

1-1-2010

# An analysis of reasonableness models for research assessments

William D. Kight  
*Walden University*

Follow this and additional works at: <https://scholarworks.waldenu.edu/dissertations>

 Part of the [Business Administration, Management, and Operations Commons](#), and the [Management Sciences and Quantitative Methods Commons](#)

---

This Dissertation is brought to you for free and open access by the Walden Dissertations and Doctoral Studies Collection at ScholarWorks. It has been accepted for inclusion in Walden Dissertations and Doctoral Studies by an authorized administrator of ScholarWorks. For more information, please contact [ScholarWorks@waldenu.edu](mailto:ScholarWorks@waldenu.edu).

# Walden University

COLLEGE OF MANAGEMENT AND TECHNOLOGY

This is to certify that the doctoral dissertation by

William Kight

has been found to be complete and satisfactory in all respects,  
and that any and all revisions required by  
the review committee have been made.

Review Committee

Dr. Aridaman Jain, Committee Chairperson,  
Applied Management and Decision Sciences Faculty

Dr. Robert Aubey, Committee Member,  
Applied Management and Decision Sciences Faculty

Dr. Lilburn Hoehn, Committee Member,  
Applied Management and Decision Sciences Faculty

Chief Academic Officer

Denise DeZolt, Ph.D.

Walden University  
2010

ABSTRACT

An Analysis of Reasonableness Models for Research Assessments

by

William D. Kight

MSEE, University of Maryland, College Park, 1984

BS, Rensselaer Polytechnic Institute, 1980

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Sciences

Walden University

February 2010

## ABSTRACT

Individuals who screen research grant applications often select candidates on the basis of a few key parameters; success or failure can be reduced to a series of peer-reviewed Likert scores on as little as four criteria: *risk*, *relevance*, *return*, and *reasonableness*. Despite the vital impact these assessments have upon the sponsors, researchers, and society in general as a benefactor of the research, there is little empirical research into the peer-review process. The purpose of this study was to investigate how reviewers evaluate reasonableness and how the process can be modeled in a decision support system. The research questions both address the relationship between an individual's estimates of reasonableness and the indicators of scope, resources, cost, and schedule as well as evaluate the performance of several cognitive models as predictors of reasonableness. Building upon Brunswik's theory of probabilistic functionalism, a survey methodology was used to implement a policy-capturing exercise that yielded a quantitative baseline of reasonableness estimates. The subsequent data analysis addressed the predictive performance of six cognitive models as measured by the mean-square-deviation between the models and the data. A novel mapping approach developed by von Helversen and Rieskamp, a fuzzy logic model, and an exemplar model were found to outperform classic linear regression. A neural network model and the QuickEst heuristic model did not perform as well as linear regression. This information can be used in a decision support system to improve the reliability and validity of future research assessments. The positive social impact of this work would be more efficient allocation and prioritization of increasingly scarce research funds in areas of science such as social, psychological, medical, pharmaceutical, and engineering.



An Analysis of Reasonableness Models for Research Assessments

by

William D. Kight

MSEE, University of Maryland, College Park, 1984

BS, Rensselaer Polytechnic Institute, 1980

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Sciences

Walden University

February 2010

UMI Number: 3396317

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3396317

Copyright 2010 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## DEDICATION

This dissertation is dedicated to the late Dr. James N. Kile, who for many years unwittingly inspired me to pursue this program. I only wish that I had started sooner so that he could have shared in my success and known my gratitude.



## ACKNOWLEDGMENTS

I would like to thank my committee members, Dr. Aridaman Jain, Dr. Lilburn Hoehn, and Dr. Robert Aubey, for their guidance and encouragement. I would also like to thank Dr. Raghu Korrapati for his suggestions and support in navigating this dissertation through the university review process. I must also thank my Walden University classmates who made this journey more than just hard work, but created a sense of community and camaraderie that made the journey fun.

## TABLE OF CONTENTS

LIST OF TABLES .....	iv
LIST OF FIGURES .....	vi
CHAPTER 1: INTRODUCTION TO THE STUDY.....	1
Background of the Study .....	2
Research Assessment .....	2
Problem Statement .....	4
Purpose of the Study .....	5
Nature of the Study .....	6
Research Questions .....	7
Theoretical Basis .....	8
Definition of Terms .....	11
Assumptions.....	11
Limitations, Delimitations, and Scope.....	12
Limitations .....	12
Delimitations .....	12
Scope .....	12
Significance of the Study .....	13
Summary .....	14
CHAPTER 2: LITERATURE REVIEW .....	16
Introduction.....	16
The State of Research Assessment .....	17
Peer Review .....	17
Portfolio Management .....	20
Project Selection .....	24
Decision Theory.....	26
Cognitive Reasoning and Judgment.....	31
Artificial Intelligence .....	38
Methodology .....	41
Policy-Capturing .....	41
Alternative Approaches .....	44
Literature Synthesis .....	46
CHAPTER 3: RESEARCH METHOD .....	48
Introduction.....	48
Research Design.....	48
Methodology .....	49
Approach.....	51
Setting and Sample .....	52
Target Population.....	52

Sampling frame .....	52
Sampling Method .....	53
Demographics and Eligibility .....	54
Sample Size .....	54
Instrumentation .....	57
Policy-Capturing Instrumentation .....	57
Validity and Reliability .....	59
Data Collection .....	61
Data Collected .....	61
Data Collection Procedures .....	61
Informed Consent .....	63
Protection of Participants .....	63
Data Analysis .....	63
Descriptive Analysis and Characterization .....	63
Policy-Capturing .....	64
Nonlinear Models .....	64
Parametric Models .....	65
Fuzzy Model .....	66
Metrics and Mapping Model .....	68
QuickEst Heuristic Model .....	69
Linear Regression Model .....	70
Exemplar Model .....	70
Neural Network Model .....	72
Summary .....	75
CHAPTER 4: RESULTS .....	76
Introduction .....	76
Analysis of the Policy-capturing Exercise .....	76
Research Question 1 .....	76
Outlier Case .....	77
Collinearity and Correlation Analysis .....	78
Scenario Analysis .....	79
Response Characterization .....	82
Test Retest Analysis .....	87
Description of the Sample .....	88
Analysis of Subsamples .....	89
Analysis of Model Performance .....	89
Research Questions 2 and 3 .....	89
Training Parameters .....	90
Simulation Parameters .....	91
Modeling Results .....	92
Model Programming Validation .....	100
User Comments on the Exercise .....	100
Summary .....	102

CHAPTER 5: SUMMARY, CONCLUSION, AND RECOMMENDATIONS .....	104
Summary .....	104
Interpretation of Findings .....	105
Policy-Capturing Exercise and Research Question 1 .....	105
Conclusions with Respect to Research Question 1 .....	106
Modeling Performance and Research Questions 2 and 3 .....	107
Conclusions with Respect to Research Questions 2 and 3 .....	110
Implications for Social Change.....	110
Recommendations for Action .....	111
Recommendations for Further Study .....	112
Concluding Remarks.....	113
REFERENCES .....	115
APPENDIX A: POLICY-CAPTURING DISPLAYS .....	123
APPENDIX B: WEB SITE PYTHON CODE.....	133
Openconsent.py.....	133
Instructions.py.....	134
Scenarios.py .....	135
Demographics.py .....	141
Closing.py .....	143
APPENDIX C: R PROGRAMMING CODE .....	145
Main Program .....	145
Exemplar Model.....	152
Fuzzy Model .....	153
Mapping Model.....	154
Neural Model .....	156
QuickEst Model .....	157
Regression Model .....	159
APPENDIX D: COPYRIGHT PERMISSIONS .....	161
APA Publications.....	161
Wiley-Blackwell Publications .....	162
CURRICULUM VITAE.....	163

## LIST OF TABLES

Table 1. Cellular Phone Value Estimations using Metrics and Mapping, Linear Regression, Exemplars, and Heuristics.....	35
Table 2. Eligibility Questions and Minimum Criteria for Participating in the Study. ....	54
Table 3. Fuzzy Indicators used to Represent Cues on a 5-Point Likert Scale .....	58
Table 4. Recommendations for Improving Validity of Policy-capturing Exercises.....	60
Table 5. Polychotomous to Dichotomous Cue Mapping. ....	66
Table 6. Attention Weighting Values versus Degree of Separation. ....	71
Table 7. Variance Inflation Factors for the Scenario Cues.....	79
Table 8. Correlation between Cues Using the Spearman Rho Statistic.....	79
Table 9. Parallel Graph Vertical Scale Values.....	82
Table 10. Test Retest Correlation across the Entire Sample Set and for Each Interval Tracked in the Exercise. ....	87
Table 11. Demographic Summary of the Sample.....	88
Table 12. Descriptive Statistics for the Session Mean Squared Error of the Primary Subpopulations.....	89
Table 13. Calculation of the Probability of Representing all Five Cue-Total Values in the Sample. ....	91
Table 14. Mean Squared Error Performance of the Models over 1000 Trials.....	93
Table 15. Tabular Results of the Pairwise Tukey Analysis of the Difference Between Means, Arranged in Order of Increasing Difference. ....	97
Table 16. Chi-Square Test of the Model Residuals Against a Normal Distribution. ....	100
Table 17. Coding Counts for Survey Open Ended Questions. ....	101
Table 18. J-Shaped (Nonlinear) Combined Model Performance Results from von Helversen and Rieskamp Study 2. ....	109

Table 19. Normalized Comparison of von Helversen and Rieskamp Combined  
Root Mean Squared Deviation Versus the Root Mean Square Error Results of this  
Study. ....109

## LIST OF FIGURES

Figure 1. The probabilistic lens model depicts the process by which exhibited behaviors are perceived by others.....	9
Figure 2. A simple illustration of a bubble chart projects ranked by risk, return, and resources.....	23
Figure 3. Conceptual illustration of Brown and Siegler’s Metrics and Mapping framework.....	33
Figure 4. Mind map illustrating the relationship of various topics presented in the literature review .....	47
Figure 5. A priori power analysis of the measured data regression. From “G*Power version 3.1.0” by Faul et al., 2007 .....	56
Figure 6. Fuzzy set membership functions describing the relationship between cues and reasonableness.....	68
Figure 7. The logistic curve activation function used in neural models .....	73
Figure 8. Prototypical neural network model for estimating reasonableness .....	74
Figure 9. The distribution of the individual session Mean Standard Error values was used as an indicator of the quality of the session responses. ....	78
Figure 10. Frequency Distribution of Scenario IDs included the data sample. ....	80
Figure 11. Normalized cumulative distribution of scenario IDs in the sample .....	81
Figure 12. Distribution of individual cue values within the sample.....	81
Figure 13. A parallel line plot for a reasonableness score of 7.....	83
Figure 14. A parallel line plot for a reasonableness score of 6.....	83
Figure 15. A parallel line plot for a reasonableness score of 5.....	84
Figure 16. A parallel line plot for a reasonableness score of 4.....	84
Figure 17. A parallel line plot for a reasonableness score of 3.....	84
Figure 18. A parallel line plot for a reasonableness score of 2.....	85
Figure 19. A parallel line plot for a reasonableness score of 1.....	85

Figure 20. Distributions of cue values corresponding to specific ratings of Reasonableness (1 = Very Unreasonable, 7 = Very Reasonable).....	86
Figure 21. Power analysis to determine the number of simulation trials required to resolve a difference between means .....	92
Figure 22. Box plot illustrating the mean squared error for each model across 1000 trials.....	94
Figure 23. Neural Model error convergence plot of the first 25 trials.....	95
Figure 24. Depiction of the pairwise difference between models mean squared error.....	96
Figure 25. Histogram plot of the squared error residuals for each model across 1000 trials.....	98
Figure 26. Normal probability plots of the squared error residuals for each model across 1000 trials. ....	99
Figure A1. Eligible participants will be given a username and password to access the exercise.....	123
Figure A2. A standard consent form with an explanation of the exercise is presented to the participant. ....	124
Figure A3. The user must agree with the consent form in order to proceed with the exercise. ....	125
Figure A4. The instructions page begins with background information for the exercise .....	126
Figure A5. The instructions then explain the participant’s role and how to answer the questions.....	127
Figure A6. The participant is presented with a series of 50 randomly generated scenarios to evaluate .....	128
Figure A7. Basic demographics information relevant to the study is collected .....	129
Figure A8. If a participant attempts to back up or enter the exercise out of sequence, the program tactfully reminds the participant not to back up .....	130
Figure A9. The participant is asked to comment on the exercise before leaving .....	131



Figure A10. The participant is provided with a point of contact for more information at the conclusion of the exercise .....132

Figure D1. APA copyright and permission guidelines .....161

Figure D2. Permission obtained from Wiley-Blackwell.....162

## CHAPTER 1: INTRODUCTION TO THE STUDY

For all of the hours and effort that go into the preparation of a research prospectus, project plan, grant application, or capital funding request, the critical first assessment is often reduced to a series of Likert scores on as little as four criteria: *risk*, *relevance*, *return*, and *reasonableness* (De Piante Henriksen & Palocsay, 2008). Furthermore, these scores are typically generated by using a peer-review model, the ubiquitous gold-standard of scientific research evaluation. Peer-review, however, is notoriously unreliable and subject to errors, bias, and fraud (Banghart, 2006; Marsh, Jayasinghe, & Bond, 2008).

Despite the prevalence of the peer-review method, and the vital impact these assessments have upon the research sponsors, researchers, and society in general as benefactors of the research, there is very little rigorous empirical research into the peer-review process (Marsh et al., 2008, pp. 160-161). This writer sought to develop a better understanding of the research assessment process, and in particular, the estimation practices that peer reviewers employ when they judge the *reasonableness* of a research plan. Reasonableness was selected as the focus this effort because it incorporates the core elements of project planning comprising cost, schedule, scope, and resource assessment.

With this knowledge, several established cognitive estimation models were evaluated for possible use within a parametric decision support system. Parametric decision support systems use mathematical and statistical techniques to evaluate the effort and resources required to perform a task where the parameters are typically indirect

indicators of the work to be done. Decision support systems can bring transparency, consistency, and equitable treatment to the assessment process that will serve to improve the reliability and validity of future research assessments.

### Background of the Study

The genesis of this study can be found in a recurring debate within this writer's organization about the effectiveness of cost and schedule planning exercises for various U.S. government-sponsored Research and Development (R&D) programs. The research staff often complains that such planning is meaningless given the uncertainty of the research outcome; some project estimates have been recognized as totally arbitrary conjectures. As one would expect, management's confidence in the research department's estimates of project cost and schedule is very low.

Without a doubt, estimating the magnitude and scope of a research task produces an uncertain result based upon information that is limited or has missing data. Nonetheless, the reality remains that these estimates are used by management to prioritize and allocate increasingly scarce research and development funds. To maximize long term value creation for the funding sponsor, managers must have a solid understanding of both the potential future value of a research program and the effort and resources required to produce a favorable outcome. Good planning can improve the chance of selecting the best programs while poor planning can lead to wasted time, money, and effort.

### Research Assessment

Planning a budget as a researcher or evaluating a budget as a manager requires estimates of both the scope of the project and the resources required. Scope describes the

extent of the work to be performed and the expected outcomes; resources include time, labor, facilities, materials, tools, and other direct costs. When the resources are properly aligned with the scope, the project is said to have a *reasonable* project plan.

Project managers, grant administrators, proposal reviewers, research review boards, and faculty members are examples of people who are frequently required to estimate the *reasonableness* of a research proposal as part of a larger research assessment. Research assessment, as its name implies, is an evaluation of one or more research projects and is a component of project management activities that include:

1. Portfolio analysis: A study of the range of research projects sponsored by an organization to ensure a diverse blend of low-risk-low-return and high-risk-high-return projects.
2. Project selection: The ranking and evaluation process that identifies projects for investment based upon metrics that may include but are not limited to features such as risk, resource availability, reasonableness, future value, market need, enabling technologies, and so on.
3. Planning and control: Periodic assessment of project performance against stated goals and adjustment of plans, resources, and milestones as may be necessary. Decision gates are often used to determine if a project should continue, be terminated, or be reactivated.

4. Project evaluation: Typically an ex post facto study of the productivity or efficiency of a research organization based upon the quality and/or quantity of research produced. Research evaluation is the same as an evaluation study that is performed to test the effectiveness of an experimental intervention or treatment.

5. Trend analysis: An extrapolation of research progress that predicts the direction and pace of scientific or technological development to forecast markets or to gauge the availability of enabling technologies.

(Kight, 2009, pp. 2-3)

This study focused primarily on the estimation of reasonableness as it applies to portfolio analysis and project selection in which the principal evaluation method used is the peer review process (Cookson & Jack, 2008); hence, peer review was the starting point of this investigation. The study further considered decision theory, cognitive modeling, and artificial intelligence as disciplines which address the underlying cognitive processes by which individuals comprehend information, reason, and ultimately make judgments. Building upon this foundation, the goal of this study was to synthesize an approach for modeling human estimates of reasonableness as a means of improving the peer review process.

### Problem Statement

This writer's interest was in the underlying estimation processes that researchers use to create proposals and plans and that peer reviewers use to arrive at their assessment of those proposals and plans. The problem this study will address is the poor inter-rater and single-rater reliability that exists in the assessment process. Marsh, Jayasighe, and

Bond (2008) asserted in their study of over 2300 grants awarded by the Australian Research Council that “the decision of whether or not to fund was based substantially upon chance, whether the random error happened to be positive or negative” (p. 162).

Considering the reliance of different stakeholders upon the peer review method and the controversy surrounding it, one might expect the subject to have been thoroughly explored. As previously noted, however, there is very little rigorous empirical research into the operationalization and effectiveness of the peer review assessment process (Marsh et al., 2008, p. 160). Citing Jefferson, Rudin, Brodney, and Davidoff, Marsh et al. (2008) contend that “good research on the peer review process was so rare that almost no conclusions were warranted, particularly about constructive alternatives and interventions designed to improve peer reviews” (p. 161).

#### Purpose of the Study

The purpose of this study was to gain a better understanding of how people evaluate reasonableness and how the process can best be modeled in a decision support system. The effort used a survey methodology to implement a policy-capturing exercise; the exercise was designed to empirically determine the relationship between the factors comprising cost, schedule, resources, and scope and the reasonableness judgments of a simulated peer reviewer community. Six cognitive models were tested against the data to determine if a viable parametric model of reasonableness could be derived. This work drew upon concepts from econometrics, decision theory, artificial intelligence, and cognitive reasoning.

### Nature of the Study

There are three principal paradigms from which a research approach may be developed: these are “(a) positivist/empiricist using quantitative methods, (b) constructivist/naturalist using qualitative methods, and (c) pragmatic using a combination of quantitative and qualitative methods” (Kight, 2009, pp. 8-9). This writer holds a pragmatic view of research which allows for the best method or methods to be used to address the specific questions being asked.

In this study, a parametric analysis of the estimation process required quantitative methods; however, research has few well defined attributes and considerably greater uncertainty. To fully investigate the problem would also require delving into the thought processes of those who estimate research well, to uncover precisely what they are looking for and the essence of how they judge a task. This is the realm of qualitative analysis hence a mixed-method approach was considered. Pragmatic considerations prevailed, however, based upon limits imposed by available time and resources; the qualitative components of the study were deferred for future work.

Within the quantitative paradigm Singleton and Straits define four principal research methods that include: (a) experiments, (b) surveys, (c) field research, and (d) available data (2005, p. 5). Field research was eliminated as a methodology because the characteristics of interest are not readily observable. Available data was seriously considered but it was found that records of proposal assessments are closely held and not generally available for review, particularly with respect to unsuccessful proposals. Experimental and survey methodologies both represented viable approaches; the survey

methodology was ultimately selected because of its suitability for collecting and analyzing data about a population (Kight, 2009).

The nature of this study is thus quantitative, using a survey methodology conducted in the form of a policy-capturing exercise (Aiman-Smith, Scullen, & Barr, 2002). The research approach was based upon the simulation and policy-capturing exercise used by von Helversen and Rieskamp (2008). A total of 40 professionals in business and academia rated the reasonableness of 50 randomly selected scenarios using a 7-point Likert scale. Scenarios were drawn from a fully orthogonal set of contrived proposals comprising 625 possible combinations of scope, resources, schedule, and cost.

The subsequent data analysis employed multiple regression techniques to assess the predictive performance of six parametric models using fuzzy set theory, neural networks, cognitive mapping, linear regression, exemplar, and QuickEst models respectively. The predictive performance of the models was quantitatively compared using the mean-square-deviation from the measured estimates of the participants.

### Research Questions

To address the problems identified and achieve the desired goals for this study, the research questions that were asked included:

1. What is the observed relationship between an individual's estimation of reasonableness and the independent predictors comprising scope, resources, cost, and schedule?



2. To what degree will various cognitive models of decision making, reasoning, and judgment predict the estimates of reasonableness by a peer review community?

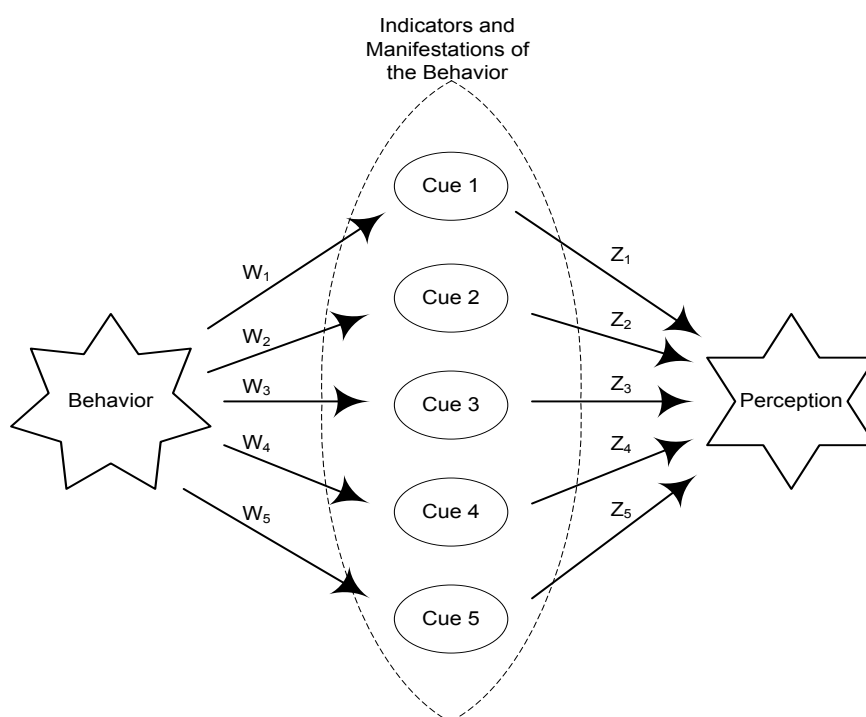
3. Which cognitive model most accurately reflects a peer reviewer's estimation of reasonableness?

These questions are descriptive in nature, hence no specific hypothesis was proposed. Descriptive studies are designed to characterize behaviors, attitudes, or experiences within a population rather than establish a relationship between variables as would be required to test a hypothesis (Singleton & Straits, 2005, pp. 65, 223). Statistical significance testing, however, was used to assess the degree to which the cognitive models predicted human behavior that was not attributable to random assignment or chance. Significance testing is considered controversial in nonexperimental settings but is recommended by Singleton and Straits as "an effective means of screening out trivialities and chance mishaps" (p. 457).

