wt. (shuttle)**
'c'	**'rtray'**	**.20**	**reduction factor (composites)**

From the tray WER parameters you have selected:

c	**wtrays**	**592.**	**wire trays wt. (shuttle)**

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 High

***Optional:* provide a reason for your uncertainty rating.**

From the tray WER parameters you have selected:

c	**rtray**	**.20**	**reduction factor (composites)**

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 High

***Optional:* provide a reason for your uncertainty rating.**

2. I choose not to rate this WER due to lack of information.

3. I choose to develop my own WER and select and rate parameters from that new WER.

Main gear

Running gear

Select the WER parameter from the following list that you want to evaluate for uncertainty.

'magrg' parameters

'c'	'1'	1.0	const
'c'	'c1'	.001	const
'c'	'cmrg'	173.	running gear const. (main)
'c'	'.14'	.14	exponent
'c'	'nw'	8.	number of main wheels, total
'c'	'.75'	.75	exponent
's'	'wland'		landed wt
'c'	'rig'	0.	reduction factor

Given WER:

c1*cmrg*wland**.75*nw**.14*(1-rig)

1. Choose parameters and rate for uncertainty.

From the magrg WER parameters you have selected:

c	cmrg	173.	running gr const. (main)

Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate 4 High

Optional: provide a reason for your uncertainty rating.

2. I choose not to rate this WER due to lack of information.

3. I choose to develop my own WER and select and rate parameters from that new WER.

Follow-up Questions
I. Ease of use and/or usefulness of methodology and questionnaire.

1. Comment on the ease of use of the methodology.

2. Do you find the methodology to be useful for a weight estimation analysis?

3. Would you prefer to use your own models (WERs or MERs)?

4. Would you find the methodology useful if adapted to your own analysis problem with your own models?

II. Uncertainty
1. Did you find the original expert's example judgments to be reasonable and understandable?

2. Does this interpretation of uncertainty (as total variation) seem logical to you?

3. Do you have any other suggestion of how to interpret uncertainty?

4. Do you have any other method or any suggestion of how to judge uncertainty?

Benchmark Questions

1. Given that a WER parameter value is based on a regression of historical data and the regression line has a good fit to the data:
What is your uncertainty rating for such a parameter?
Rate the degree of uncertainty that you associate with this parameter:

Low 2 Moderate4 High

2. Given that a WER parameter value is based on someone else's analysis or experiment (for example a study at Marshall Space Flight Center or at Johnson Space Center, etc.):
What is your uncertainty rating for such a parameter? Explain your assumptions about the data source if that is an important consideration to you.

Low 2 Moderate4 High
Explanation (if required):

3. Given that a WER parameter is a reduction factor that has been validated using actual structures or by some other analytical techniques:
What is your uncertainty rating for such a parameter?

Low 2 Moderate4 High

4. Given that a WER parameter is based on a known design (such as the current space shuttle) and the new structure is assumed to be similar:
What is your uncertainty rating for such a parameter?

Low 2 Moderate4 High

5. Given that the subsystem structure being analyzed is not well-defined (i.e. very early in the conceptual design phase) and the WER parameter is estimated:
What is your uncertainty rating for such a parameter?
Low 2 Moderate4 High

{for STEP 3 above}
***Uncertainty* here has been defined (or interpreted) as the total amount of variance for a design parameter from an initial design point estimate. In other words, given the nature of the WER parameters and what they represent, what is the potential <u>range</u> of a specific parameter value (assuming the variable is continuous). Specify the <u>range</u> in terms of a total percentage (i.e. total variation or total uncertainty). For example, the quantity of 20% would represent a total variation of -10% to +10% around the point estimate.**

Keep this definition in mind as you attempt to rate each of the WER parameters. Ultimately, your rating would be used to calculate an upper bound and a lower bound around the point estimate or most likely value.

Provide a Quantitative explanation of your understanding of Low, Moderate and High uncertainty. *CIRCLE ONE NUMERICAL CHOICE FOR EACH.*

The amount of uncertainty or variation that I associate with <u>Low Uncertainty</u> is:

Less	**5%**	**7.5%**	**10%**	**15%**	**20%**	**More**

The amount of uncertainty or variation that I associate with <u>High Uncertainty</u> is:

Less	**15%**	**20%**	**30%**	**40%**	**50%**	**More**

The amount of uncertainty or variation that I associate with <u>Moderate Uncertainty</u> is:

Less	**10%**	**15%**	**20%**	**25%**	**30%**	**More**

For ratings of 2 or 4 on the Qualitative rating sheet:

the midpoint between Low and Moderate will be used for a 2 rating

the midpoint between Moderate and High will be used for a 4 rating

VITA

Richard W. Monroe
Department of Engineering Management
Old Dominion University
Norfolk, VA 23529

Richard W. Monroe earned his bachelor's in industrial engineering technology with honor from Southern Technical Institute in Marietta, Georgia in 1975. He completed his M.S.in Engineering Management in 1990 at Western New England College, Springfield, Massachusetts. He has seventeen years of experience in industry including three years with Southwire Company in Carrollton, Georgia and Jewett City, Connecticut, five years with Spalding Sports Worldwide in Chicopee, Massachusetts and over five years with Stanadyne Automotive Corp. in Windsor, Connecticut. While pursuing his doctorate, he has taught at Old Dominion University in the College of Engineering and Technology for the past three years. He has also taught as an adjunct instructor in the College of Business and Public Administration for the past year.