### Theoretical Basis

The policy-capturing methodology derives from the work of Egon Brunswik who pioneered probabilistic functionalism, which later became known as the lens model of perception (Aiman-Smith et al., 2002; Doherty, 2006; Dorsey & Coovert, 2003; Hammond, 1996). The lens model describes the judgment of behaviors as a two-step process: In the first step, a behavior manifests itself to varying degrees in the form of cues or indicators; in the second step, the cues or indicators are perceived by the individual judging the behaviors (Doherty, 2006). The lens analogy stems from the

representation of the process as shown in Figure 1, which resembles the rays of light emanating from a source, passing through a lens, and perceived by the receiver; as with any lens, there is some distortion. The perception of a behavior is thus a function of the original behavior, the observable manifestations of those behaviors, and the interpretive transformations that occur between source and destination, all within the environmental context at the moment of observation.



*Figure 1.* The probabilistic lens model depicts the process by which exhibited behaviors are perceived by others. Adapted from “Probabilistic functioning and the clinical method” by K. R. Hammond, 1955, *Psychological Review*, p. 260. Copyright 1955 by the American Psychological Association.

Kenneth Hammond first applied Brunswik’s probabilistic function theory to clinical judgment studies in 1955 (Doherty, 2006; Hammond, 1996). Hammond (1955) concluded that the clinician is not a reader of instruments, but in fact is an “instrument to

be understood in terms of a probability model” (p. 262). Hoffman (1960) subsequently formalized the modeling approach in his paper *The Paramorphic Representation of Clinical Judgment*.

The term policy-capturing appears to have originated at the Personnel Research Laboratory at Lackland Air Force Base, Texas (Stewart, Ward, Naylor, & Cooksey, 1998). Ward describes how the research team was attempting to capture the policies used by personnel counselors responsible for assigning servicemen to their respective jobs (Stewart et al., 1998, p. 3). The consensus of Stewart et al. appears to be that judgment analysis is the more general and correct term but that the term policy-capturing is widespread and useful for bibliographic searches (p. 1).

Although the predominant functional relationship described in the literature for judgment analysis is multiple regression, Hammond (1996) explains that any appropriate functional relationship could be used (p. 245). It is this fact that led this writer to the work of von Helversen and Rieskamp (2008). Von Helversen and Rieskamp developed and presented the metrics and mapping framework of Brown and Siegler (1993) as an alternative model of cognitive judgment applied to quantitative estimation. The metrics and mapping framework posits that quantitative estimates are a combination of ordinal and numeric data (von Helversen & Rieskamp, 2008, p. 73).

As a compound estimate of scope, resources, and time, reasonableness is a fundamentally quantitative, albeit dimensionless, estimate. This study built upon von Helversen and Rieskamp’s work by testing the metrics and mapping framework against a compound estimate. The research design thus invoked a conceptual framework that

merged the conventional linear additive theory informing judgment analysis with the cognitive models depicting human estimation to produce a parametric model of reasonableness.

### Definition of Terms

*Cognitive mapping:* A model of quantitative estimation based upon the total number of cues that map to known reference points.

*Cue:* A feature or indicator that an individual uses, consciously or subconsciously, as a factor in their reasoning process.

*Policy-capture:* A statistical method using linear regression to establish a functional relationship between decision factors and decision outcomes.

*Reasonableness:* An estimate of the degree of belief that a project can be completed within the constraints imposed by the scope, cost, schedule, and resources proposed without incurring unnecessary risk or expense.

*Research assessment:* An evaluation of a research project as a member of a portfolio or for consideration for new or continuing funding. It is not used in the context of post research evaluation.

*Research portfolio:* A collection of research programs, typically representing a diverse mix of high and low risk programs, that is managed by a company or an organization.

### Assumptions

Assumptions are those factors influencing the study's results that are believed to be true. The assumptions for this study were that: (a) the participants reported their

experience accurately, (b) the cues selected were representative of the true factors that influence an estimate of reasonableness, and (c) the participants interpreted the scenarios in the manner that the researcher intended.

### Limitations, Delimitations, and Scope

#### *Limitations*

Limitations are factors which constrain the study but are not under the control of the researcher. The known limitations for this study were: (a) the contrived scenarios did not necessarily represent a realistic evaluation environment to the reviewer, (b) a convenience sample was used, and (c) self-administered electronic surveys were subject to coverage and nonresponse bias.

#### *Delimitations*

Delimitations are constraints imposed by the researcher to limit the scope of the research effort. The delimitations for this study included: (a) scenarios for the policy-capturing exercise were generalized and did not include domain specific references, and (b) a fractional factorial design (Aiman-Smith et al., 2002, p. 401) was used to reduce the number of scenarios presented to a single participant.

#### *Scope*

This study was an exploration of how people estimate reasonableness and how well six cognitive models of estimation can predict such estimates. A survey methodology was used to conduct a policy-capturing exercise across a self-administered convenience sample of graduate students, faculty, and staff participating in Walden University's participant pool for research studies. In the policy-capturing exercise,

participants were asked to judge the reasonableness of 50 randomly contrived scenarios in which the scope, resources, cost, and schedule parameters were varied. The mean-square-deviation between the cognitive models and the measured data was used to determine cognitive model performance. A web-based survey instrument developed by this writer was used for presentation of the scenarios and collection of the participant responses.

### Significance of the Study

From a management perspective, the benefit of this work lies in its potential to enhance the accuracy of research assessment and to improve single-rater reliability in the grant review process. There is no doubt that research assessment is fraught with uncertainty; maximizing value creation, however, requires that such uncertainty be mitigated. The first step is to understand the underlying research assessment processes. The policy-capturing exercise provided an empirical base from which we may begin to characterize at least one component, that being reasonableness estimation.

The second step toward mitigating uncertainty is to improve the reliability of the research assessment process. The six cognitive models evaluated in this study, or a novel variation thereof, may provide the key to standardizing research assessment in a rigorous and meaningful way. The use of decision support tools has been previously demonstrated to improve upon the success rate of conventional expert reviewers (Galbraith, DeNoble, Ehrlich, & Kline, 2007). Using tools incorporating components of the models described in this study, business management can benefit from greater confidence in portfolio analyses and more reliable project management.

With respect to our society at large, the positive social impact of this work can be found in potentially more efficient allocation and prioritization of increasingly scarce research funds in areas of science such as social, psychological, medical, pharmaceutical, and engineering. Not only can better decision support tools improve the research grant success rate by improving the reliability of the assessment process, they can improve transparency in the award process. The Washington Post reported that nearly 33% of the grants awarded from 2004 through 2009 to specialized HIV/AIDS support groups in Washington, D.C. have gone to programs “wracked by questionable spending, practices, and services” (Cenziper, 2009); the use of decision support systems in the grant assessment process can reduce the funds lost to fraud, waste, and abuse.

### Summary

In this first chapter, the concept of estimating reasonableness as a component of the overall research assessment process was introduced along with the underlying theory that supports how individuals make estimates, how research is assessed, and how methods for the assessment process might be improved through parametric modeling, artificial intelligence, and decision support systems. The need for research in this area is demonstrated by the relative lack of consistency in the research assessment process which relies heavily on the reviewer’s ad hoc estimation of key project parameters.

In chapter 2, current topics in research assessment, decision theory, cognitive reasoning and judgment, artificial intelligence, and their application to decision support systems are reviewed. The relationship of these topics to the research questions will be explored to create a foundation for the research design presented in chapter 3. Results of

the policy-capturing exercise and the analysis of the cognitive modeling are presented in chapter 4 followed by a discussion of findings, conclusions, and recommendations in chapter 5.



CHAPTER 2:  
LITERATURE REVIEW

Introduction

This literature review has touched upon more than 1200 articles, reports, and textbooks. The majority of the reference material has been obtained through Walden University's EBSCOHost subscription using the *Academic Search Premier*, *Business Source Premier*, *SocINDEX*, and *PsychARTICLES* databases. Additional reference material was obtained through the University of Maryland Baltimore County Albin O. Kuhn Library. A limited number of references were obtained through web searches of university document repositories, professional organization web sites, and use of the search engine CiteSeer<sup>X</sup><sub>beta</sub> sponsored by Pennsylvania State University and the Google<sup>TM</sup> Scholar document search engine.

The search strategy evolved over time beginning with keyword searches that included terms such as *peer-review*, *research assessment*, *decision theory*, and *research management*. Queries using these keywords led to a few initial papers that produced several more relevant keywords, but more importantly, additional papers. Tracing the referenced papers, as well as papers citing the referenced papers produced a wealth of information. Of particular value was the discovery of the *Journal of High Technology Management Research*, published by Elsevier, and the journal *Research-Technology Management*, published by the Industrial Research Institute, both of which focus exclusively on issues pertaining to research management and which yielded several key papers. As the topic began to focus on individual judgment and estimation, keywords

such as *judgment theory*, *cognitive reasoning*, and *policy-capturing* yielded papers that were more directly related to the questions at hand.

The most significant papers will be considered in this literature review, beginning with those papers that support this dissertation's thesis that a study of the research assessment process is warranted. An exploration of work in the foundational subjects of decision theory and cognitive reasoning follows, including a discussion of artificial intelligence. Although artificial intelligence may appear to be a peripheral topic, work in this area includes the implementation of the parametric models that are envisioned to be a product of this study. The literature review continues with a discussion of recent methodological studies that guide the research design proposed in chapter 3. A synthesis and summary conclude the review and introduce the remaining sections of this dissertation.

## The State of Research Assessment

### *Peer Review*

Despite claims of bias and poor reliability, peer review is the gold standard of research project and proposal assessment (Cookson & Jack, 2008); Marsh, Jayasinghe, and Bond (2008) suggest that peer review is so important that it is considered by some to be "sacrosanct, above reproach, and not subject to serious scrutiny" (p. 160). Peer review is established as the primary evaluation method for journal articles, grant proposals, dissertations, job applications, and even promotions, yet there is very little empirical research in this area (p. 160).

Addressing this issue, Marsh et al. investigated the peer review process using data obtained from the Australian Research Council. A total of 2,331 proposals and their corresponding 10,023 evaluations by 6,233 reviewers were studied for reliability and bias (p. 161); the authors reported a single-rater reliability of only 0.15 (pp. 161-162). In an earlier paper, Jayasinghe, Marsh, and Bond (2003) examined the same database of Australian Research Council grant proposals and determined that there was no evidence that single rater reliability was any better or worse with respect to the hard sciences versus the social sciences (p. 297).

Laudel (2006) examined the peer review process from an interdisciplinary perspective noting that the lack of research into the peer review process can be attributed to “the general reluctance by funding agencies and assessors to grant access to their ‘black box’ [review process]” (p. 67). Nonetheless, Laurel was able to perform a qualitative case study of the process from which he concluded that interdisciplinary tension between applicant and reviewer specializations impacted the peer review process (p. 67); no specific recommendations, however, were made to overcome this issue.

Seeking to enhance the peer review process, Obrecht, Tibelius, and D’Aloisio (2007) tested the notion that a committee discussion phase would improve reliability; they found that although it was counter intuitive, adding a committee discussion phase to the review process did not improve reliability. The authors studied the records of 775 fellowship applications, of which 157 were awarded, surveyed 46 committee members, and used 6 committee members in a focus group (p. 82). They concluded that peer

reviews should be based upon a structured review of the best available predictors (p. 89); although they did not elaborate upon what those predictors might be.

Ennis (2007) explored the use of an enhanced peer review process to improve rater reliability. Enhanced peer review was derived from the Calibrated Peer Review (CPR™) developed at UCLA for scoring the Critical Thinking Assessment Test (pp. 1-2); the concept behind CPR™ is that reviewers are trained using exemplars on how to consistently apply the scoring rubric. Enhanced peer review differs from CPR™ in that it is group oriented and focused on critical thinking. Although Ennis' study was primarily concerned with the impact of critical thinking skills training upon the reviewers themselves, the technique illustrates the potential reliability improvements through the use of exemplars and simple training exercises.

Other writers have recently discussed the need for improving the peer review process (Bellingan, 2007; Schwartz, Mastin, & Martin, 2006). Bellingan advocates the improvement of peer review in an editorial about biological research while criticizing cost saving policy changes being proposed by the Research Counsel of the United Kingdom (RCUK). Of particular interest in Bellingan's editorial is the radical nature of the RCUK changes that include prohibiting resubmission of proposals, limiting proposal submissions, consolidating funding, and pre-screening submissions prior to full review. Bellingan argues that only pre-screening has merit (2007, p. 8), but as will be discussed later, such screening exercises are performed rapidly with less information and must therefore be performed carefully.

In a similar editorial, Schwartz, Mastin and Martin (2006) cite the need for improving peer review “to greatly facilitate our ability to identify and fund the best research” (p. A270). As with so many papers on this subject, Schwartz et al. hail the peer-review process as essential, acknowledge its limitations, and call for its improvement. Banghart, having completed an extensive analysis of peer review in a social sciences context, calls for further research noting that “the consequences of the peer review process are so broad reaching that it is important to better understand the process and to ensure that the process performs the task it is intended to perform” (2006, p. 105).

The need for research in the area of peer review is well established, both in terms of the understanding yet to be gained and the significance to the research community at large; as Bellingan concludes, peer review is “vital” to the decision making process (2007, p. 8). In the next several sections, current research into the application of the peer review method and its implementation will be explored with an emphasis on the estimation of a project’s reasonableness as a component of the peer review process.

### *Portfolio Management*

Although peer review is commonly associated with the publication of journal articles, the peer review method, along with its use of scoring criteria, is deeply rooted in other important research assessment practices. Portfolio management is the practice of selecting and maintaining a diverse collection of research and development programs that balances risk against potential return. A key aspect of portfolio management therefore is the technology assessment that must be performed in order to rank and classify candidate programs.

Cooper, Edgett, and Kleinschmidt (2001) performed an exploratory study of over 200 businesses in North America to identify the types of portfolio management used in industry and their relationship to portfolio performance. Financial methods based upon net present value and return on investment and strategic fit dominated the results; the authors found that over 77% of the companies surveyed used financial methods and over 64% used strategic fit as components of their portfolio analysis (p. 367).

Faulkner (1996) describes the use of the Black-Scholes option pricing formula as one means of valuing a research portfolio. Options pricing models are derived from the investment banking industry and recognize the degree of uncertainty as an opportunity for financial gain rather than a risk of loss (§5). In a more recent paper, MacMillan, Van Putten, McGrath, and Thompson (2006) contended that the Black-Scholes formula overvalues R&D projects and suggest instead the Mac Van Adjusted Option Method using real options when cost uncertainty is high (p. 36).

In contrast De Piante Henriksen and Palocsay described the financial methods as formidable, requiring a trained analyst, and “largely ignored by real-world R&D organizations” (2008, p. 530); De Piante Henriksen and Palocsay pointed instead to scoring methods as the primary evaluation tool (p. 530). The difference lies in the scope of the studies; Cooper et al. focused on technology development businesses which necessarily have the interest and resources to perform a full financial analysis whereas De Piante Henriksen and Palocsay addressed the larger R&D community that includes academic and pure research.

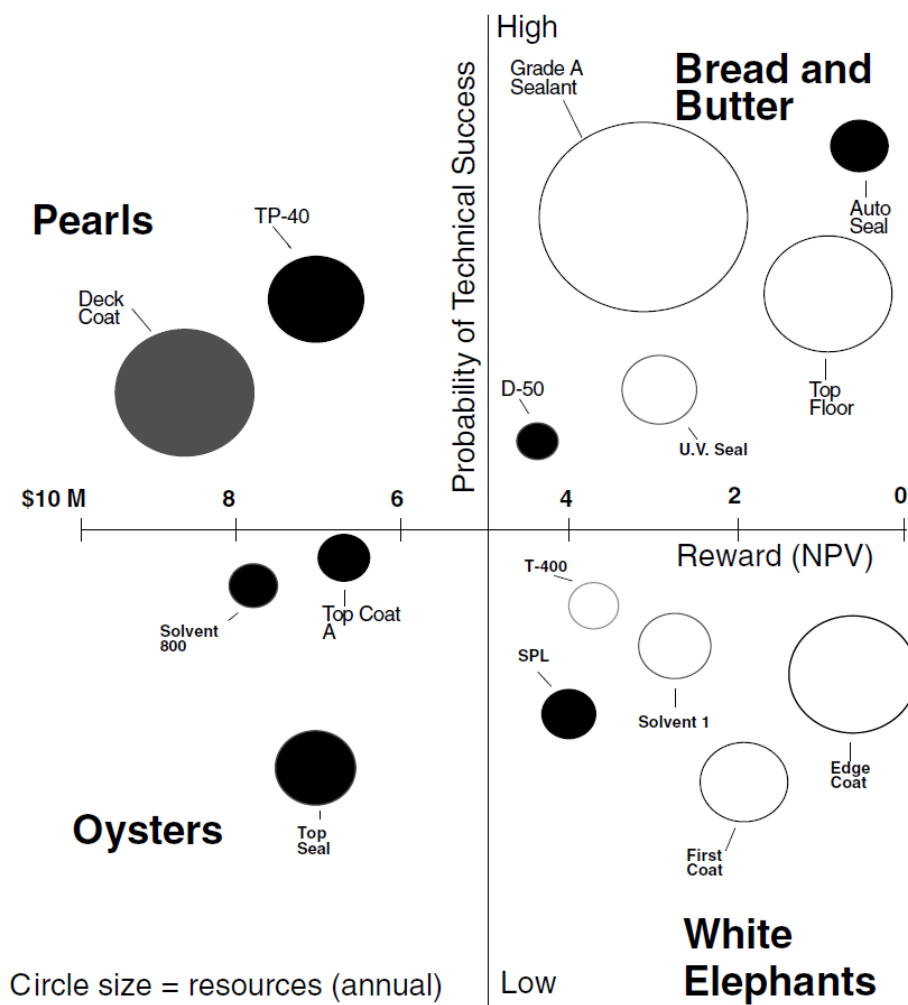
Cooper et al. found that nearly 38% of the businesses surveyed incorporated a scoring system in which various criteria are estimated on a Likert type scale (p. 371). A well documented example of a research portfolio scoring system is the PORTMAN system that was developed by the RAND Corporation for the Office of Naval Research (Silberglitt et al., 2004). Candidate projects are scored by a team of experts in terms of *Capability*, *Performance Potential*, and *Transition Probability* (p. xiii). *Capability* is a measure of the project's alignment with the organization's goals; *Performance Potential* is nominally equivalent to the project's expected return; *Transition Probability* is an estimate of the likelihood that the project will succeed based upon technical risk *and* plan realism.

Of particular relevance to this discussion is the use of explicit scale descriptions in the scoring criteria to rate the reasonableness of the project plan. Silberglitt et al. crafted a set of six questions, two per criteria, and developed descriptive scales that incorporated the relevant factors for each question. For example, question 6 asks the reviewer to rate the transition plan for the project by choosing one of the following:

1. Well conceived and appears to be implementable.
2. Some problems with cost, schedule, or other fielding burdens.
3. Major problems with cost, schedule, or other fielding burdens.
4. Is severely flawed or nonexistent (Silberglitt et al., 2004, p. 19).

Cooper et al. also cited bubble diagrams as a popular supporting tool used by 41% of the businesses surveyed (2001, p. 371). Bubble diagrams map a project against three criteria; two of which represent the X and Y axis of the chart, and the third criteria

depicted by the relative size of the bubble. The most common bubble chart places risk on the X axis and reward on the Y axis with the resources required to implement the program reflected in the size of the bubble (p. 372). Estimates of risk, return, and resource consumption are thus required to populate the chart. A simple illustration of a risk/reward bubble chart is shown in Figure 2.



*Figure 2.* A simple illustration of a bubble chart projects ranked by risk, return, and resources. From “Portfolio management for new product development: results of an industry practices study” by R. Cooper, S. Edgett, and E. Kleinschmidt, 2001, *R&D Management*, 31(4), p. 369. Reprinted with permission.



Portfolio management is important because it can be shown that businesses with a managed mix of radical high risk / high return programs and incremental low risk low return programs perform better than those that do not (P. Srivastava, 2006, p. 82). In both the scoring approach, and the bubble chart approach, the research assessment is performed by using estimates generated by domain experts. Even the highly formulaic Black-Scholes and Mac Van options valuation methods described at the beginning of this section ultimately rely on subjective estimates of risk (Benaroch, Jeffery, Kauffman, & Shah, 2007, p. 133). Srivastava, Towery, and Zuckerman (2007), in a survey of federal portfolio management practices, concluded that the need to improve upon these techniques is supported by a “growing consensus among science and policy makers that theories, models, and tools borrowed from other fields can and should be used to make management of research portfolios more scientific” (p. 153).

### *Project Selection*

Project selection has a strong peer review component and many similarities with portfolio analysis. The level of detail however is greater because project selection requires more information than simply risk and reward as might be used in a portfolio bubble chart. Consistent with the previous discussion on the unreliability of peer review, with respect to picking successful projects in the technology business, “expert reviewers provide little or no predictive value” (Galbraith et al., 2007, p. 134). In a study of 69 post 9-11 high technology defense programs evaluated over a period of 3 years, Galbraith et al. compared the predictions of expert reviewers against three parametric models, a

Bayesian Data Reduction Algorithm, an Artificial Neural Network, and a linear regression algorithm.

The experts scored each program in terms of reasonableness, technical merit, potential return, competing technologies, and ability to deliver by using an 11-point Likert scale (Galbraith et al., 2007, pp. 127-128). The parametric models used eight variables derived from the project plan that included age of the firm, number of employees, annual revenue, business diversity, education of the team, stage of development, external funding, and partnerships (p. 128). A program was considered successful if a product was launched, additional SBIR funding was received, or a cooperative development license was received (p. 128).

Galbraith et al. found that expert reviewers provided no statistically significant contribution to the parametric models; their results indicated that reasonable selection models can be created from organizational variables alone (2007, p. 134). The authors concluded that “the overall validity of early stage screening by experts needs further investigation” (p. 134). It is important to note that Galbraith et al. point out that their results are not representative of all expert assessments; whereas their study addressed a screening phase, the experts would have had more information in a full review or due diligence investigation (p. 135) .

Scoring by expert review panels, however, remains a fact of research life; relying heavily on estimation, reviewers are asked to score projects on a number of criteria using Likert scales or by assigning points. Scoring is flexible, straightforward, and can produce a credible result. De Piante Henriksen and Palocsay (2008) contended that scoring “has

been found to be the most widely used quantitative technique for evaluating R&D projects” (p. 530).

De Piante Henriksen and Palocsay created a Microsoft Excel<sup>®</sup> based decision support tool that integrates selection criteria scores, criteria weights, and program costs to produce an aggregate merit score and rank. A six question survey is used to rate *risk*, *relevance*, *reasonableness*, and *return* (2008, p. 532). Of interest is the fact that De Piante Henriksen and Palocsay make a distinction between research return, programmatic return, and business return, recognizing that a given project may have value from an academic or program perspective that does not directly translate to the financial bottom line.

### Decision Theory

A decision is any judgment that involves a choice between two or more options – should I buy the red shirt or the blue shirt? Is this the correct route? Is the score on this rating a low, a high, or something in between? Decision theory is the study of how people make decisions. Decision theory emerged as a discipline during the latter half of the 20<sup>th</sup> century drawing upon the sciences of economics, statistics, psychology, sociology, operational research, and philosophy (Hansson, 2005, p. 6).

Hansson (2005) described two primary approaches to the field; the first is *normative* and the second is *descriptive*. Normative theory describes the way that a decision should be made whereas descriptive theory describes how a decision is made (p. 6). This dissertation is concerned with both approaches as it is necessary to understand how a reviewer estimates reasonableness in order to improve upon the way that the reviewer should estimate reasonableness.

The research to initially be considered in this literature review addresses the normative approach to decision theory. The field is rich with current studies into Bayesian methods, analytic networks, heuristics, expected utility formulations, and other models of decision making, several of which will be discussed here. Research that addresses the descriptive approach to decision theory will follow in the discussion of cognitive reasoning and judgment. Relevant studies into the implementation of both normative and descriptive theory are then explored in the section on Artificial Intelligence.

Meade and Presley (2002) presented a case study of the analytic network process as a project selection tool. The analytical network process is a generalized form of Thomas Saaty's analytical hierarchy process (p. 60). In the analytical hierarchy process, a decision is broken down into sub-goals or objectives, with each sub-goal further decomposed as necessary to reach specific measurable attributes; alternatives are then fed into the model and numerical comparisons are made at each node until a decision is reached at the top of the hierarchy (p. 60).

The analytic network process expands upon the analytic hierarchy process by introducing bi-directional relationships rather than strictly propagating the decision from bottom to top (Meade & Presley, 2002, p. 60). Relationships between attributes at the same level can also be represented to account for interdependencies among the various criteria. Meade and Presley applied the analytic network process to the specific case of a small high-tech company that develops high-speed printing presses (p. 61). The decision

to be addressed was whether to upgrade the current model or develop an entirely new model.

Meade and Presley (2002) present the complete derivation of the analytic network process model for the printer manufacturer and illustrate its use. The authors report that the manufacturer was satisfied with the model and ultimately selected an upgrade for the current printer (p. 65); they also suggested that additional work could have been done to develop the criteria and interactions used in the model. However, the authors did not report upon the success or failure of the manufacturer's decision to upgrade hence there is no data as to the reliability of the model. Most notably, despite the mathematics and formulation of the result, Meade and Presley note that the model is ultimately limited by the subjective opinions of the model user who is responsible for setting the criteria weights, and in some case, the criteria values (p. 66).

An approach to account for the subjective uncertainties encountered by Meade and Presley (2002) can be found in the use of *fuzzy logic*. Mateou and Andreou (2008) proposed a novel methodology decision support system using evolutionary fuzzy cognitive maps. The authors' approach integrated aspects of decision support systems, expert systems, fuzzy cognitive maps, and genetic algorithms. In this context, decision support systems assist the decision maker, expert systems replace the decision maker, fuzzy cognitive maps integrate the belief systems of fuzzy logic with the relationship processing capability of neural networks, and genetic algorithms optimize the weights and relationship coefficients by using random evolution.

Mateou and Andreou (2008) applied their fuzzy logic model to a simulation of the decision to accept or reject the 2004 Annan Plan that aimed to reconcile the political divide between the Greek and Turkish Cypriots on the island of Cyprus. As a decision support tool, the model allowed users to adjust various parameters to determine the effect upon the outcome, as an expert system, the model successfully predicted the rejection of the plan (pp. 161-162). It is important to note, however, that the model required domain experts to identify the fundamental parameters and to estimate weights and activation values (p. 161).

Chin, Yang, Guo, and Lam (2009) developed a decision analysis approach that builds upon conventional multi-attribute decision analysis by incorporating a belief decision matrix using the Dempster-Shafer mathematical theory of evidence (p. 143). The Dempster-Shafer theory is often interpreted as an extension of Bayesian probability theory (Bossé, Roy, & Wark, 2007, p. 176); its principal advantages being the representation of uncertain measures in terms of belief rather than probability and its ability to combine information intuitively (p. 176). Rakowsky (2007), in a tutorial based on Dempster-Shafer theory, described the model as being particularly well suited for simulating the “uncertain judgment of experts” (p. 579).

Chin et al. incorporated a unique feature into their model that accounts for interval-valued judgments; whereas conventional belief models specify a single value, interval-valued assessments allow the user to enter an estimated range of values (2009, p. 144). The model development and supporting mathematical analysis are well documented in sufficient detail to replicate the approach. The authors follow with a detailed case

study in which the model is applied to a design decision considered by a Hong Kong electrical appliance manufacturer.

As with the Mateou and Andreou (2008) case study, Chin et al. considered the case study to have been a success, but they were unable to generalize any conclusions about the validity of the model due to the case study's limited scope (2009, p. 153). Furthermore, the authors noted that the complexity of the model demands a software implementation (p. 153); the observation is consistent with the De Piante Henriksen and Palocsay's (2008) argument discussed earlier that such models don't get much use for precisely this reason (p. 530).

The last of the normative models to be considered in this review is based upon a traditional Bayesian probabilities model but it incorporates a novel statistical method to account for uncertainties and incomplete data. Hovanov, Yudaeva, and Hovanov (2009) model the "uncertain choice of admissible (from the point of view of appropriate *NNN-information* [expert data] ) probabilities and weights by a random choice from corresponding sets of probabilities and weights" (p. 858). In other words, uncertain or incomplete data is modeled by the expected value of the information as aggregated from multiple expert sources.

Hovanov et al. presented a case study using the model based upon forecasting the change in share price of the Russian oil company LUKoil (2009, p. 860). Five sources of expert information were used as inputs and the model produced a "rather exact prediction" of  $\approx 12.48\%$  versus the actual value of  $\approx 13\%$  (p. 861). In an interesting twist, the authors make no attempt to explain the success of the model and in fact

conclude that the result is not meaningful (p. 861); they remain assured, however, that the model is effective but don't explain why (p. 861).

The central point behind this discussion on normative decision theory models is that despite the formulization and rigor of the normative models, they ultimately rely on human estimation of some aspect of the model. That isn't to say that the models are not effective; numerous reports suggest that decision models consistently outperform basic human judgment (Banghart, 2006; De Piante Henriksen & Palocsay, 2008; Galbraith et al., 2007; Karren & Barringer, 2002; Marsh et al., 2008; von Helversen & Rieskamp, 2008). In the next section, cognitive models that address how people actually make decisions as opposed to how they should make decisions will be explored.

#### Cognitive Reasoning and Judgment

Broome (2008) described reasoning as “a process that starts from some mental state of yours, and brings you to a new mental state” (p. 121). Such a definition would be considered minimal by Cooper (1987) who suggested that a more complete definition includes the notion of utility maximization (p. 395). Rolf (2006) defined reasoning to include symbolic representation of the past, present, and future: “most of our knowledge and decisions rely on our capacity for reasoning about matters not immediately present to our senses” (p. 136).

Studies of reasoning date back to Aristotle and his discussions of logic and rational choice (Drolet & Suppes, 2008, p. 31; Sahlin & Vareman, 2008, p. 42). Judgment analysis, as a product of reasoning, has its origins more recently in the 20<sup>th</sup> century beginning with the work of Egon Brunswik and his theory of *probabilistic functionalism*



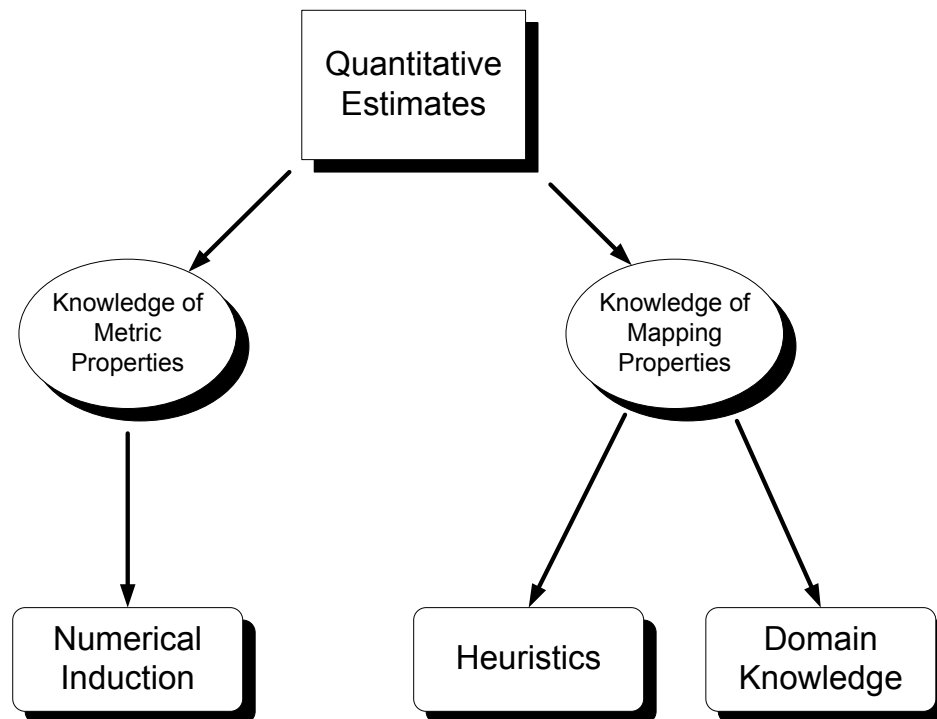
(Brunswik, Hammond, & Stewart, 2001, p. 3). One uses reasoning when deciding whether to buy a new car or take a new job; judgment is a less concrete term in that it implies reasoning with abstract, uncertain, or incomplete information. Whereas reasoning studies typically deal with logical arguments and conclusions, judgment studies involve problem solving, comparisons, and evaluations (Poltzer, 2004, p. 94).

This dissertation draws upon the literature in cognitive reasoning and judgment to study how people make judgments with respect to estimation; estimation is the reasoning process by which people decide upon a quantitative value judgment from incomplete and uncertain information. A foundational study in quantitative estimation is the work of Brown and Siegler (1993) who developed a general framework, referred to as the *metrics and mappings framework*.

Approaching the issue from a psychological perspective, Brown and Siegler (1993) considered heuristics, reasoning, and induction as elements of a theory for quantitative estimation that integrates the numerous decision and reasoning processes involved in producing a quantitative estimate (p. 512). Heuristics is the use of rules and formulations as a problem solving methodology; domain expertise is the application of specific relevant knowledge; intuitive statistics refers to “how people induce descriptive statistical properties” (pp. 512-513).

Brown and Siegler proposed that all three approaches are involved; the core theme is that two kinds of quantitative information are required to form a real-world estimate: (a) metric knowledge, and (b) mapping knowledge (1993, p. 514). Metric knowledge is numeric and statistical in nature; mapping knowledge is ordinal and relative

in nature; both describe general characteristics as opposed to specific information about an entity. Figure 3 illustrates the framework and the relationship of its components.



*Figure 3.* Conceptual illustration of Brown and Siegler’s Metrics and Mapping framework. From “Metrics and mappings: A framework for understanding real-world quantitative estimation” by N. R. Brown and R. S. Siegler, 1993, *Psychological Review*, 100(3), p. 514. Copyright 1993 by the American Psychological Association.

The central premise of Brown and Siegler’s work is that quantitative estimates are formed from the combined knowledge of the distribution of the quantitative values, and the relative ordering of entities within that distribution (1993, p. 529). Their experiments found that metric information and mapping information are independent of each other; different types of information can affect either metric or mapping process, both, or neither (p. 529). Brown and Siegler contended that their model was generalizable to any quantitative estimation (p. 529).

Unfortunately, Brown and Siegler did not reduce their metrics and mapping framework to an operational form. To apply the metrics and mapping framework to real-world estimation problems, Von Helversen and Rieskamp (2008) developed a computational form of the model and tested it by asking university students to estimate the toxicity of a collection of fictional bugs. The computational model first established the ordinal mapping relationship of the reference objects based upon the sum of the cue values that are present and groups objects with like numbers of cue values. The model then established a metric value for each group with like numbers of cues by averaging the criteria of interest across the reference objects in the group. The criteria value of a new object was estimated by counting the cues, and assigning the average value of the group with a like number of cues (von Helversen & Rieskamp, 2008, p. 74).

Von Helversen and Rieskamp illustrated the model using an example involving cell phones (2008, p. 74). Consider four reference phones A through D with known prices and various features that may include a digital camera, internet access, low weight, large display size; the objective is to estimate the price of two others, Psi and Omega, based upon the relevant cues. The information is detailed in Table 1; estimated values using multiple regression, exemplars, and a heuristic model called QuickEst are also provided for comparison.

Table 1

*Cellular Phone Value Estimations Using Metrics and Mapping, Linear Regression, Exemplars, and Heuristics.*

	A	B	C	D	Psi	Omega
Cue	Cue Values					
Camera	-	-	-	√	√	√
Internet	-	√	√	-	√	-
Weight	-	-	√	√	√	-
Display	√	-	-	√	-	-
	Selling Price					
Criterion	10	20	30	100	?	?
	Modeled Price (estimations)					
Mapping	15	15	30	100	100	15
Regression	10	20	30	100	110	90
QuickEst	15	15	20	50	30	15
Exemplar	10	20	30	100	30	43

*Note.* Adapted from “The mapping model: A cognitive theory of quantitative estimation” by B. von Helversen, and J. Rieskamp, 2008, *Journal of Experimental Psychology: General*, 137(1), p. 74. Copyright 2008 by the American Psychological Association.

In this example, phones A and B each satisfy one cue, and thus belong in the same group; the criteria value for the group is averaged between the two yielding a value of 15. Phones C and D have two and three cues, respectively, so each phone represents a group with a criteria value corresponding to the single entry in each group. Phone Psi has three features, and thus assumes an estimated value of 100, corresponding to the value of Phone D, the lone entry in the group. Phone Omega has one feature and assumes an estimated value of 15, corresponding to the average of the two phones in the single cue group.

Von Helversen and Rieskamp (2008) found that the metrics and mapping model outperformed other conventional models when the criteria being estimated followed a J-shaped curve (p. 89). In a linear environment, the model performed equally well with the regression model and was only outperformed when the criteria was perfectly linear (p. 89). In view of the success of the model and their experimental approach, von Helversen and Rieskamp's paper will form the foundation of this dissertation's study into estimations of reasonableness. Their work, however, only considered a one-dimensional estimation problem using dichotomous cues; this work will evaluate the metrics and mapping model as applied to a multi-dimensional problem using continuous cues:

$$\textit{Reasonableness} = f(\textit{scope}, \textit{resources}, \textit{cost}, \textit{schedule})$$

Scope is a measure of the effort necessary to complete the task, resources is an estimate of the tools, staff, facilities, and experience available, cost is the projected cost of the program, and schedule is the projected time allocated to complete the task. The primary limitation of the metrics and mapping model identified by von Helversen and Rieskamp is the reliance on historical values (p. 91); the model has no ability to extrapolate, however, this behavior is consistent with human inability to extrapolate values in anything other than a pure linear environment (p. 91). Other concerns included performance in the presence of complex criteria functions and non-binary or continuous cue values.

Von Helversen and Rieskamp's (2008) model will be applied in this dissertation to a policy-capturing study of research and development estimation. Policy-capturing is a

method for determining the functional relationship between decision parameters and decision outcomes; it is concisely defined by Aiman-Smith et al. (2002) as:

Policy-capturing is a regression-based decision-capturing methodology in which participants are asked to make decisions in response to a series of decision or problem solving scenarios presented by the researcher. The researcher regresses the decision outcomes on the values of one or more cues embedded in the scenarios and then uses the resulting regression weights to make inferences about the respondents' judgment policies. (p. 390)

Related research has been conducted in a number of areas. For instance, Beckstead (2007) examined the effects of type II errors as a function of the number of cues; LaVoie, Bourne, and Healy (2002) evaluated the effects of *seeding*, or training with known data points, upon the reliability of the metrics and mapping model; Newell and Broder (2008) look closely at the interdisciplinary connections that tie cognitive modeling and decision theory together, and Huang and White (2005) present a policy-capturing study of relevance judgments using five criteria.

Karren and Barringer (2002) identified 37 studies over the previous 25 year period that used policy-capturing as the primary decision analysis tool (p. 339). Whereas the cognitive models described so far represent human reasoning and judgment as theory, the field of artificial intelligence operationalizes human reasoning and judgment as functional computational algorithms. Coppin (2004) defined artificial intelligence as “using methods based on the intelligent behavior of humans and other animals to solve complex problems” (§1.2). In the sections which follow, current research in artificial intelligence will be explored and considered in the context of reasoning, judgment, and estimation.

## Artificial Intelligence

A product of the research into cognitive reasoning and judgment is that people are generally poor decision makers but that with the aid of structured procedures, human decision making can be improved. Galbraith, DeNoble, Ehrlich, and Kline (2007) clearly demonstrated that computer models outperform experts in a preliminary screening scenario, much like what would be used in a preliminary project selection or portfolio analysis process (p. 135). The massive software and computing power introduced in the last 20 years makes it possible to provide even a field user with robust decision support tools (Rolf, 2006, p. 135).

Although the specific class of software referred to as decision support software is generally passive, in that it guides and directs the user to a decision (Mateou & Andreou, 2008, p. 151), there is much to be learned from the larger class of computer reasoning known as artificial intelligence. Artificial intelligence is the field of study devoted to replicating human reasoning by using computer algorithms. Artificial intelligence techniques can be used to enhance parametric modeling methods and decision support tools by improving upon the classic linear regression methods commonly used today.

Bolton, Astwood, and Campbell (2002) conducted a study evaluating the use of fuzzy logic as a policy-capturing approach; fuzzy systems have been described as better models of uncertainty in human reasoning than traditional probability models (p. 541). To evaluate the approach, the authors devised a policy-capturing scenario using three dichotomous variables to describe the usability of several advance distance learning applications; a total of eight scenarios were required to represent a fully orthogonal

combination of the variables. The scenarios were presented in each of six contexts for a total of 48 scenarios. A questionnaire was used to assess the participant's judgment regarding each scenario. (p. 542).

Bolton et al. performed a policy-capture using both linear regression techniques and a fuzzy model implementation. Their initial results did not find a statistically significant difference between the fuzzy logic approach and the traditional linear regression model (2002, p. 543). The authors qualify their results however, noting that they only had six participants in the study; G\*Power 3.0 calculates a sample size of 77 would be required for  $\alpha=0.05$ ,  $effect\ size=0.15$ ,  $power=0.80$ ,  $predictors = 5$  (Faul, Erdfelder, Lang, & Buchner, 2007). Bolton et al. concluded that fuzzy logic has at least a *superficial* consistency with human reasoning processes (2002, p. 543).

Dorsey and Coovert (2003) performed a similar experiment seeking to test the hypotheses that fuzzy models performed better than linear and nonlinear regression models (p. 3). The study involved 10 participants in a policy-capturing exercise judging five variables presented in 110 scenarios. In this study, the participants were evaluating merit pay based upon performance, group performance, importance, tenure, and current salary level (pp. 4-5).

Unlike the previous study by Bolton et al. (2002), Dorsey and Coovert (2003) were able to establish a significant performance advantage over both linear (p. 8) and non linear models (p. 9). The authors acknowledged that the improvement, although significant, was relatively small (p. 9). Some of this improvement may be due to the fact that the fuzzy models used more free variables than used in the regression models.



Dorsey and Coovert suggest that ultimately “the power of fuzzy modeling may be in the flexibility afforded to model judgment strategies of arbitrary complexity” (p. 9).

A key feature of fuzzy logic lies in its ability to represent abstract quantities such as High, Medium, and Low in terms of membership functions that describe the relative degree of belief in the value (Dorsey & Coovert, 2003, p. 2) . Fuzzy models are implemented using inference engines that operate in five basic steps: (a) *fuzzification*, (b) application of *fuzzy rules*, (c) generate outputs from the fuzzy rules, (d) combine outputs and *infer* a result, and (e) *defuzzify* the result (p. 3). Fuzzification is the conversion from a specific value to a relative value within a fuzzy set using a *membership function*; defuzzification is simply the reverse process. A membership function defines the degree of belief in a value on a scale from 0 to 1; fuzzy rules then calculate output values based upon the degree of belief. Thomaidis, Nikitakos, and Dounias (2006) provide a concise and informative background on fuzzy sets and calculations.

There are numerous examples of fuzzy logic based decision support systems. Vasant (2005) describes a multi-criteria manufacturing decision support system using eight variables and a logistic *S-curve* membership function. Thomaidis et al. (2006), mentioned previously, developed a multi-criteria decision system for evaluating information technology projects. Huang, Chu, & Chiang (2008) applied fuzzy sets to the analytic hierarchy process in support of a research and development project selection tool. Yu, Wang, & Lai (2009) use fuzzy sets to evaluate financial risk in a group decision making context.

Mateou and Andreou (2008) introduce yet another dimension to fuzzy processing and artificial intelligence through the use of genetic algorithms. Genetic algorithms self optimize through random mutation of the various membership functions and rules to achieve the best possible model performance. The studies listed are merely representative of the work being performed in artificial intelligence; their inclusion in this review is to inform the study of new and different ways to implement the anticipated parametric model of reasonableness.

## Methodology

### *Policy-Capturing*

The implementation of this study is as much a topic of current research as is the subject matter itself. Politzer (2004) cautions that “from a methodological point of view, the experimental study of thinking is among the most difficult in cognitive psychology to carry out” (p. 16). The experimental representation of the scenario, the environment, and the execution of the test can interfere significantly with the cognitive process being evaluated. Politzer suggests two precautionary measures that include a macroanalysis and a microanalysis (p. 16).

Macroanalysis seeks to determine the participant’s preconceptions regarding the expectations of the researcher with respect to the participant’s ability to perform the task. Macroanalysis is performed in the context of the specific problem content and the relationship between the researcher and the participant (Politzer, 2004, p. 16). Microanalysis then examines the processes used by the participant to remove ambiguities,

resolve implied meanings, and establish referential assignments in the course of interpreting the problem or its premise (p. 16).

Notwithstanding the challenges identified by Politzer (2004) the survey-based policy-capturing approach remains the method of choice for this dissertation. The technique is well documented by Aiman-Smith et al. (2002) and also by Karren and Barringer (2002). Aiman-Smith et al. (2002) presented a well organized tutorial that addresses research design, implementation, and interpretation of policy-capturing studies. Karren and Barringer (2002) provided an extensive review of previous policy-capturing studies and go into considerable depth on study realism, cues, sample size, and validity.

Aiman-Smith et al. (2002) instructed the novice policy-capturing researcher to consider (pp. 396-400):

1. The number of cues and scenarios: more than five cues become unrealistic, particularly if an orthogonal design is used.
2. The values of the cues: when using *polychotomous* cues, the number of cues values will increase the number of scenarios required and hence reduce the number of independent cues that can be tested.
3. The range of the cues: cue range distributions should be consistent and normalized if possible; if some cues have wide ranges, and others narrow, the participant will focus on the wide range cues.
4. Demand effects: introduce a small amount of extraneous information to reduce the likelihood that participants will deduce the nature of the test.

5. Cue correlation: the need for realistic scenarios involving cues that are often correlated must be traded off with the potential adverse effects of multicollinearity.

6. Length and complexity of scenarios: scenarios must be long enough to be interesting and realistic, but not so long as to induce boredom.

7. Start-up effects: experience indicates that it requires 8 to 12 responses to begin generating stable results. A short training session is recommended to familiarize participant with the process.

8. Fatigue: Fractional designs have been shown to be less fatiguing than fully orthogonal designs, but require more participants. In any case, the maximum number of scenarios should not exceed 100.

9. Number of respondents: In a nomothetic study, a statistical power analysis based upon a large effect size can be used. Nomothetic studies typically achieve a power level equal to traditional survey methods with fewer observations per group.

10. Post task Data Collection: a series of post-task open-ended questions about the test, understanding of the cues, and reasoning process is recommended to facilitate interpretation of the results.

Aiman-Smith et al. acknowledged that a policy-capturing study requires a number of tradeoffs and they direct readers to a number of sources for additional information. Despite its complexity of implementation, the method can provide valuable insight for understanding a decision maker's reasoning process. Attention to detail will enhance the validity of a policy-capturing design and the value of the knowledge gained (2002, p. 408).

*Alternative Approaches*

Whereas the survey methodology in general and the policy-capturing exercise more specifically are relatively straightforward to implement, other approaches have been used to study judgment and decision making practices. Gehris (2008) for example, addressed the number of questions and the number of judges to be used in a policy-capturing study using a mathematical simulation to evaluate the relationship between the numbers of questions, the number of judges, and model performance. The policy-capturing model used by Gehris was designed to capture the judgment patterns of a group of experts answering a series of fixed questions to establish and statistically characterize a knowledge base; this approach differs from the cue manipulation described by Aiman-Smith et al. (2002) but the simulation techniques provide an additional model for evaluating the metric and mapping concept.

Banghart (2006), Galbraith et al. (2007), and Marsh et al. (2008) each used available data to analyze the research assessment process. These types of studies can provide detailed insight into the assessment process but they are rare. In some cases, as in the study by Galbraith et al. (2007), the authors are working with partial data; only proposals that resulted in a grant award were considered in the analysis. Research assessment “is normally practiced in relatively small settings with secrecy and anonymity. The details of the findings of peers are generally not available for inspection” (Banghart, 2006, p. 4).

In the consideration of alternative approaches, if the intent of the study is to understand the essence of how people reason and judge something, then the incorporation

of qualitative methods is warranted. In a previous evaluation of methodologies for this study (Kight, 2009), the case study method and narrative method were identified as relevant qualitative methods. When qualitative and quantitative methods are combined, the resulting mixed-method model provides researchers the flexibility to leverage the strengths of each and minimize their weaknesses (Johnson & Onwuegbuzie, 2004, p. 16).

Johnson and Onwuegbuzie (2004) presented an overview of the mixed method approach with a detailed comparison and contrast between qualitative, quantitative, and mixed method paradigms. Of the strengths identified for the mixed-method approach, the ability to add insight and understanding to the study stands out among the others as applicable to this study. Of the weakness identified, the additional time required to incorporate the qualitative component into the study is significant.

Teddlie and Tashakkori (2006) described a four-criteria typology for classifying mixed method research based upon: (a) number of methods – monomethod or mixed method, (b) number of phases – monostrand or multistrand, (c) implementation process – concurrent, sequential, conversion, or combination, and (d) stages of mixed integration approaches - all stages, experimental only, or other combination (p. 13). Teddlie and Tashakkori recommend a sequential quasi-mixed method multistrand design approach for solo investigators as it is easier to manage the phases separately and the study is more predictable (2006, p. 22).

The distinction as a quasi-mixed-methods study stems from the structure of the study which collects both quantitative and qualitative data and incorporates both quantitative and qualitative analysis, yet does not integrate the inferences (2006, p. 25).

The authors make the distinction noting that the perceived difference between mixed model and mixed-method approaches is diminishing as evidenced by the revised definition in the *Call for Papers of the Journal of Mixed Methods Research* that reads “mixed methods research is defined as research in which the investigator collects and analyzes data, integrates findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or program of inquiry” (p. 15).

### Literature Synthesis

The above review of the literature concerns the aspects of the research estimation process that this study will examine. Figure 4 was created using a brainstorming process called mind mapping to illustrate the connections between these aspects from a broad interest in research estimation to assessments, methods, and scoring, to reasonableness and estimates of reasonableness. The problem space is informed generally by a broad range of disciplines, and the research design is built upon cognitive theories of estimation, policy-capturing methodologies, and artificial intelligence as an implementation.

This study addressed a notable gap in the literature concerning the estimation of what constitutes a reasonable proposal. The majority of books and papers reviewed by this writer that instruct readers on how to evaluate a research proposal cite reasonableness as a critical element in the evaluation but fail to explain precisely how to do that. This study aims to contribute toward answering that question. In the next chapter, the anticipated research design and study implementation will be explained in detail.

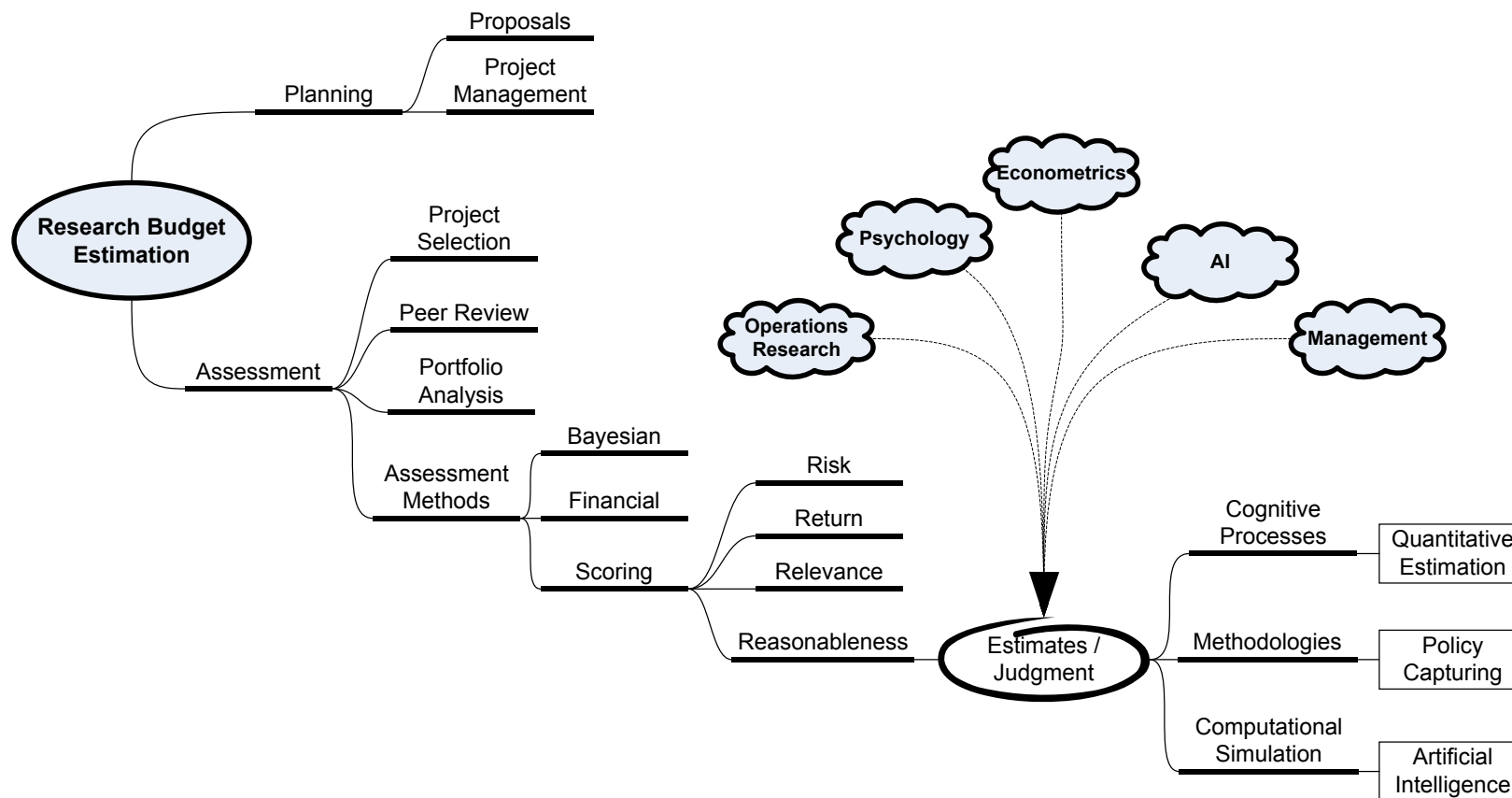


Figure 4. Mind map illustrating the logical relationship of various topics presented in the literature review.



## CHAPTER 3: RESEARCH METHOD

### Introduction

The purpose of this study was to gain a better understanding of how people evaluate reasonableness and to explore how these factors can best be modeled in a decision support system. This chapter describes the methodology employed in the study and the researcher's justification for using this approach. The chapter begins with a detailed explanation of the research design approach followed by a discussion of the target population and sampling methodology. The policy-capturing instrument was created by this writer for this application; its organization and content are described and sample pages of the instrument are illustrated in Appendix A. This chapter concludes with a description of the data collection procedures and the data analysis used to generate this study's findings.

### Research Design

The selection of a research design approach rarely happens in isolation; the process is typically iterative as the topic, research questions, and the review of the literature evolve. Such was the case with this research design which initially began with a mixed-methods approach but converged into a quantitative design. It was not for any deficiency in the qualitative approach anticipated for this study; on the contrary, there is much to be gained from a qualitative analysis. The decision was, in the spirit of mixed methods, a pragmatic one.

Singleton and Straits (2005) introduce their text on social research design with a list of the five principal issues that drive research design: (a) structure and state of the discipline, (b) social problems, (c) personal values of the researcher, (d) social premiums, and (e) practical considerations (pp. 43-44). Of these issues, practical considerations primarily influenced the final design approach. Resource limitations as well as “the skill of the researcher and the availability of relevant data ... will shape both the nature and scope of the problem that the researcher can pursue” (p. 44). In this case, the data required to perform a detailed narrative analysis of proposal evaluations was closely held and not available to this writer.

### *Methodology*

To proceed with the general topic of research assessment while still contributing in a meaningful way to the literature required that the data be created via empirical means. There were several options available; the four principal methodologies in quantitative social research included: (a) experimentation, (b) surveys, (c) observation, and (d) analysis of available data (Singleton & Straits, 2005, p. 153). Each methodology had its strengths and weaknesses hence the selection of a particular methodology was driven by the research questions, the purpose of the study, and the available resources.

In this study, the research questions addressed two different issues:

1. The establishment of an empirical baseline that reflects the measured relationship between the independent variables of *cost*, *schedule*, *scope*, and *resources* and the dependent variable *reasonableness* for the population of interest.

2. The evaluation of several cognitive models of reasoning and judgment to determine the quality of each model and to identify the model with the best performance.

The first issue describes a measurement task and no treatment was involved. As such, experimentation was ruled out for there was no causal relationship of interest to be tested. The analysis of existing data was also ruled out because the necessary data was closely held and not available. As a data gathering exercise, survey methods and field observation remain were the remaining possible options.

The nature of the data, however, did not lend itself to field observation without being particularly intrusive or burdensome. One could conceivably observe and interact with individuals as they were going about the process of assessing a research proposal, but such a process would have required an excessive amount of time and would likely have imposed upon the participants. As such, the remaining option to use survey methodology was determined to be the most practical way to gather the necessary information from a large pool of participants.

The second issue required the analysis of various cognitive models to determine their predictive performance as compared to the measured data. The models were developed in the data analysis phase of the project following data collection. The measured data was then edited, coded, and cleaned to format the data for use to eliminate extraneous entries or values. The models used in this study are described later in this chapter.

### *Approach*

The linear additive policy-capturing technique described by Aiman-Smith et al. (2002) was the foundation for this study's research design. The policy-capturing technique used a survey methodology to collect data on the estimation of reasonableness as a function of *scope*, *resources*, *cost*, and *schedule* by presenting a series of contrived proposal scenarios. The collected data was then compared to predictions generated by von Helversen and Rieskamp's (2008) metrics and mapping model of quantitative estimation along with several other popular cognitive models.

There are other influential or relevant cues in the estimation process that can vary with circumstance. Aiman-Smith et al. (2002) cautioned, however, against using more than five cues (p. 396). Their recommendation is consistent with Cowan's work on mental capacity which suggests that people can at best handle three to five chunks of information at a time (2001, p. 174). A total of four cues were therefore used in this study and they were based upon De Piante Henriksen and Traynor's (1999) review of over 50 papers on project-selection scoring systems in which they defined reasonableness as "whether or not the level of resources proposed will permit successful completion of the project objectives on time and within budget" (p. 163).

The focus of the policy-capturing exercise was to determine the relationship between the manipulated cues of scope, resources, cost, and schedule and the judgment criteria reasonableness. To isolate the evaluation of reasonableness from the evaluation of the cues themselves, the scenarios were presented as if the individual cues have been previously reviewed and rated by a team of experts. The scenarios used generic

descriptions and relative measures of scope, resources, cost, and schedule to eliminate a need for domain specific knowledge.

### Setting and Sample

#### *Target Population*

The primary sample was drawn from the population Walden University students, staff, and faculty enrolled as of September 2009 that were participating in the Center for Research Support participant pool. Additional participants working in research or management were also solicited directly by this writer. The participant pool is open to all Walden University students, faculty, and staff; since participation is voluntary, its composition may change over time, and therefore, precise demographic information can only be known at the conclusion of data collection.

Walden University is a U.S.-based online institution with a reported graduate student enrollment of 26,615 full-time and 5,358 part-time students (The Higher Learning Commission, 2009). The school comprises an ethnically and culturally diverse population primarily pursuing graduate level work. Furthermore, the nature of online graduate study attracts a population that is likely to be 30 years of age or older and employed (Ebersole, 2004). In this study, 42.5% of the respondents were 25 to 45 years of age, 57.5% were older than 45, and no respondents were under age 25.

#### *Sampling frame*

The sampling frame is by definition the “the set of all cases from which the sample is actually selected” (Singleton & Straits, 2005, p. 116). All members of the sampling frame were solicited to participate in the study. In this study the sampling frame

consisted of those members of the Walden Participant Pool and those individuals solicited by this writer that agreed to participate in the policy-capturing exercise and passed the eligibility screening questions.

### *Sampling Method*

A convenience sample was employed for the policy-capturing exercise based upon voluntary participants from the Walden University participant pool and individuals directly solicited by this writer. All respondents who joined the study and met the minimum eligibility requirements were included in the sample. Screening questions were used to improve the validity of the study to ensure that participants had the requisite program management and research planning experience.

The Walden University Research Center administers the participant pool and periodically advertises via email to student, faculty, and staff about opportunities to use the participant pool for both research and to support other researchers as participants. Joining and participating in the pool is completely voluntary; participants remain anonymous and are identified only by a system assigned identification number.

Upon log-in to the participant pool website, participants reviewed a short description of available studies and could register for studies they were interested in supporting. If the participant met the desired eligibility criteria, then instructions for completing the exercise were supplied. Researchers using the participant pool are not allowed to directly solicit Walden University students, faculty, or staff to join the pool or to participate in a specific study.

The study was open on the participant pool website for 60 days. During this period, additional qualified participants known to this writer were solicited via e-mail communication. A snowball sampling method was used, whereby initial contacts referred new contacts who refer additional contacts and so on (Singleton & Straits, 2005, p. 138) .

### *Demographics and Eligibility*

The ideal participant was a doctoral student or faculty member who was familiar with research programs and had experience in planning and estimating research programs. Participants were expected to be an ethnically diverse group of males and females representing ages 25 to 75 years of age. Eligibility was screened through a series of qualification questions that are listed in Table 2:

Table 2

#### *Eligibility Questions and Minimum Criteria for Participating in the Study.*

Eligibility Question	Minimum Criteria
Are you or have you ever been engaged in the development or evaluation of research proposals?	Yes
How many years of project management experience do you have?	2
How many years of experience do you have in planning or evaluating research projects?	2

### *Sample Size*

The power of a statistical test is a measure of the likelihood that the test will produce statistically significant results; or in other words, detecting a difference when one actually exists (Cohen, 1988, p. 1). Statistical power is tightly coupled with the desired significance  $\alpha$ , the sample size  $n$ , and the effect size ES (p. 14). In this study, the statistical test of interest is the multiple coefficient of determination  $R^2$ ; a measure from 0

to 1 of the proportion of variation in the dependent variable *reasonableness*, explained by the independent variables, *scope*, *resources*, *cost*, and *schedule*. The corresponding hypothesis test to determine if a significant relationship exists is then:

$$H_0 : R^2 = 0$$

$$H_1 : R^2 \neq 0$$

Aiman-Smith et al. (2002) recommend using power analysis to determine the sample size required for a policy-capturing exercise (p. 399). Using G\*Power version 3.1.0, as shown in Figure 5, to determine a multiple regression sample size for  $\alpha = 0.05$ ,  $effect\ size = 0.35$ ,  $power = 0.80$ , and  $predictors = 4$ , a recommended sample size of 40 was determined (Faul et al., 2007). The effect size of 0.35 was based upon the small, medium and large effect sizes of 0.02, 0.15, and 0.35 recommended by Cohen (1992) for multiple regression or partial correlation tests (p. 157). The choice of a large effect size was based upon Aiman-Smith et al. (2002) who note that because each respondent is making a series of judgments, thus reducing the standard error, “policy-capturing studies can be expected to yield larger effect sizes than more traditional survey research designs” (pp. 399-400). Upon collecting the desired number of responses, a post hoc power calculation was performed; the results of which are discussed in chapter 4.



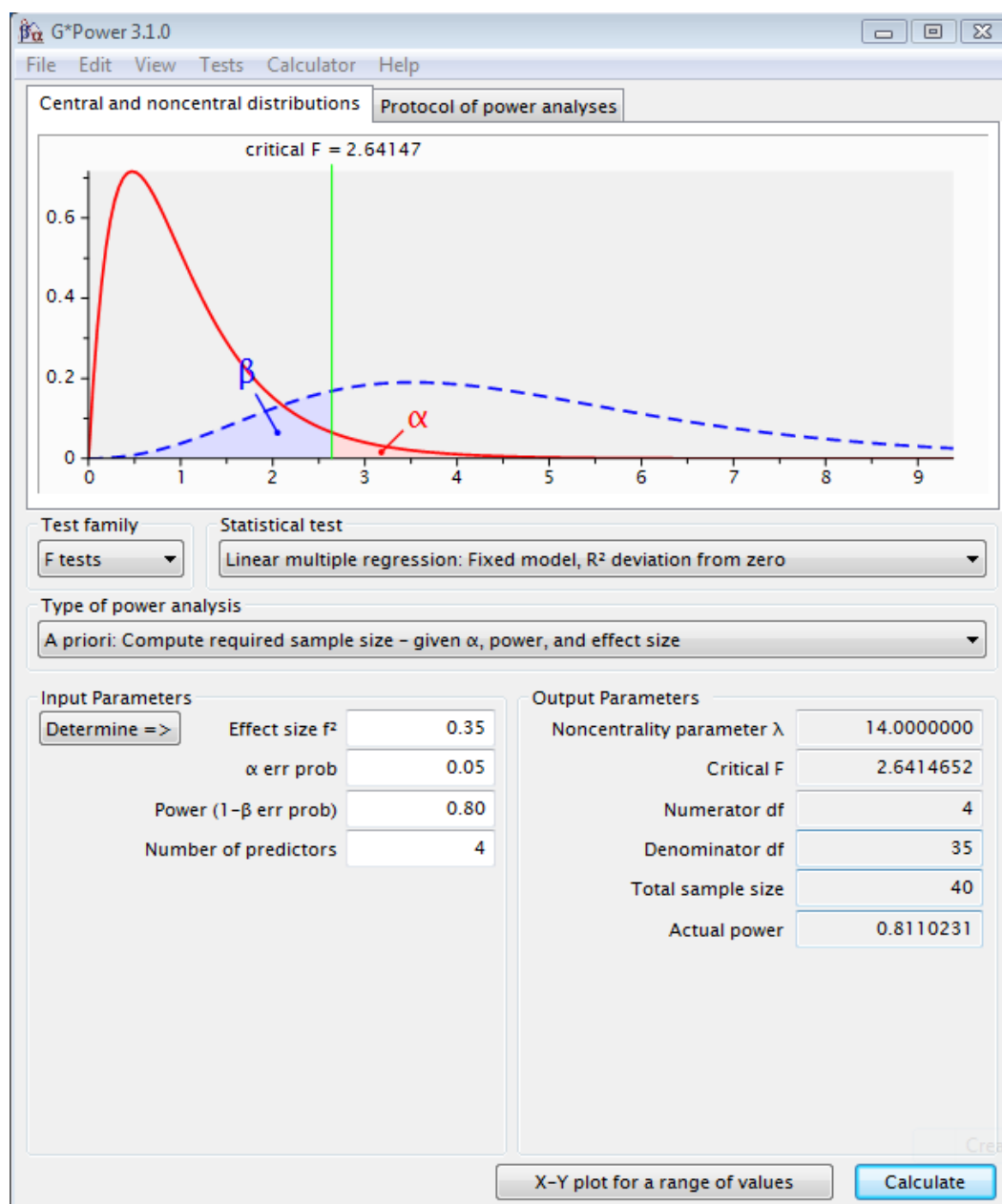


Figure 5. A priori power analysis of the measured data regression; produced with the computer program “G\*Power version 3.1.0” by Faul et al., 2007.

## Instrumentation

### *Policy-Capturing Instrumentation*

The policy-capturing exercise was performed using a custom survey-like instrument built with Python CGI scripts (Python Software Foundation, 2009a). Participants were directed to a web site which, upon proper authentication, prompted the participant through the policy-capturing exercise. The policy-capturing exercise comprised six sections:

1. Introduction and informed consent statement. The study would exit if the consent statement was not satisfied.
2. Instruction for completing the exercise.
3. Training questions – a total of 10 randomly selected scenarios intended to introduce and familiarize the participant with the structure of the exercise; respondents judged reasonableness on a 7-point Likert scale.
4. Policy-Capturing exercise – a total of 40 randomly selected scenarios will be presented, 4 of which will be repeated in the series; respondents judged reasonableness on a 7-point Likert scale.
5. Demographic questions.
6. Open ended post exercise assessment questions and instructions for obtaining additional information on the study.

Screen captures of each section of the exercise are provided in Appendix A. The participant was not aware of the transition from training questions to the actual questions used for the policy-capturing exercise. Scenarios that were repeated as part of the test

retest reliability measurement were given different names and descriptions so that participants would not recognize the duplication.

Scenarios were presented to the participant in the form a report describing a project that had been analyzed by a team of reviewer's and scored on each of the four cues. The cues were presented as ratings on a 5-point Likert Scale; fuzzy indicators were used for the cue scales as shown in Table 3. The participant, acting in the role of the senior reviewer, was then asked to score the reasonableness of the project using a numeric 7-point Likert Scale.

Table 3

*Fuzzy Indicators are used to Represent Cues on a 5-point Likert Scale.*

Cue	Likert Scale Indicators				
Scope	Insignificant	Incremental	Leading	Radical	Extreme
Resources	Very Weak	Weak	Moderate	Strong	Very Strong
Cost	Very Low	Low	Appropriate	High	Very High
Schedule	Very Short	Short	Appropriate	Long	Very Long

Participants were presented with a total of 50 scenarios. The number of scenarios was based upon the recommendations of Aiman-Smith et al. (2002) suggesting a 10:1 scenario-cue ratio (p. 398) and a start-up allowance of 8 to 12 scenarios (p. 399). Within the set, four scenarios were replicated in order to measure test-retest reliability (p. 400).

There are a total of 625 possible scenarios that can be constructed using four cues with five levels each. In a full factorial design, all 625 scenarios would be reviewed by each participant; in view of the impractical nature of such an exercise, this study used a confounded factorial design (Karren & Barringer, 2002, pp. 348-349) in which a randomly selected subset of the possible scenarios were reviewed by each participant.

The scenario list was created when the participant enters the exercise using the Python function `sample=random.sample(population,k)` thus each participant in the study evaluated a unique set of scenarios. Python's random number generation is based upon the Mersenne Twister algorithm (Python Software Foundation, 2009b)

### *Validity and Reliability*

Converting theory to actual practice so that the test can be performed is called *operationalization*; *construct validity*, is a measure of how well the operationalizations reflect the variables and concepts being tested. Construct validity incorporates a number of components; these components include subjective measures such as *face validity* and *content validity*, and objective criterion related measures such as *convergent validity* and *discriminant validity* (Singleton & Straits, 2005, pp. 98-103; Trochim, 2001, pp. 66-68).

Face validity is a simple judgment that the test appears to measure what it is intended to. Content validity addresses the completeness of the measure – is the test fully addressing and accounting for all of the relevant aspects of the measure. Although subjective measures such as face and content validity are useful, the objective criterion related measures stand up better to scrutiny by others (Singleton & Straits, 2005, p. 99). Convergent validity considers how closely the test aligns with similar tests or theoretical constructs. Discriminant validity, in the same fashion, would expect different results from tests that should theoretically be dissimilar.

Practices for enhancing the content validity of the policy-capturing technique can be found in the guidelines provided by Karren & Barringer (2002). Their relevant recommendations are listed in Table 4 along with their corresponding application to this

study. The key issue however is that the scenarios need to include “salient and realistically defined cues” (p. 338).

Table 4.

*Recommendations for Improving Validity of Policy-capturing Exercises.*

Recommendation	Implementation
Cognitive studies have demonstrated that decisions should be limited to four to eight variables (Karren & Barringer, 2002, p. 344).	This study uses the four variables cost, schedule scope and resources.
The validity of a policy-capturing exercise can be improved with thoughtful selection of the relevant decision criteria, preferably synthesized from a number of sources (p. 345).	Criteria selection based on published research by De Piante Henriksen and Palocsay (2008)
Cues should be presented with value ranges representative of the real attributes being simulated and be minimally correlated where possible and (p. 347).	Likert scales with realistic ranges were used. Maximum cross correlation was measured at -.068 between cost-scope and schedule-scope.
Elimination of unrealistic scenarios in a fractional factorial design can also contribute to the validity of the study (p. 345).	The study included all possible combinations; none were eliminated.

Reliability of the respondent answers can be checked by replicating scenarios throughout the survey and comparing same-person-same-scenario answers. In this study, 50 randomly selected scenarios are presented in sequence; the scenarios in positions 10, 11, 12, and 13 are then repeated in positions 20, 30, 40 and 50 respectively. Karren and Barringer (2002) have cited reliabilities between 0.75 and 0.90 using this technique in various policy-capturing exercises (p. 353). Karren and Barringer also cite a 1987 study by Hollenbeck and Williams in which a policy-capturing exercise was repeated one

month later, achieving a test-retest reliability of 0.72 (p. 353). An overall test retest reliability of 0.80 was achieved in this study.

### Data Collection

#### *Data Collected*

Demographic data collected included:

1. Age
2. Years of experience in preparing research plans and proposals
3. Years of experience in evaluating research plans and proposals
4. Education level
5. Professional Background – Academic/Government/Industry

Policy-capturing data collected included:

1. Fifty reasonableness estimates on a 7-point Likert scale.

Post-test data collected included open-ended questions for:

1. Description of the respondent's estimation process.
2. Respondent's opinions of the test instrument and procedure.
3. General comments and observations.

Data collected for file and record management included:

1. Randomly assigned unique session identifier.
2. Date/Time of the session.

#### *Data Collection Procedures*

Eligible participants were provided with a web site address, username, and password for access to the policy-capturing exercise. The participant was asked to allow

30 minutes to complete the exercise in a single sitting. The username and password were the same for all participants within two classes: (a) participant pool, and (b) direct invitation. It is not possible to identify participants by their logon or password information. Authentication was used to protect the site from accidental or malicious access to the exercise.

The exercise opened with an introduction and informed consent form; acknowledgement of the informed consent was required to proceed with the exercise. If the participant did not affirmatively acknowledge the informed consent, the session would exit. If the participant agreed with the statement of consent, the exercise proceeded to a page with detailed instructions on what to expect, how to interpret, and how to respond to the scenarios. The user was again be provided an estimate of the time required to complete the exercise which was expected to be less than 30 minutes.

A total of 50 randomly selected scenarios were presented to the participant. The four scenarios in the series at positions 10, 11, 12, and 13, were duplicated to measure test-retest reliability at positions 20, 30, 40, and 50 respectively. The position numbers refer to the location of a scenario within the series with the first scenario at position 1 and the last scenario at position 50. The replicated scenarios were given new names and descriptions but retained the same scoring. The participant was asked to rank each scenario using a 7-point Likert scale; a counter allowed the participant to monitor the number of scenarios completed.

The exercise proceeded to a series of demographic questions. Responses were not required to these questions and the participant could exit the exercise at this point if

desired or continue. The exercise closed with an open-ended request for feedback on the participant's thought processes, opinion of the exercise, and general comments followed by a thank you page and directions for requesting additional information.

### *Informed Consent*

All participants were required to acknowledge an informed consent statement prior to accessing the survey that addressed the following:

1. Background and purpose of the study
2. Procedures
3. Voluntary nature of the study; compensation if applicable.
4. Risks and Benefits
5. Confidentiality
6. Contacts – researcher and the IRB

### *Protection of Participants*

All data was collected anonymously, no identifying information was collected, and participation was voluntary. No risk of participation was identified. The participant could exit the study at any time and was under no obligation to participate.

### Data Analysis

#### *Descriptive Analysis and Characterization*

General descriptive statistics were prepared to characterize the raw data and associated demographics of the respondents. Cross correlation testing of the data was performed to determine if there were any potential collinearity issues with the primary cues. Maxwell (2000) cautions that cross correlation factors of 0.3 are not uncommon in



behavioral testing and that failure to account for multicollinearity can result in under sampled and under powered studies (p. 439).

### *Policy-Capturing*

Policy-capturing is a research technique that was first introduced by Hoffman (1960) as a means of understanding how “clinicians utilize information at their disposal to arrive at judgments or decisions” (p. 116). Hoffman proposed a linear additive model in which a decision  $J$  could be described as a sum of weighted information sources  $X_i$  where  $i = 1, 2, \dots, k$ :

$$J = A_0 + \sum_{i=1}^k A_i X_i$$

Determination of the coefficients  $A_{0\dots k}$  is accomplished using conventional multiple regression. The reference data points are collected in a policy-capturing exercise in which the subject(s) of interest are presented with a series of controlled scenarios; the number and nature of the scenarios are a function of the research objectives. Aiman-Smith et al. (2002) elaborate on the numerous issues pertaining to policy-capturing design and their work has formed the basis for this study design.

### *Nonlinear Models*

Whereas policy-capturing has been developed using a linear additive model, von Helversen and Rieskamp (2008) found that certain nonlinear cognitive models can outperform a linear-regression model when the judgment criteria follows a *power law* function (p. 91). The finding is significant because “power law distributions... are among the most prevalent distributions encountered in everyday life” (p. 77). This study sought

to determine if similar model performance could be attributed to the estimation of *reasonableness* in a research assessment application.

### *Parametric Models*

The six parametric models considered included: (a) Fuzzy Set Model, (b) Metrics and Mapping Model, (c) QuickEst Heuristics Model, (d) Linear Regression Model, (e) Exemplar Model and (f) Neural Network Model. Each of these had been proposed as a cognitive model of human judgment and was applied in this study to the estimation of reasonableness. The performance analysis of the models was patterned after the metrics and mapping study performed by von Helversen and Rieskamp (2008). The mean-square-deviations between the predictions of each parametric model and the policy-capture data were calculated and analyzed.

There are some differences in the computational form of the models used in this study versus those used by von Helversen and Rieskamp (2008). Von Helversen and Rieskamp used dichotomous (binary valued) cues whereas this study used polychotomous (multi-valued) cues; von Helversen and Rieskamp had anticipated this condition and suggested dichotomization of the cues as a possible adaptation (p. 91). The QuickEst model used by von Helversen and Rieskamp also required a similar adaptation. The polychotomous cues were dichotomized for both the metrics and mapping and QuickEst models as shown in Table 5.

Table 5

*Polychotomous to Dichotomous Cue Mapping.*

Polychotomous to Dichotomous Cue Model					
Scope	Insignificant 0	Incremental 1	Leading 1	Radical 1	Extreme 0
Resource	Very Weak 0	Weak 0	Moderate 1	Strong 1	Very Strong 1
Cost	Very Low 0	Low <i>Note 1</i>	Appropriate 1	High <i>Note 2</i>	Very High 0
Schedule	Very Short 0	Short <i>Note 1</i>	Appropriate 1	Long <i>Note 2</i>	Very Long 0

*Note 1.* Cue = 1 If Scope is Insignificant or Incremental, Cue = 0 otherwise.

*Note 2.* Cue = 1 If scope is Radical or Unrealistic, Cue = 0 otherwise.

A small degree of parameter correlation was introduced to improve the validity of the dichotomization: The cue values for cost and schedule have a minor dependence upon the scope of the task resulting in a cost-schedule correlation of  $\rho = 0.028$  and a cost-scope and schedule-scope correlation of  $\rho = -0.068$ . This dependency reflects the way that the perceived assessment of scope would be expected to affect a perceived assessment of cost or schedule. A rating of low cost might be deemed acceptable for an incremental or insignificant task, but not for a radical or unrealistic task. Likewise, a rating of high cost might be acceptable for a radical or unrealistic program, but not for an insignificant or incremental task.

*Fuzzy Model*

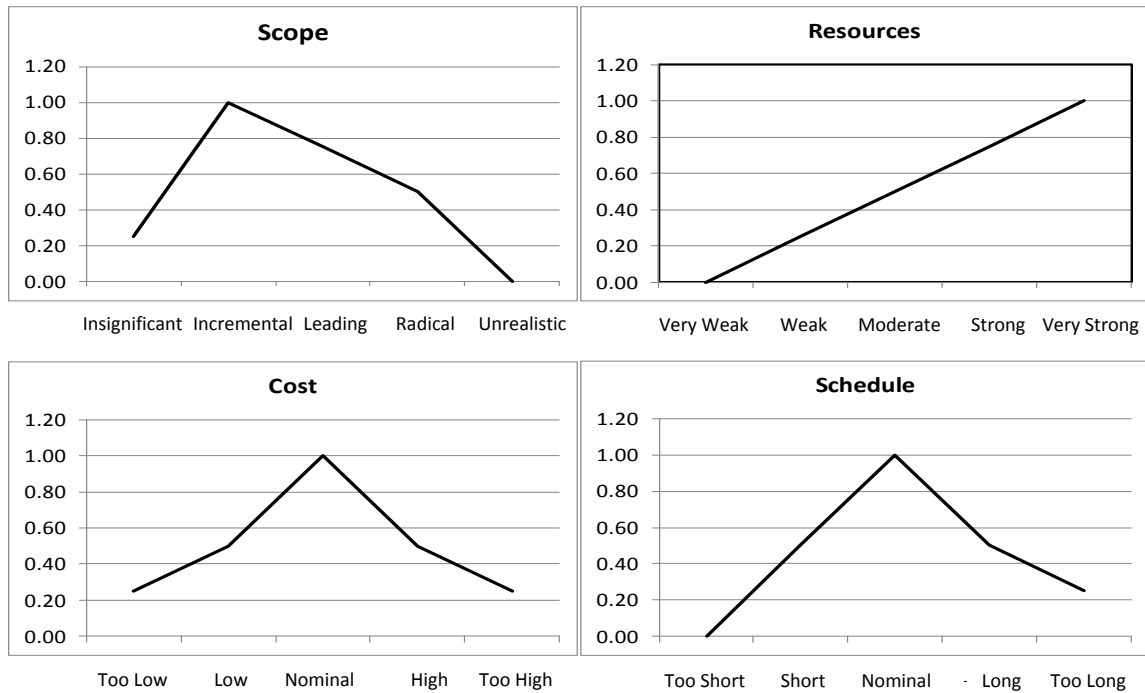
The fuzzy model was initially established as an arbitrary reference for testing the various algorithms but was subsequently incorporated into the analysis. In the fuzzy model, a reasonableness score is calculated as the sum of the membership values

corresponding to four membership functions. Each membership function defines the fuzzy rank assigned to each cue:

$$Score = f_{Scope}(S) + f_{Resource}(R) + f_{Cost}(C) + f_{Sched}(T)$$

Performance of the fuzzy model was characterized by determining the mean-square-deviation between the predictions of the model and the collected data.

The initial membership functions, illustrated in Figure 6, were developed by this writer and represent the perceived relationship of each cue to an estimate of reasonableness. Although reviewed by others with experience in proposal assessment, a more formal Delphi approach for defining the membership functions should have been used. Another approach, not included in this study, but worthy of future consideration, is the genetic algorithm method for optimization of fuzzy membership functions described by Mateou and Andreou (2008).



*Figure 6.* Fuzzy set membership functions describing the relationship between cues and reasonableness.

### *Metrics and Mapping Model*

The metrics and mapping model was implemented using the cue definitions detailed in Table 4. For any given scenario, the positive cues are counted and the scenario is placed in a category corresponding to the total number of positive cues; there are five possible categories corresponding to  $S = (0, 1, 2, 3, 4)$ . The estimated value for a scenario is then equal to the average of value of the reference scenarios within the category. This method requires at least one reference scenario per category to be effective.

The metrics and mapping model must be trained since it uses known values as references for predicting unknown values. Training used a within sample cross validation method in which 250 records were randomly selected for use as the reference scenarios and the performance of the model was validated against 250 additional randomly selected

records. Following von Helversen and Rieskamp's methodology (2008, p. 77), this process was repeated 1000 times and the mean-square-deviation for each model was averaged across all trials.

#### *QuickEst Heuristic Model*

Von Helversen and Rieskamp (2008) used the QuickEst heuristic model proposed by Hertwig et al. (1999) as one of their reference cognitive models. The QuickEst heuristic assumes that "people process cues sequentially and stop searching as soon as a cue has a negative cue value" (von Helversen & Rieskamp, 2008, p. 76). The model assumes that cues are positively correlated with the associated criteria.

For each cue, the average criterion value is calculated across all objects that have a corresponding negative cue value, irrespective of the value of other cues; this value is called the *nil mean size* (von Helversen & Rieskamp, 2008, p. 76). In a similar fashion, the average value of the criterion is calculated across all objects with a positive cue value; this value is called the *conditional positive mean* (p. 76). The cues, not the objects, are then ordered by their nil mean size.

An estimate of an object is made by examining the cues in the order of increasing nil mean size. If a cue has a positive value, then the next cue is examined. If the cue has a negative value, then the nil mean size for that cue is used as the estimate. If all of the cues are positive, then the conditional positive mean of the last cue is used as the estimate (von Helversen & Rieskamp, 2008, p. 76). Estimates are generally rounded up to the next spontaneous number, which is a multiple of a power of 10; however such rounding is not appropriate for this application which has a relatively small dynamic range.

### *Linear Regression Model*

The linear regression model calculates optimal coefficients for each cue such that the sum of squared deviations between the data and the predicted value is minimized. The estimation  $J$  given  $K$  cues of value  $X$  takes the linear additive form:

$$\hat{J} = \sum_{k=1}^K \beta_k X_k + \beta_0$$

Determination of the coefficients will be accomplished using the standard R linear model  $lm(\dots)$ . Performance of the model is characterized by the mean-square-deviation between the data collected and the predicted values of the model.

### *Exemplar Model*

Von Helversen and Rieskamp (2008) used an exemplar based model in which estimates are based upon the degree of similarity between an object in question and the reference objects maintained in memory (p. 75). An estimate is the average of the exemplar criterions weighted by the degree of similarity between the object in question and each of the exemplars in the form:

$$\hat{J}_p = \frac{\sum_{i=1}^I S(p, i) j_i}{\sum_{i=1}^I S(p, i)}$$

Where  $\hat{J}_p$  is the estimate for object  $p$ , and  $S$  is the similarity between object  $p$  and exemplar  $i$  with a criteria value of  $j_i$ . Similarity is computed as a product of the cue values and attention weight parameters across each of the cues:

$$S(p, i) = \prod_{k=1}^K d_k$$

Where  $d_k = 1$  if the object cue matches the exemplar and  $s_j$  if they do not. The value  $s_j$  is an attention weight parameter assigned to each cue and corresponds to the relative importance of any given cue. Von Helversen and Rieskamp elected to keep all the attention weights uniform to simplify the model implementation (2008, p. 76) . A large value of  $s_j$  implies a cue is not particularly significant and its presence or absence does not alter the model estimation of a particular value.

To accommodate the polychotomous cues used in this model, the exemplar similarity function was adapted to compute the similarity between the scenario being estimated and each of the exemplars with a graduated attention function. When an object cue matches the exemplar precisely, the cue is weighted at  $d_k = 1.0$ , as the cues become more dissimilar, as in having a greater distance between the object and exemplar ratings, the attention weighting factor is reduced accordingly. The attention weighting function used for this study is listed in Table 6.

Table 6

*Attention weighting values versus degree of separation.*

	Distance Weight factors (attention factors)				
Distance	0.00	1.00	2.00	3.00	4.00
Weight	1.00	0.50	0.25	0.10	0.01

The Exemplar model required a large number of reference scenarios from which it can calculate the estimated value of a scenario in question. A limitation of the exemplar model is that it cannot extrapolate beyond the high and low end criteria; hence the reference scenarios should include an upper and lower bound if available.



The performance of the exemplar model was characterized in the same fashion as the metrics and mapping model. A within sample cross validation was used in which a randomly selected subset of the data was used as exemplars, and the performance was validated against a randomly selected test set.

### *Neural Network Model*

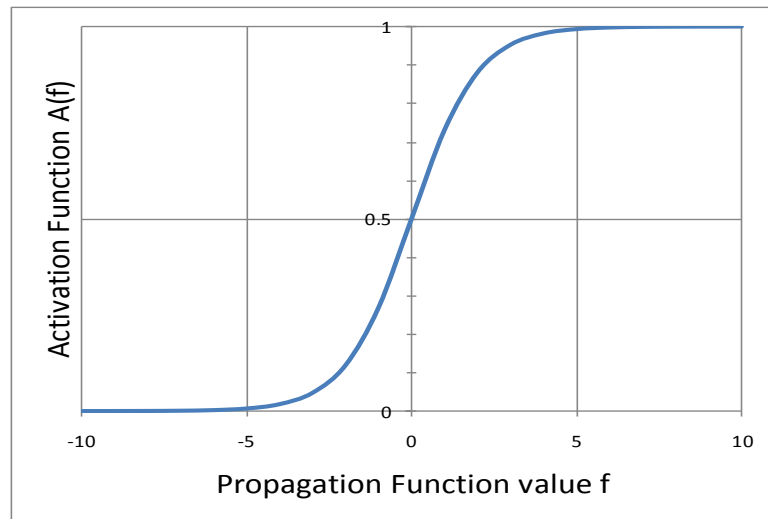
Neural networks are mathematical models that have been inspired by the function of the human neuron (Von Altrock, 1997, p. 144). Although it isn't possible to create intelligence, neural networks are very good at pattern recognition and complex problems (Nissen, 2003, p. 2). The most common neural element is an artificial neuron that consists of a *propagation function* and an *activation function*. The propagation function computes the weighted sum of the various inputs  $X_i$  and then adds an offset bias  $\Theta$ , known as the background activation level, in the form:

$$f = \theta + \sum_{i=1}^N w_i X_i$$

The activation function is typically an S-curved sigmoid or logistic function illustrated in Figure 7 and mathematically defined as:

$$A(f) = \frac{1}{1 + e^{-2s(f+t)}}$$

where  $s$  is a steepness factor and  $t$  adjusts the center of the function. Other common activation functions include the threshold step function and the hyperbolic tangent function (Nissen, 2003, p. 2).



*Figure 7.* The logistic S-curve activation function used in many neural models.

A neural network model is formed when multiple artificial neurons are linked together. The most common implementation of the neural network model is the multilayer feed-forward design that is arranged in an input layer, one or more hidden layers, and an output layer (Nissen, 2003, p. 5). The configuration for this study is illustrated in Figure 8; the optimal number of internal nodes and layers is typically determined by trial and error during the training phase (Saha, 2002, p. 12).

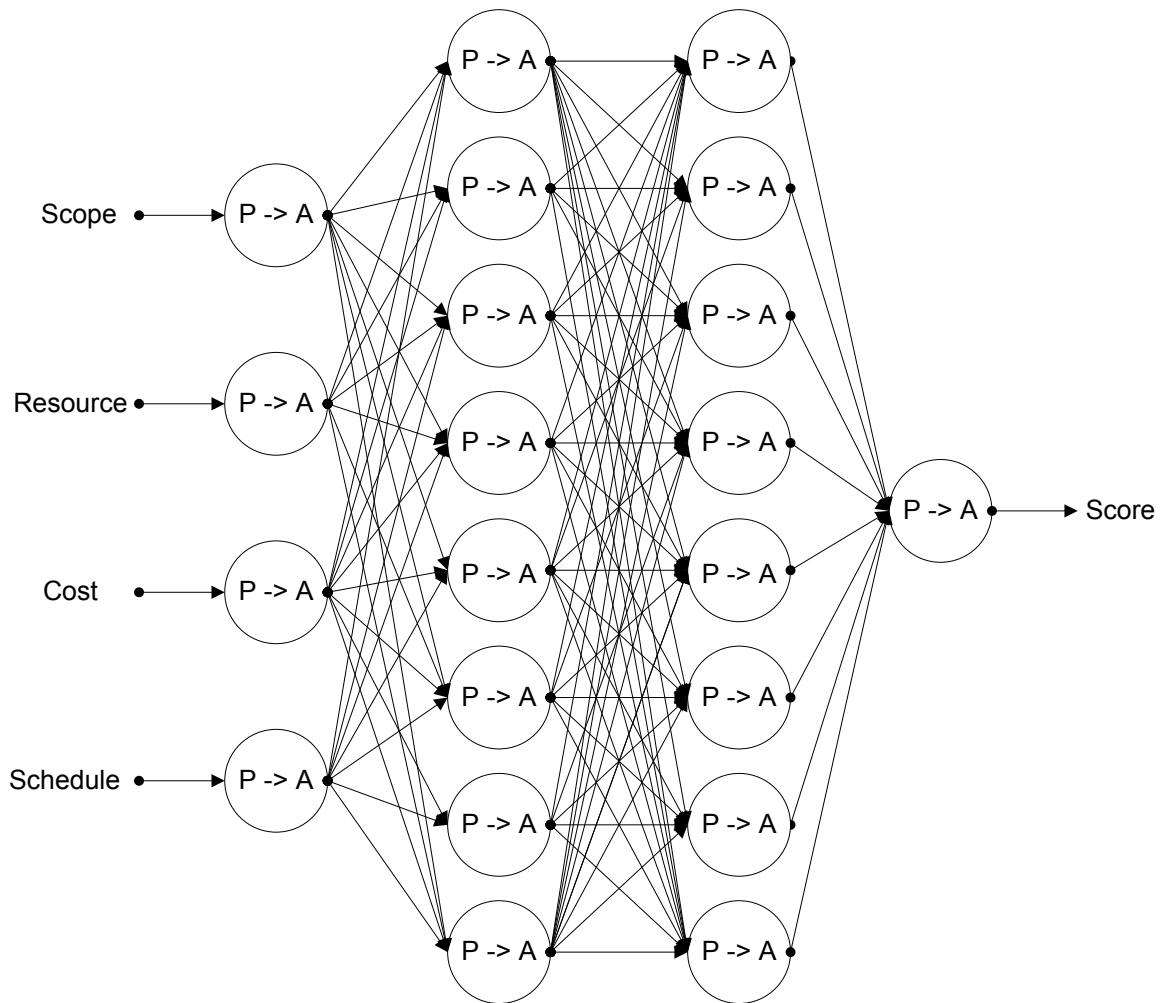


Figure 8. Neural network model used for estimating reasonableness.

The propagation function weights and offsets are determined by training the model. The training process involves iterative optimization of the propagation weights to reduce the mean-square-error of the training set against known data; the most common optimization method being the back-propagation algorithm (Nissen, 2003, p. 7).

For this study, the R package *AMORE*, a neural network model developed by Limas, Mere, Gonzalez, Ascacibar, Espinoza, and Elias (2007) will be used. The neural model was trained and characterized by using a cross validation method that uses selected

random samples within the data and validated against a randomly selected test set. Performance was determined by calculating the mean-square-deviation between the predicted values and the collected data.

### Summary

The policy-capturing approach to understanding how people estimate and integrate information as a component of their decision making process is an established and popular method (Karren & Barringer, 2002, p. 338). Despite concerns about the validity of the method, it remains an effective and practical means of gaining insight into the decision process (p. 338). To improve the study's validity, the policy-capturing exercise was designed consistent with the recommendations of Aiman-Smith et al. (2002) and Karren and Barringer (2002), both providing extensive guidance on the nuances of the technique.

Implementation of the policy-capturing exercise required the approval of the dissertation committee, the Walden University Research Review, and the Walden Institutional Review Board for Ethical Compliance. Approval to conduct the study was granted on 20 August, 2009; the Institution Review Board approval number was 08-20-09-0336690.

## CHAPTER 4:

### RESULTS

#### Introduction

This chapter is organized to present the data from the policy-capturing exercise to include the general descriptive statistics and graphic visualization of the data set. The demographic information collected from the sample is also presented, although the limited response prevented a significant analysis. The results of the cognitive modeling analysis follow and the chapter concludes with a summary of the results and post hoc power analysis.

The required number of participants needed based upon the preliminary power analysis was obtained and the quality of the data was acceptable; only one response was discarded. The results were generally consistent with this writer's expectations and support findings presented by von Helversen and Rieskamp (2008) in their mapping model study, upon which this effort was based.

#### Analysis of the Policy-capturing Exercise

##### *Research Question 1*

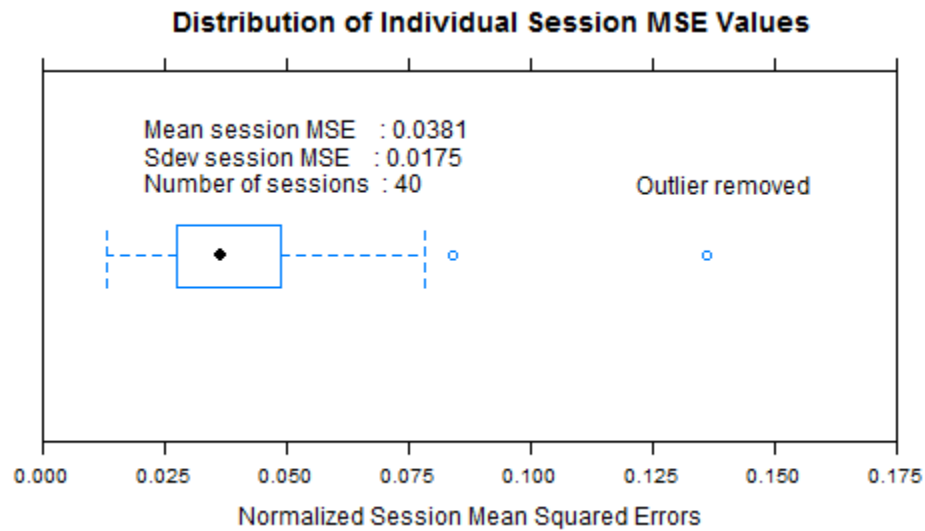
The first research question posed in this study sought to understand the observed relationship between an individual's estimation of reasonableness and the independent predictors comprising scope, resources, cost, and schedule. A policy-capturing technique using a survey methodology was applied to this question. The policy-capturing exercise was open to both the Walden University participant pool and personal invitees for a period of 2 months. A total of 61 people began or attempted the exercise, 41 people

completed the exercise, and 1 response was removed from the set as an extreme outlier. The first nine responses to each 50-question survey were eliminated, yielding a total 1640 data points; the initial questions are intended to familiarize the participant with the exercise and therefore not used in the analysis.

#### *Outlier Case*

In the outlier case, the respondent rated nearly all scenarios at the extremes of the scale (i.e., very reasonable or very unreasonable). The case did not appear to indicate a deliberate attempt to skew the data, but rather a possible misunderstanding of the exercise objectives. The session mean squared error for the outlier respondent as compared to the overall data set was more than twice that of the highest mean squared error of any other respondent.

Session mean squared error was used as a general measure of data quality during the exercise; the value was determined from the mean of the squared residuals for each respondent against a simple 2<sup>nd</sup> degree polynomial regression that was performed across the entire data set. Figure 9 illustrates via a box plot the distribution of the individual session mean squared error values with the outlier shown.



*Figure 9.* The distribution of the individual session Mean Standard Error values was used as an indicator of the quality of the session responses.

#### *Collinearity and Correlation Analysis*

The data set was also examined for collinearity and correlation between any of the cues. Using the R function *vif* in package *car* (Fox, 2009), the Variance Inflation Factor (VIF) was determined for each of the cues and the resulting values are shown in Table 7. The relatively low set of variance inflation factor values indicates very little collinearity among the cues as a value greater than 10 is a common rule of thumb for evidence of multicollinearity (Cohen, Cohen, West, & Aiken, 2003, p. 423).

Table 7

*Variance Inflation Factors for the Scenario Cues*

	<i>Variance Inflation Factor</i>
Scope	1.000351
Resource	1.000326
Cost	1.000465
Schedule	1.000306

In a similar fashion, the R function *cor* in package *stats* was used to calculate the correlation between cues. The values are presented in Table 8; consistent with the VIF calculation, there is no apparent correlation between any of the cues in the data set. The magnitude of the largest cue/cue correlation is 0.016.

Table 8

*Correlation Between Cues Using the Spearman Rho Statistic.*

	Scope	Resource	Cost	Schedule
Scope	1.000			
Resource	0.007	1.000		
Cost	0.013	-0.016	1.000	
Schedule	-0.012	-0.009	-0.010	1.000
Response	0.161	0.405	-0.111	0.050

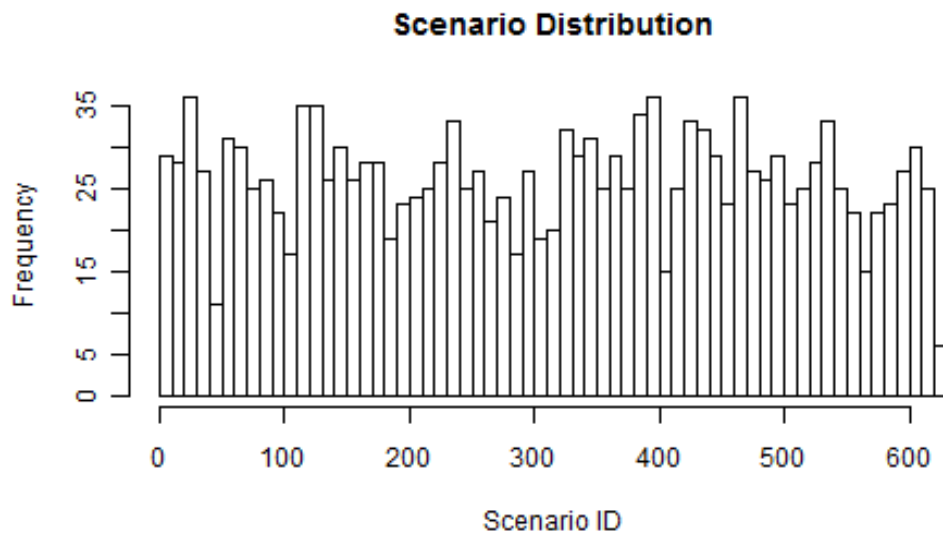
*Scenario Analysis*

The combination of test scenarios for each session was randomly generated at the time the policy-capturing exercise was administered. The resulting overall distribution of scenario values presented in the sample set is illustrated in Figure 10 and the cumulative



sum of the distribution is plotted against a straight line in Figure 11. Distributions for each of the cue parameters individually are illustrated in Figure 12.

To validate the uniformity of the algorithm used to generate the scenarios; a larger sample of 10,000 scenarios was created using the same python random sequence generator used in the web site code. A chi-square test of the resulting sequence distribution yielded a  $X^2$  value of 47.25 on 62 degrees of freedom for a p-value of .902; the p-value indicates a 90% likelihood the distribution would be uniform.



*Figure 10.* Frequency Distribution of Scenario IDs in the data sample.

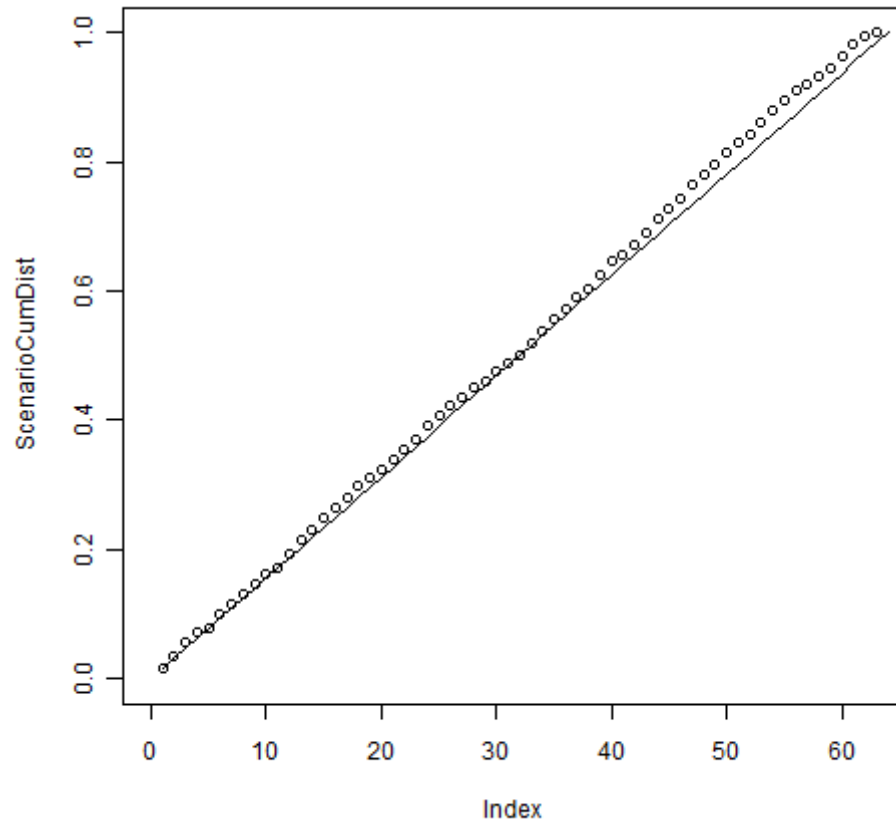


Figure 11. Normalized cumulative distribution of scenario IDs in the sample.

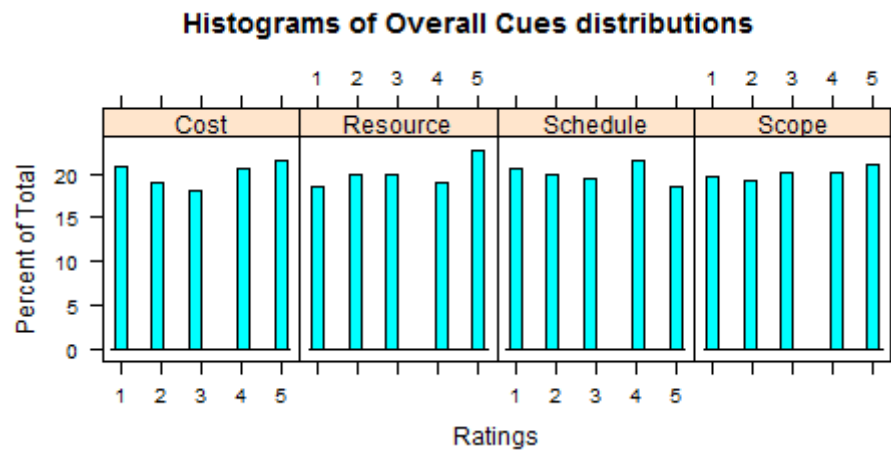


Figure 12. Distribution of individual cue values within the sample.

### *Response Characterization*

Visualizing multidimensional data presents a unique challenge. Functions with two and three variables are easily plotted as lines and surfaces; however, as the number of dimensions increases visualization becomes more complex. A proven technique for visualizing multidimensional data is the parallel graph that can “transform multivariate relations into 2D *patterns*” (Inselberg, 2005, p. 158). The technique is “well suited for visual data exploration and analysis” (p. 158) as well as applications in decision support systems (p. 165).

Figures 13 through 19 illustrate a series of parallel graphs for each of the seven possible reasonableness score responses. The parallel graphs have four vertical axes corresponding to the cue values presented to the participant; the values are listed in Table 9. Each combination of cue values that evoked a given reasonableness score is plotted as a line connecting the corresponding points on the vertical axes.

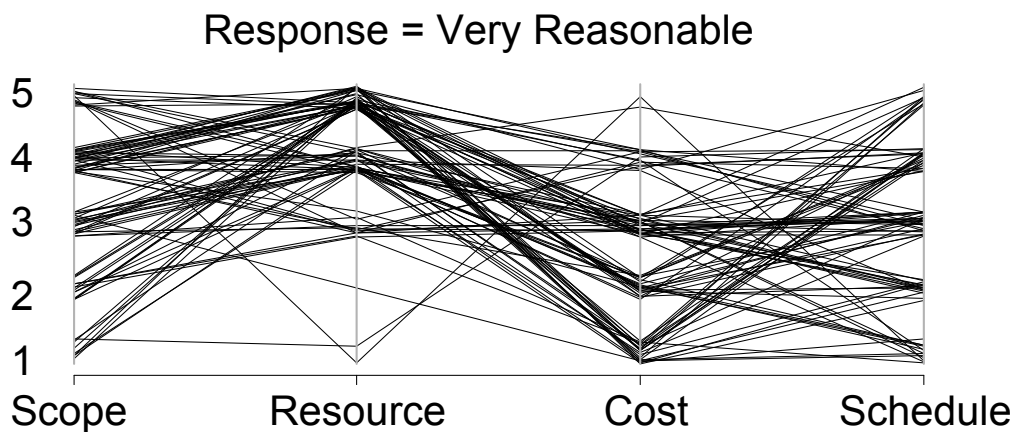
Table 9

#### *Parallel Graph Vertical Scale Values.*

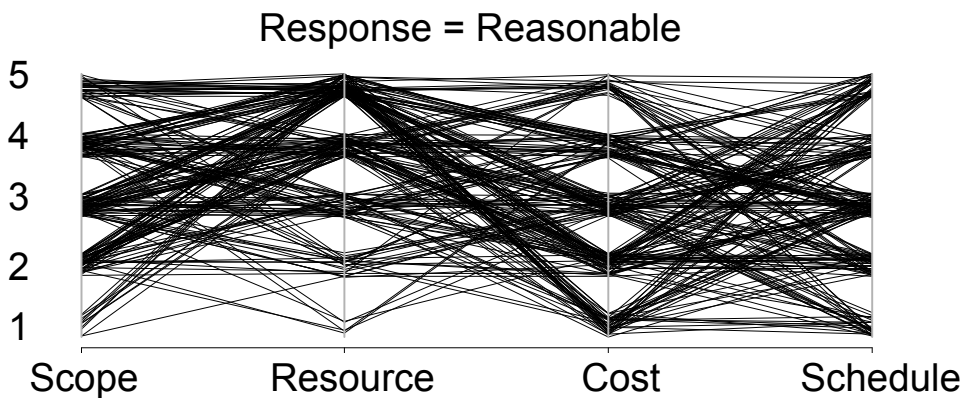
Cue	1	2	3	4	5
Scope	Insignificant	Incremental	Leading	Radical	Extreme
Resources	Very Weak	Weak	Moderate	Strong	Very Strong
Cost	Very Low	Low	Appropriate	High	Very High
Schedule	Very Short	Short	Appropriate	Long	Very Long

A small amount of random variation known as jitter was added to the cue values for each data record to facilitate visualization by randomly dithering the integer cue values by  $\pm 20\%$ ; the displayed lines would otherwise sit on top of each other and it would not be possible to get a sense of the weight and distribution of the cues. For

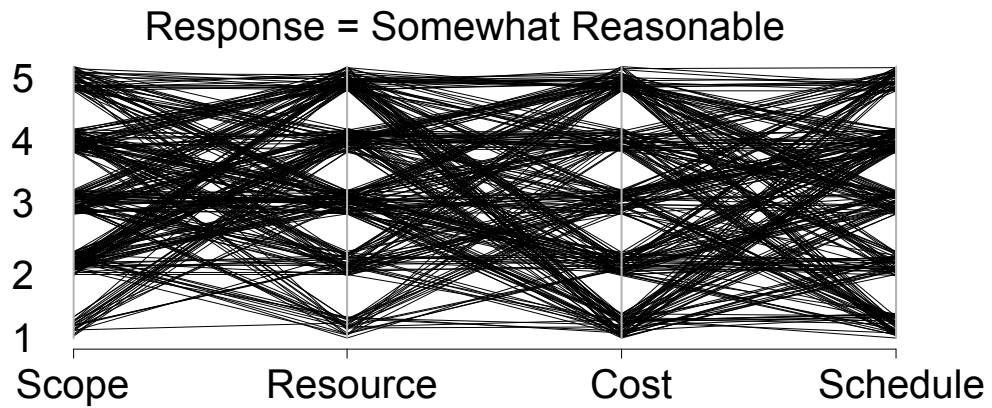
example, a data record with cues values of 3, 4, 2, and 2, corresponding to scope, resource, cost, and schedule might be plotted using values of 2.98, 4.03, 2.11, and 1.92. As might be expected, the data records associated with the central values of the reasonableness scores 3, 4, and 5 are widely distributed and do not reveal any specific patterns. The extreme values of very reasonable and very unreasonable do, however, indicate distinct relationships with the cue values.



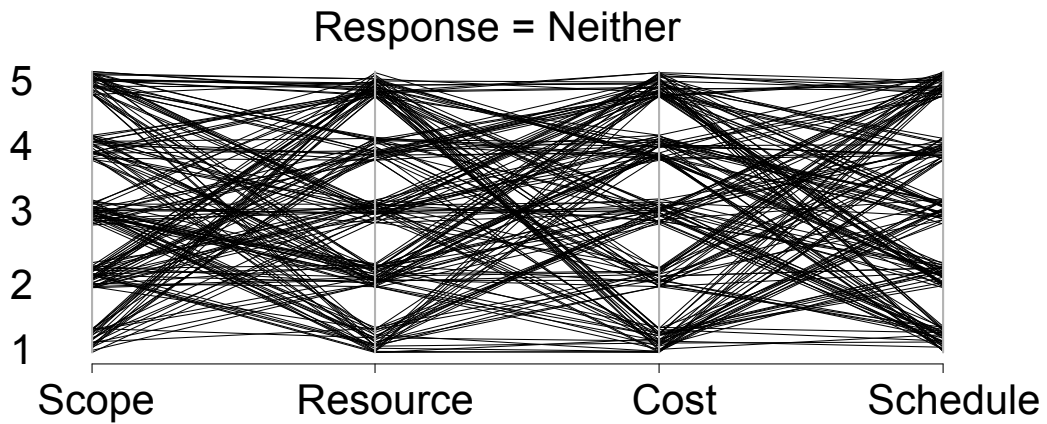
*Figure 13.* A parallel line plot for a reasonableness score of 7.



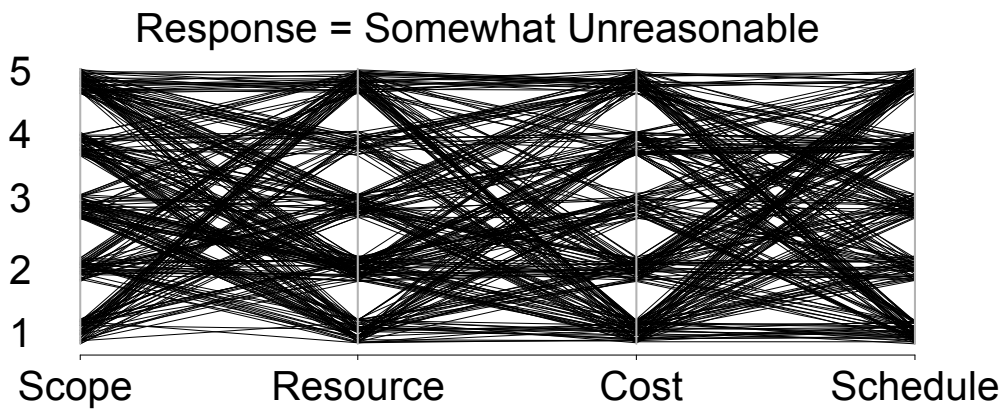
*Figure 14.* A parallel line plot for a reasonableness score of 6.



*Figure 15.* A parallel line plot for a reasonableness score of 5.



*Figure 16.* A parallel line plot for a reasonableness score of 4.



*Figure 17.* A parallel line plot for a reasonableness score of 3.

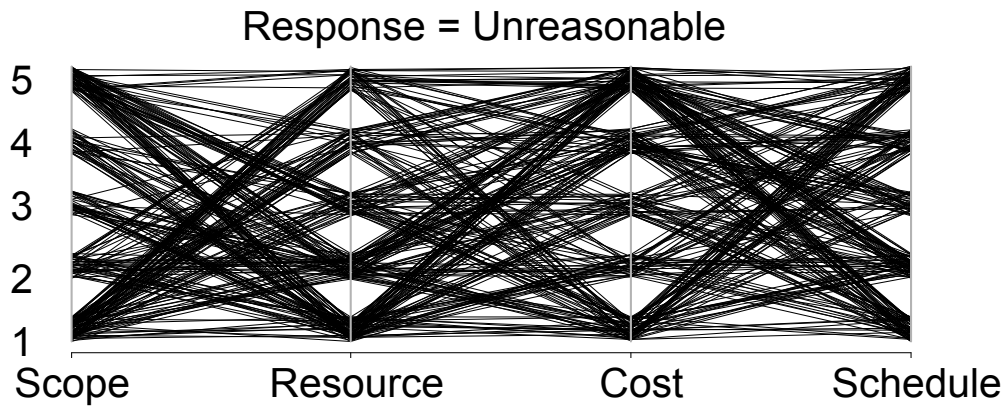


Figure 18. A parallel line plot for a reasonableness score of 2.

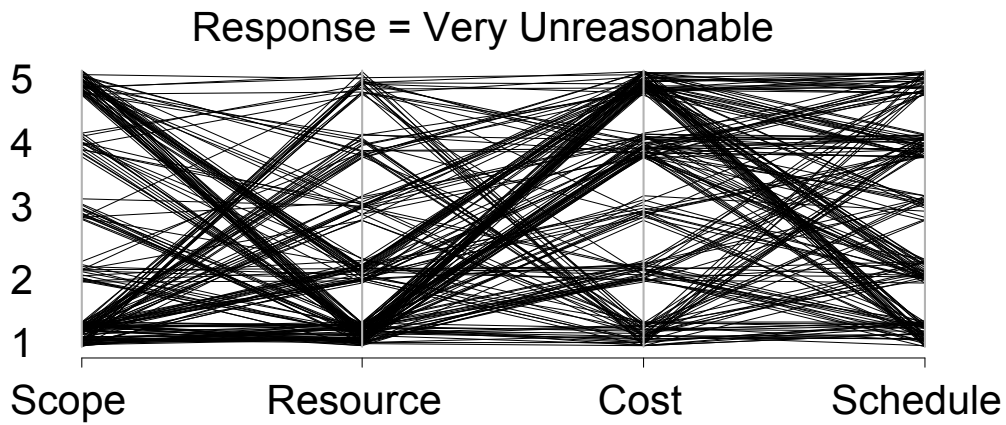


Figure 19. A parallel line plot for a reasonableness score of 1.

A second method of visualization is presented in Figure 20. For each of the four cues, a histogram of the cue settings that evoked the given reasonableness response is presented. The percentage of total on the vertical axes is relative to the number of records tallied for the specific reasonableness score. As a case in point, one can see that for those scenarios judged by the participants as being *very reasonable* (score = 7), nearly 50% had been described as having *very strong* resources.

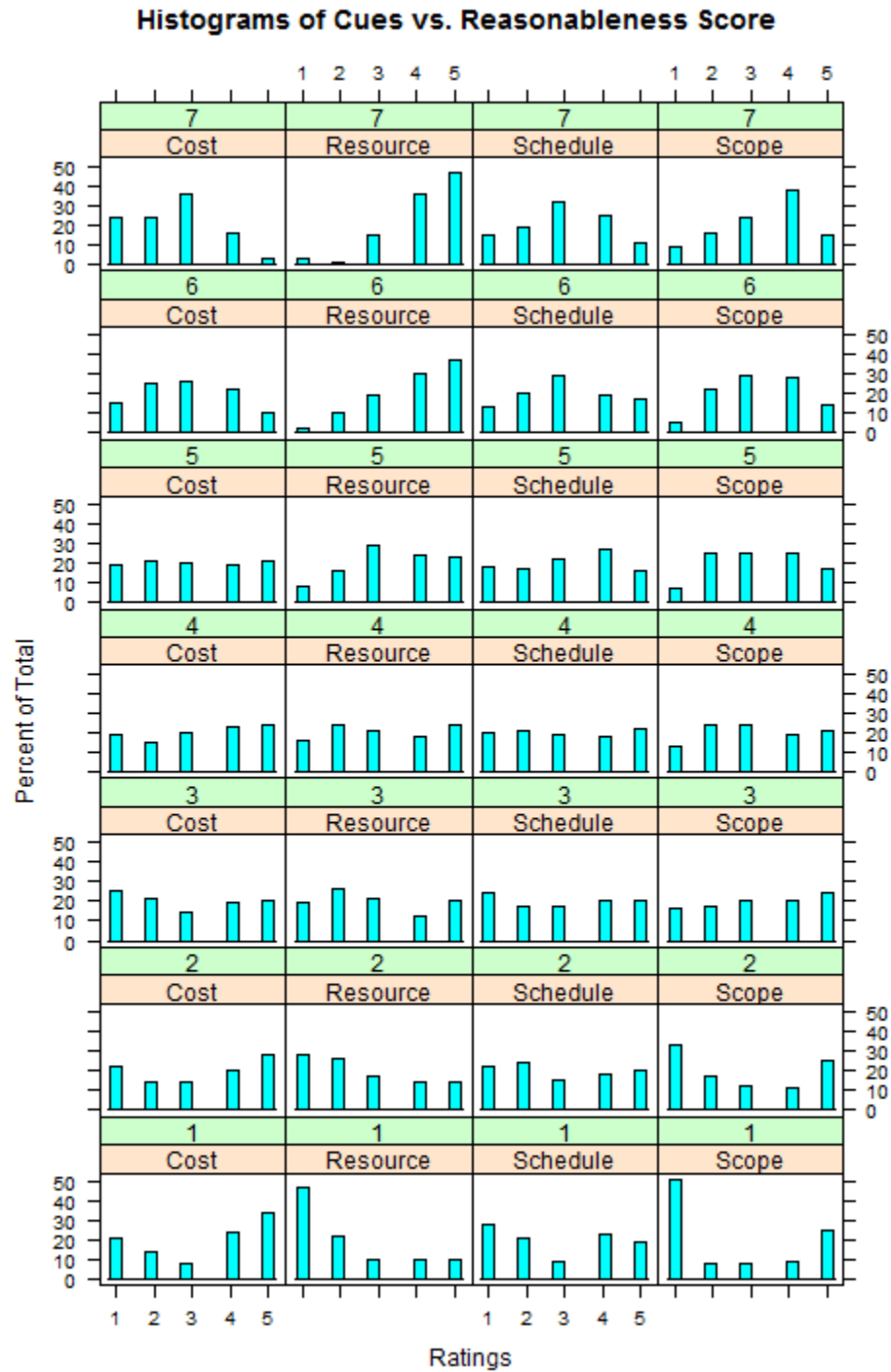


Figure 20. Distributions of cue values corresponding to specific ratings of Reasonableness (1 = Very Unreasonable, 7 = Very Reasonable).

The histogram format does not directly reveal information about the interrelationship between the cue values and their corresponding reasonableness score; however, it is possible to gain a sense of how each cue relates individually to reasonableness. For example, one might conclude that reviewers are more forgiving of an underpriced proposal (low cost) than they are of a short schedule when considering a strong proposal.

#### *Test Retest Analysis*

To determine the test retest validity of the data set, the policy-capturing exercise intentionally repeated four scenarios at different intervals through the survey. The scenario sequence for each session was randomly generated without substitution to create a sequence of 50 unique cue combinations. The scenarios at positions 10, 11, 12, and 13 were then repeated at positions 20, 30, 40, and 50 respectively. The correlation between the initial response and subsequent response is an indication of test retest validity; a value of 0.70 or greater is considered reliable (Domino & Domino, 2002, p. 43) The results of the test retest correlation analysis for this study are shown in Table 10 with an overall test retest correlation of 0.80 achieved.

Table 10

*Test Retest Correlation for the Entire Sample and for Each Interval in the Exercise.*

	<i>Test / Retest Correlation</i>
All Samples	0.803401
Scenario 10 to 20	0.782696
Scenario 11 to 30	0.717671
Scenario 12 to 40	0.910030
Scenario 13 to 50	0.780160



*Description of the Sample*

Demographic information was requested from each respondent to include age group, total years of general management experience, total years of research experience, educational level, and occupation. A summary of the data collected from the demographic survey questions is provided in Table 11.

Table 11

*Demographic Summary of the Sample*

	<i>Count</i>	<i>Percent of Sample</i>
<b>Age Group</b>		
< 25 yrs	0	0.0%
25 – 45 yrs	17	42.5%
> 45 yrs	23	57.5%
<b>Project Planning Experience</b>		
< 5 yrs	6	15.0%
5 – 15 yrs	14	35.0%
> 15 yrs	20	50.0%
<b>Research Planning Experience</b>		
< 3 yrs	12	30.0%
3 – 10 yrs	14	35.0%
> 10 yrs	14	35.0%
<b>Education Level</b>		
No response	1	2.5%
Bachelors	10	25.0%
Masters	18	45.0%
Ph.D.	11	27.5%
<b>Professional Background</b>		
Engineering	9	22.5%
Academic	8	20.0%
Manager	14	35.0%
Software	3	7.5%
Other	6	15.0%
<b>Source</b>		
Walden Participant Pool	11	27.5%
Invitation	39	72.5%

### *Analysis of Subsamples*

Although the sample size was limited, some insight can be gained by examining the session mean squared error calculation that was used to initially assess the survey responses. Values for the principal subpopulations are shown in Table 12. Considering the two extreme mean values of mean squared error corresponding to Engineers and Faculty Members respectively, it is not possible to reject the null hypothesis  $H_0 : \mu_1 - \mu_2 = 0$ , hence there was no significant difference detected in the two subpopulations. A  $t$ -test to check for a significant difference in the means at  $\alpha = 0.05$  yields a p-value of 0.396 for the null hypothesis and a 60% likelihood the means are equal.

Table 12

#### *Descriptive Statistics for the Session Mean Squared Error of the Primary Subpopulations.*

	Min	1st Qtr.	Median	Mean	3rd Qtr	Max	Std Dev	Count
Engineers	0.027	0.032	0.036	0.042	0.039	0.087	0.019	9
Faculty	0.013	0.030	0.038	0.035	0.043	0.051	0.013	8
Manager	0.012	0.031	0.032	0.039	0.054	0.076	0.019	14
Invite	0.015	0.028	0.032	0.038	0.043	0.081	0.017	29
Pool	0.014	0.022	0.035	0.037	0.046	0.086	0.021	10
Full	0.013	0.027	0.036	0.038	0.047	0.084	0.018	40

### Analysis of Model Performance

#### *Research Questions 2 and 3*

The second and third research questions both addressed model performance. The second question asked to what degree various cognitive models of decision making, reasoning, and judgment will predict the estimates of reasonableness by a peer review community and the third question asked which cognitive model most accurately reflects a

peer reviewer's estimation of reasonableness. To answer these questions, six cognitive models were tested against the data collected in the policy-capturing exercise. Details on the training and testing results follow.

### *Training Parameters*

Four of the six models evaluated in this study required training; a process whereby randomly sampled data records are used to set the coefficients for the predictor function. The number of training records used in the simulation trials was based upon the metrics and mapping model which has the most stringent requirements; the metrics and mapping model requires at least one representative sample for each of the possible dichotomized cue totals.

The number of training records was increased in increments of 50 until the probability that all five cue totals were represented in the sample exceeded 99.99%. At 250 training records, the likelihood that all five cues totals were represented is 99.998%; the value is based upon obtaining cue total of four which is the least represented of the five possible cue totals. The cue totals are listed in Table 13 along with the probability of selecting one out of 625, the probabilities of not selecting the scenario in 250 trials, and the corresponding probability of selecting at least one in 250 trials.

Table 13

*Calculation of the Probability of Representing all Five Cue-Total Values in the Sample.*

Cue Total	Count	Will find P%	Won't find 1-p	P: Won't find in 250 Trials	P: Will find in 250 Trials
	36	5.76%	94.24%	3.62092E-07	99.99996%
1	170	27.20%	72.80%	3.41071E-35	100.00000%
2	254	40.64%	59.36%	2.36305E-57	100.00000%
3	138	22.08%	77.92%	8.17026E-28	100.00000%
4	27	4.32%	95.68%	1.60432E-05	99.99840%

#### *Simulation Parameters*

The number of simulation trials is based upon a power analysis of the number of samples required to ascertain a difference between means. As previously cited, mean squared error is used as a measure of model performance; using preliminary data based upon a 100 trial run, the narrowest difference in means occurred between the Fuzzy and Mapping models. The means and standard deviations of the preliminary trial run were applied to G\*Power 3.1 to determine the sample size required to resolve a difference between the two means; G\*Power 3.1 predicted the minimum sample size to be 860 as illustrated in Figure 21, hence the number was rounded to 1000.

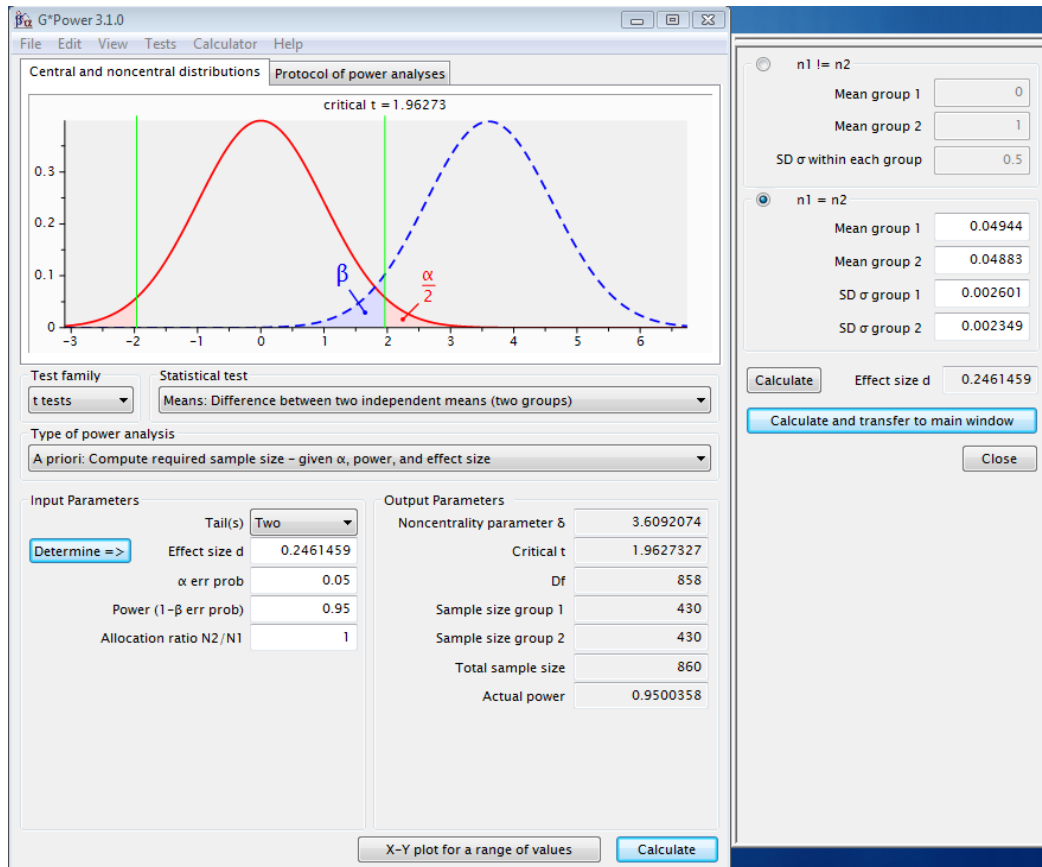


Figure 21. Power analysis to determine the number of simulation trials required to resolve a difference between means; produced with the computer program “G\*Power version 3.1.0” by Faul et al., 2007

### Modeling Results

A simulation trial comprises four steps: (a) randomize the order of the sample data set of 1640 records, (b) use records 1 through 250 to train the models, (c) use records 251 through 500 to test the models by calculating the squared error between each model’s predicted value of reasonableness and the corresponding record’s measured value of reasonableness, and then (d) for each model, calculate the mean squared error as the average of the 250 discrete squared error measurements obtained in step c.

The simulation was repeated for 1000 trials and the results for each model were accumulated. Figure 22 illustrates a box plot of the mean squared error results for each of the models and the corresponding numerical results are tabulated in Table 14. The metrics and mapping model developed by von Helversen and Rieskamp (von Helversen & Rieskamp, 2008), the Fuzzy model, and the Exemplar model all outperformed the conventional linear regression.

Table 14

*Mean Squared Error Performance of the Models over 1000 Trials.*

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Sdev
QuickEst	0.048	0.058	0.062	0.062	0.065	0.082	0.005
Neural	0.046	0.062	0.068	0.069	0.075	0.109	0.009
Mapping	0.037	0.046	0.049	0.049	0.051	0.060	0.004
Linear	0.043	0.051	0.053	0.053	0.056	0.065	0.004
Fuzzy	0.035	0.046	0.049	0.049	0.052	0.064	0.004
Exemplar	0.038	0.045	0.047	0.047	0.050	0.058	0.003

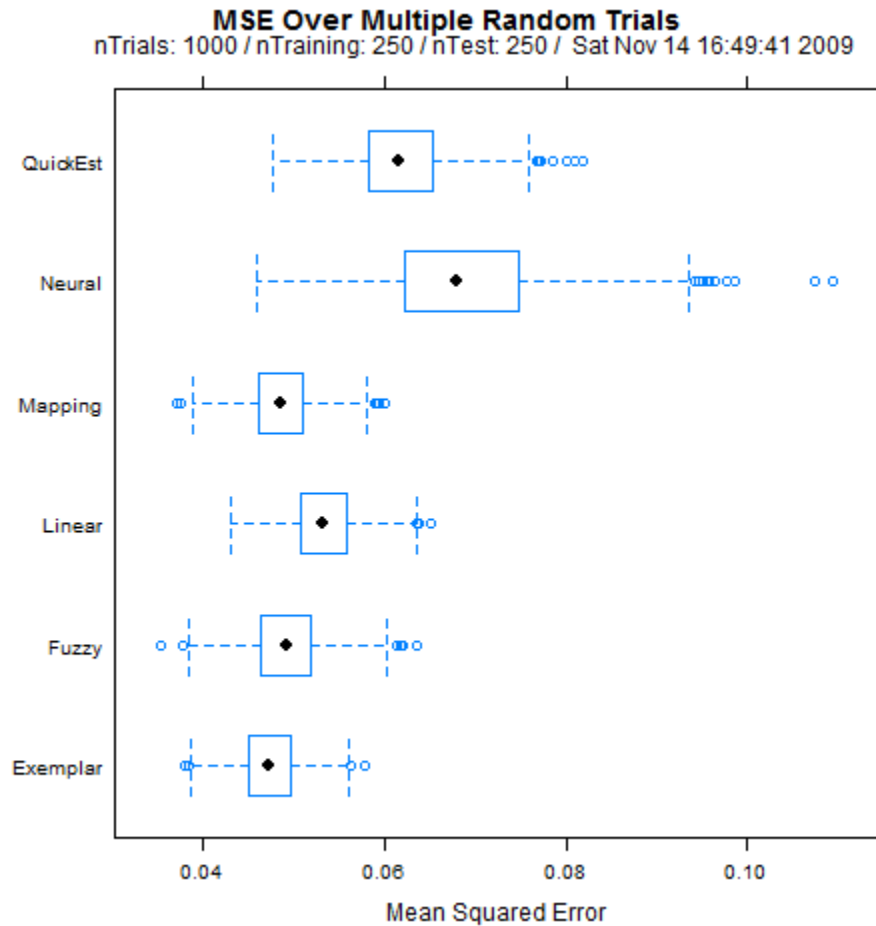
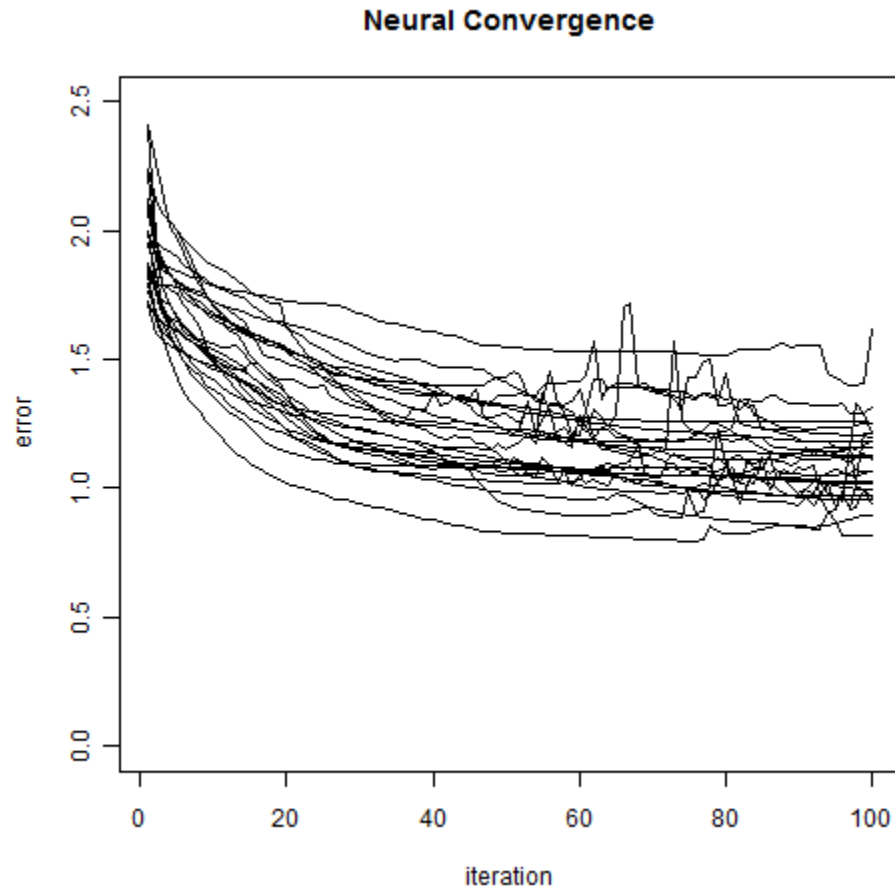


Figure 22. Box plot illustrating the mean squared error for each model across 1000 trials.

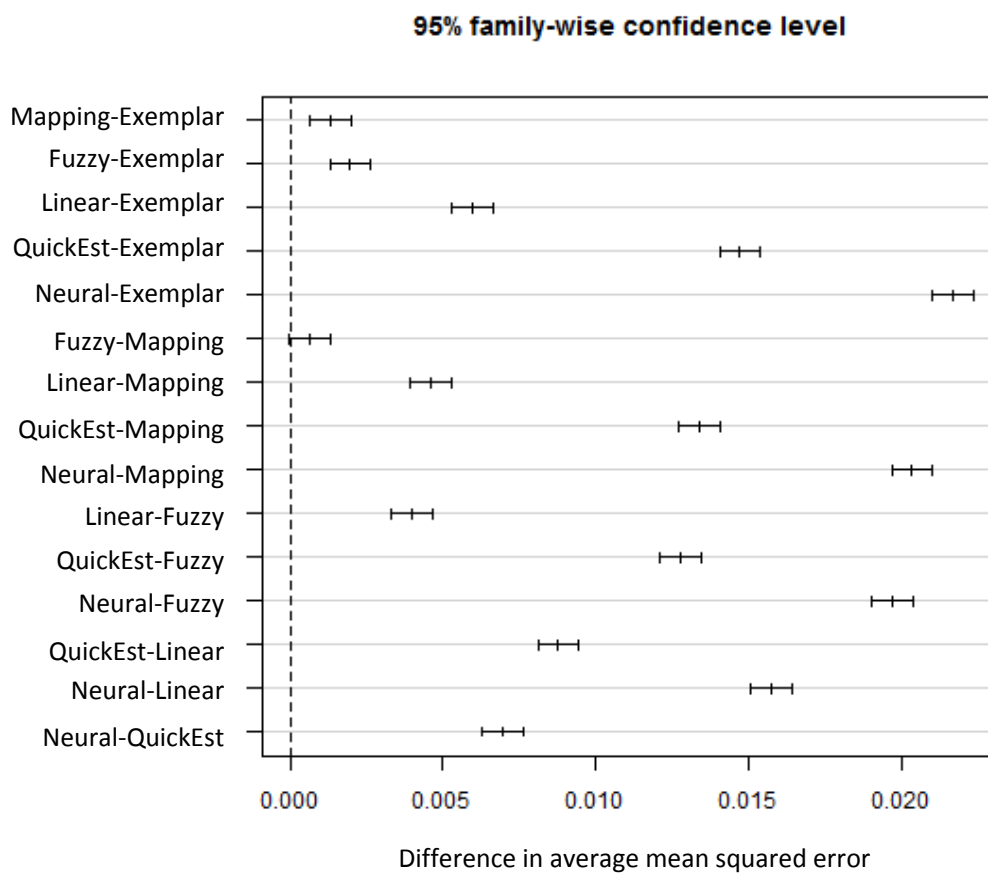
The neural model demonstrated the greatest variance and the greatest mean squared error. To determine if the neural model was given a sufficient training interval, the error convergence for the first 25 trials was captured and plotted as shown in Figure 23. Neural models unfortunately are subject to converging upon local minima, overshooting the correct answer, and over fitting the data. Saha suggests *trial and error* as an effective method for adjusting learning rate and momentum to achieve the lowest possible error (Saha, 2002).



*Figure 23.* Neural Model error convergence plot of the first 25 trials.

One can see from the previous box plot in Figure 22 that the exemplar, mapping and fuzzy models outperformed the linear, QuickEst, and neural models; to quantitatively compare the model's performance, a pairwise Tukey test for differences in the means was created using an  $\alpha$  of 0.05. The results are illustrated graphically in Figure 24 and tabulated in Table 15. No statistically significant difference in the mean squared error performance was detected between the mapping and fuzzy models.





*Figure 24.* Depiction of the pairwise differences between models means squared error.

Table 15

*Tabular Results of the Pairwise Tukey Analysis of the Difference between Means, Arranged in Order of Increasing Difference.*

	Difference	Lower	Upper	p adj
Mapping-Exemplar	0.0013	0.0007	0.0020	0.0000003
Fuzzy-Exemplar	0.0020	0.0013	0.0026	0.0000000
Linear-Exemplar	0.0059	0.0053	0.0066	0.0000000
QuickEst-Exemplar	0.0147	0.0140	0.0154	0.0000000
Neural-Exemplar	0.0217	0.0210	0.0223	0.0000000
Fuzzy-Mapping	0.0006	0.0000	0.0013	0.0773011
Linear-Mapping	0.0046	0.0039	0.0053	0.0000000
QuickEst-Mapping	0.0134	0.0127	0.0141	0.0000000
Neural-Mapping	0.0203	0.0197	0.0210	0.0000000
Linear-Fuzzy	0.0040	0.0033	0.0047	0.0000000
QuickEst-Fuzzy	0.0128	0.0121	0.0134	0.0000000
Neural-Fuzzy	0.0197	0.0190	0.0204	0.0000000
QuickEst-Linear	0.0088	0.0081	0.0094	0.0000000
Neural-Linear	0.0157	0.0151	0.0164	0.0000000
Neural-QuickEst	0.0070	0.0063	0.0076	0.0000000

As an additional measure of the goodness of fit, the residuals of the first 25 trials were analyzed to determine if the errors were normally distributed. Histograms of the residual distributions are shown in Figure 25 with the corresponding normal probability plots shown in Figure 26. Chi square tests against a normal distribution are tabulated in Table 16; each of the models produced similar p-values with respect to normalcy with an average value of 0.238.

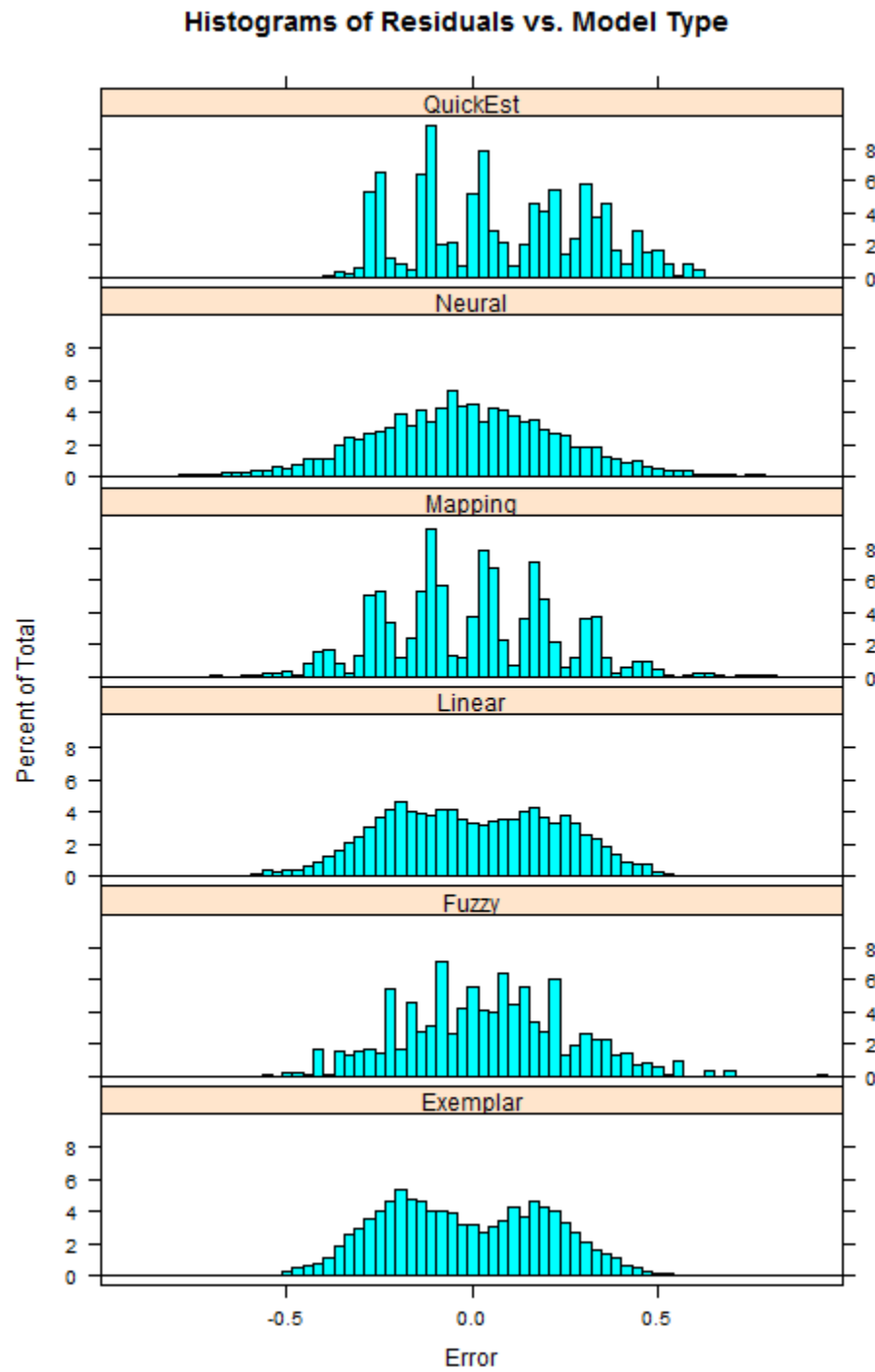


Figure 25. Histogram plot of the squared error residuals for each model across 1000 trials.

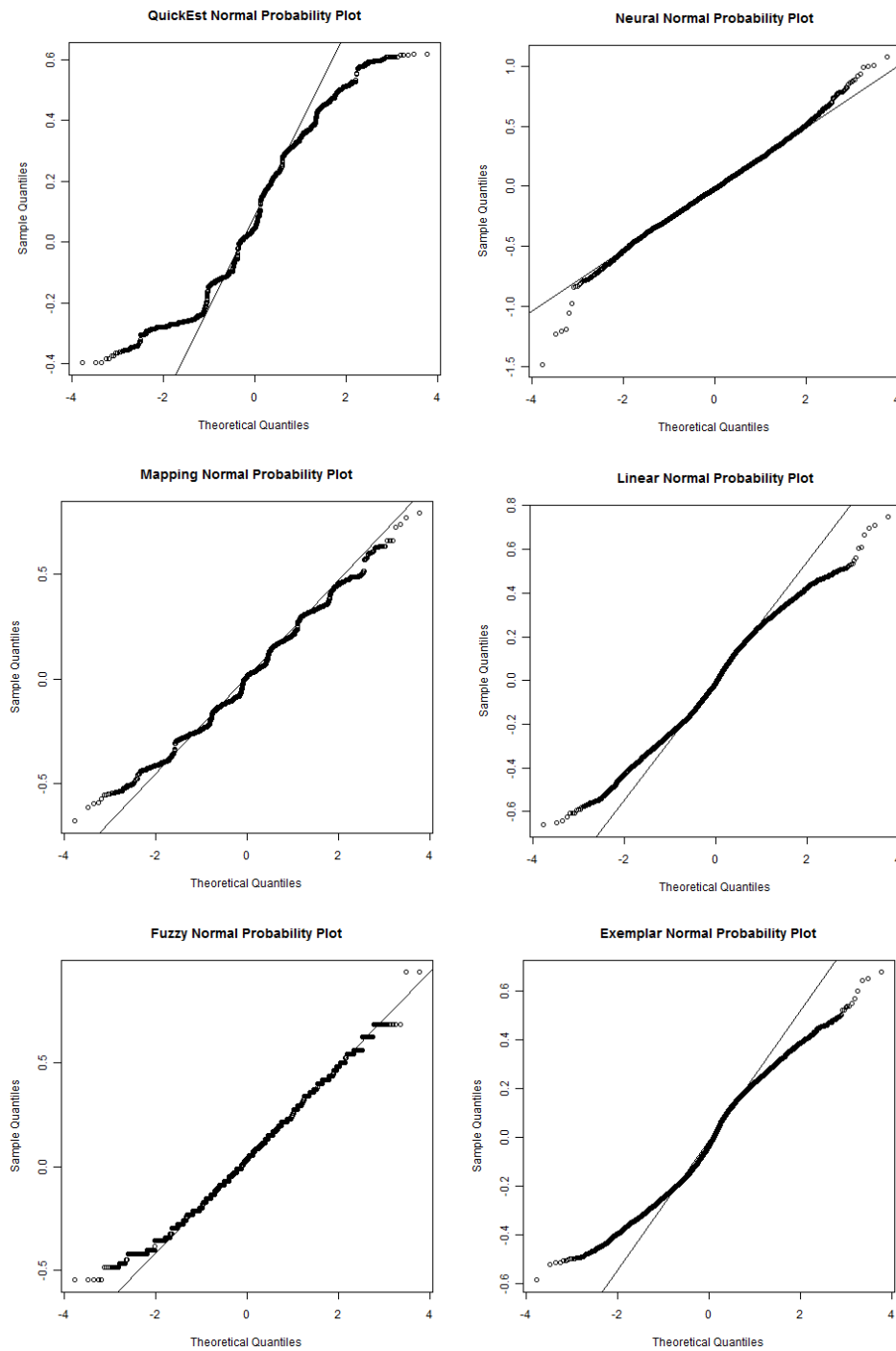


Figure 26. Normal probability plots of the squared error residuals for each model across 1000 trials.

Table 16

*Chi-Square Test of the Model Residuals against a Normal Distribution.*

	Residual Mean	Residual SDev	Chi Square	p-value
QuickEst	0.0835	0.2341	110	0.2322
Neural	-0.0176	0.2597	154	0.2503
Mapping	-0.0016	0.2201	210	0.2344
Linear	-0.0020	0.2307	210	0.2344
Fuzzy	0.0381	0.2177	224	0.2417
Exemplar	-0.0179	0.2171	156	0.2335

### *Model Programming Validation*

As an additional step to ensure that the model programming was done correctly, a sample of test records were created based upon the fuzzy model constructs. The models were then run using the R programming language that was used for the analysis in this study, and again in a separate Excel® spreadsheet. The Excel® spreadsheet provides better visibility into the underlying model calculations and it is easier to validate the intermediate steps used in the calculations. The two approaches generated identical results as expected.

### User Comments on the Exercise

Three open ended questions were presented at the conclusion of the exercise to solicit the participant's comments on estimation in general, the survey in particular, and any other aspect of the exercise that might come to mind. Responses were coded using Weft-QDA, a qualitative data analysis tool, to identify the major themes (Fenton, 2006); Table 17 presents a tabulation of the coding results for each of the closing questions.

Table 17

*Coding Counts for Survey Open Ended Questions.*

	Estimating Practices	Survey Comments	Other Comments
Value proposition needed	19	2	1
Resources Important	16	1	0
Cost Important	3	0	0
Balance Important	5	0	0
Schedule was Least Important	4	0	0
Suggestions for improvement	0	4	1
Evaluate as a portfolio	0	3	1
Exercise was reasonable	0	8	1
Exercise was too long	0	1	1
Exercise was difficult	0	2	0

The first question asked participants to comment on their most important consideration when evaluating reasonableness. Although *value* was not one of the criteria covered in the study, it was the most often cited consideration with 22 comments suggesting that knowing something about the potential return of a project was necessary in order to assess its reasonableness.

In the introduction to this study, a proposal was said to be measureable in as little as four criteria comprising *risk*, *relevance*, *reasonableness*, and *return* (De Piante Henriksen & Palocsay, 2008). It was the intention of this study to examine reasonableness as a function of its programmatic components of cost, schedule, resources, and scope; it is apparent that it was very difficult for the participants to separate that judgment from the large value proposition offered by the proposal.

The second factor most often cited was the perceived strength of the resources to perform the task; 17 comments suggested that the strength of the resources was the most important factor in their assessment. This is consistent with the work of Marsh et al.

(2008) that demonstrated the positive correlation between an author's professional standing, institutional standing, and the rating received in a peer review, all other things being equal. Marsh et al. note, however, that it remains uncertain if the correlation is a source of validity or a source of bias (p. 164).

Remaining comments included various combinations of scope, resource and cost related criteria suggesting that balance was the principal concern. It is worth noting that schedule was the only criteria to be specifically identified as not being important by four respondents. The lack of emphasis on schedule as a criterion is evident in the policy-capturing responses previously illustrated in Figure 20. Schedule had the lowest correlation with the participant response with a value of 0.050 as compared to 0.161 for scope, 0.405 for resources, and -0.111 for cost.

The second question addressed the effectiveness of the policy-capturing exercise implementation. Eight of the participants felt the exercise sufficiently portrayed the scenarios; four had suggestions for improvement; and three felt the exercise was too long or difficult. Suggestions centered on providing additional information with the cue values explaining the rationale for a given cue score.

### Summary

The purpose of this study was to gain a better understanding of how people evaluate reasonableness and how the process can best be modeled in a decision support system. The study successfully used a policy-capturing exercise to gain insight into reasonableness assessments and demonstrated the superior performance of the exemplar, fuzzy, and mapping models over conventional linear regression in predicting

reasonableness scores. The significance of these results and recommendations for further work are discussed in the next chapter.



CHAPTER 5:  
SUMMARY, CONCLUSION, AND RECOMMENDATIONS

Summary

This study had its origins in an ongoing debate over the ability to estimate the cost and schedule aspects of a research program. The subsequent investigation into the topic revealed much information spanning proposals to production to publication. Given this writer's interest in decision support systems in general and artificial neural networks in particular, the investigation ultimately led to developing a better understanding the proposal assessment process and tools that could enhance the consistency and fairness of that process.

A review of the literature documented the need for work in this area citing significant problems with the peer review and grant approval processes. Although Marsh et al. (2008) found inter-rater reliability to be as low as 0.15 in a study of more than 10,000 proposal evaluations, they could find very little empirical research addressing the problem (pp. 160-162). The literature also introduced a number of novel approaches to measuring and modeling human judgment, including policy-capturing and the metrics and mapping model proposed by von Helversen and Rieskamp (2008), that became the basis for this study.

To investigate these issues, three research questions were posed:

1. What is the observed relationship between an individual's estimation of reasonableness and the independent predictors comprising scope, resources, cost, and schedule?

2. To what degree will various cognitive models of decision making, reasoning, and judgment predict the estimates of reasonableness by a peer review community?

3. Which cognitive model most accurately reflects a peer reviewer's estimation of reasonableness?

Following the approach used by von Helversen and Rieskamp (2008) empirical data was collected in the form of a simulated judgment exercise (policy-capturing) then compared against the predictions of six different computer models. As in the von Helversen and Rieskamp study, the metrics and mapping model outperformed conventional linear regression; the exemplar and fuzzy models also outperformed linear regression whereas the neural and QuickEst models did not.

### Interpretation of Findings

#### *Policy-Capturing Exercise and Research Question 1*

Research question 1 asked; "What is the observed relationship between an individual's estimation of reasonableness and the independent predictors comprising scope, resources, cost, and schedule?" To address this question, a policy-capturing exercise was used to gather data on how people make judgments regarding reasonableness when presented with varying combinations of scope, resources, cost and schedule.

The policy-capturing exercise produced a total of 1640 estimates of reasonableness by 40 participants. Analyzing the results of the policy-capturing exercise revealed an unexpected relationship between cost and reasonableness. It was apparent

that respondents were very forgiving of insufficiently funded cost estimates when judging proposals; on the other hand, an overpriced proposal resulted in a poor score. This is not a rational behavior because an underfunded project cannot supply the resources needed for a successful completion. A more rational response would have been to forgive a higher price with the potential for producing additional work while reducing risk due to unforeseen problems or delays.

The respondent's need for knowledge of the social or economic value of the project also confounded the results, as respondents expressed a strong desire to know if the proposal they were evaluating was even worth evaluating in the first place. These findings support the notion that proposal evaluations have a strong emotional component. Decision support tools can effectively mitigate those emotional responses; hence the need for the second phase of this study which focuses on cognitive models.

The policy-capturing exercise also demonstrated the strong relationship between resources available and the reasonableness of the proposal. The ability to perform the work, regardless of other cue values, was a primary driver in determining a score. Marsh et al. identified this same relationship but commented that they could not infer if the relationship was a source of bias or validity (2008, p. 164). Because no information about the proposal authors was provided in the scenario, one may infer that name recognition bias alone is not the driving factor in the assessment.

#### *Conclusions with Respect to Research Question 1*

From the findings of the policy-capturing exercise, the following conclusions may be drawn regarding research question 1:

1. The strength of the team and availability of resources was the strongest driver in assessing the reasonableness of a proposal.
2. Irrational behaviors, such as favoring lower cost proposals over higher cost proposals are evident in the judgment process.
3. The social context of the work in terms of its benefit to industry, science, or society is a key factor.

#### *Modeling Performance and Research Questions 2 and 3*

Research question 2 asked; “To what degree will various cognitive models of decision making, reasoning, and judgment predict the estimates of reasonableness by a peer review community?” Research question 3 builds upon question 2 and more specifically asks; “Which cognitive model most accurately reflects a peer reviewer’s estimation of reasonableness?” A modeling and computer simulation was used to address these questions.

The modeling exercise produced an unexpected result in that the artificial neural network model performed poorly in relation to the other models, both in its ability to produce an accurate result and in its consistency. Neural network models, however, can be complex and difficult to optimize; although some adjustments to the learning rate and momentum parameters were made based upon recommendations from Reed and Marks (1999, ch. 6) the model evaluated in this study did not substantially deviate from its default configuration. Without further work on the neural model configuration, which was beyond the scope of this study, no significant conclusion can be drawn with respect to its performance.

Of particular interest is the performance of the mapping, QuickEst, and exemplar models relative to the mapping model study conducted by von Helversen and Rieskamp (2008); in the mapping study, the authors compared the predictive performance of the mapping, QuickEst, exemplar, and linear regression models against estimation data produced by a artificial exercise in which students were asked to estimate the toxicity of fictitious bugs. Four different studies were conducted, the second of which most closely resembles the policy-capturing exercise and simulation used in this study.

Recognizing that the scale of the von Helversen and Rieskamp data differs from that used in this study, the relative error associated with each of the models when used in a non-linear environment is presented in Table 18. The reference to *J*-shaped describes the power law response curve that was modeled in the study; old profiles are the training set and new profiles are the test / validation set. In the von Helversen and Rieskamp (2008) study, the mapping model outperformed the linear regression, QuickEst, and exemplar models when applied to the test / validation set (p. 85); a normalized comparison against the results of this study is provided in Table 19. Mean squared error, used as the performance measure in this study, has been converted to root mean square error in the table.

Table 18

*J-Shaped (Non-Linear) Combined Model Performance Results from von Helversen and Rieskamp Study 2.*

	Mapping	Regression	QuickEst	Exemplar
Old Profiles				
RMSD	160	139	244	165
$SD_{\text{RMSD}}$	35	36	33	35
$r^2$	0.68	0.76	0.33	0.68
$SD_r^2$	0.13	0.11	0.09	0.12
New Profiles				
RMSD	174	172	246	184
$SD_{\text{RMSD}}$	43	58	51	42
$r^2$	0.38	0.50	0.25	0.37
$SD_r^2$	0.19	0.18	0.14	0.15
Total Profiles				
RMSD	99	186	156	118
$SD_{\text{RMSD}}$	13	17	21	18
$r^2$	0.77	0.36	0.44	0.70
$SD_r^2$	0.13	0.08	0.11	0.09

*Note.* Adapted from “The mapping model: A cognitive theory of quantitative estimation”, by B. von Helversen and J. Rieskamp, 2008, p. 85. Copyright 2008 by the American Psychological Association.

Table 19

*Normalized Comparison of von Helversen and Rieskamp Combined Root Mean Squared Deviation Versus the Root Mean Square Error Results of this Study.*

	von Helversen and Rieskamp (2008)	This study
Mapping	0.53	0.96
Exemplar	0.63	0.94
QuickEst	0.84	1.08
Linear	1.00	1.00
Neural	n/a	1.14
Fuzzy	n/a	0.96

The models in this study did not exhibit the range of variation found in the von Helversen and Rieskamp study although on a gross scale, the relative performance was similar, with the mapping and exemplar models outperforming the linear and QuickEst models. In addition to the fact that the models were simulating different behaviors, the lack of variation in this study may have been due to the larger number of training records used. In this study the training records were selected to ensure full coverage of the cue values and it is possible that the models may have been over fitted to the data.

#### *Conclusions with Respect to Research Questions 2 and 3*

The findings of the modeling and computer simulation lead to the following conclusions regarding research questions 2 and 3:

1. The fuzzy, exemplar, and mapping models produced nominally equivalent results with a normalized mean squared error of less than 0.050.
2. The fuzzy, exemplar, and mapping models performed better than the linear regression, neural network, and QuickEst heuristic models with normalized mean squared errors of 0.053, 0.069, and 0.062 respectively.

#### Implications for Social Change

The National Science Foundation manages a research budget of over \$6 billion and issues over 10,000 grant awards a year out of over 42,000 proposals received (National Science Foundation, 2009a, 2009b). Funds are allocated to support both the physical and social sciences for the specific purpose of promoting “the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense” (National Science Foundation, 2009b). Though sizeable, these funds are not

unlimited; fiscal year 2009 federal funding for R&D with respect to gross domestic product declined for the fifth year in a row (American Association for the Advancement of Science, 2008).

With a 25% award rate, it is imperative that these awards be made as accurately and fairly as possible. Of those proposals awarded, only a fraction will actually succeed. In a 3-year longitudinal study of 69 US Department of Defense research grant awards, Gailbrath et al. reported that only 52% of the grants were considered successful (2007, p. 128). Gailbrath et al. also note that although they studied the success rate the awarded grants, the study did not consider the rejected proposals that might have been successful.

The positive social impact of this work can therefore be found in the potential development of decision support tools that yield better grant decisions. Better grant decisions yield more efficient allocation and prioritization of research funds in all areas of science including social, psychological, medical, pharmaceutical, and engineering. Better use of funds results in more money being available to support additional work to advance the national health, prosperity, and general welfare.

#### Recommendations for Action

Two principal communities may find this work to be of value but for different reasons. Those who are preparing research grant proposals will gain insight into the emphasis they must place on their proposals; this study lends credence to advice that is typically presented anecdotally including emphasizing a well planned and lean cost proposal, selecting a challenging topic, and ensuring adequate resources in place to complete the proposed effort.



Those who administer grant proposals and are responsible for awards will see the potential for emotional responses in the evaluation process and the diversity of opinions that can result. The need for better decision tools to improve the evaluation process is evident. Using new approaches demonstrated in this study that have been shown to improve upon traditional methods of modeling these processes, further research into peer review, judgment analysis, and decision support is strongly recommended.

#### Recommendations for Further Study

This study focused on the relationship between cost, schedule, scope, and resources without considering how one would evaluate these components. A logical extension of this work would be an investigation into each component individually. It was also apparent from the participant comments that the research, economic or social value of the project was a factor that needed to be included in their assessment. This may be more of a psychological factor than a true component of the judgment because value is considered elsewhere in the overall assessment; the participants however were uncomfortable making the assessment without the information.

The policy-capturing exercise survey instrument could stand to be improved based upon feedback received from the participants and upon further consideration by this writer. The policy-capturing survey design incorporated numerical scores in each scenario for each cue to facilitate the interpretation of the cue's intended value; these scores were based upon the reasoning used to create the fuzzy membership functions and may have resulted in a scoring bias that favored the fuzzy model. The format of the

scenario visual presentation, as well as the number of scenarios presented also requires further consideration.

With respect to the cognitive models considered in this study, the exemplar, mapping, and fuzzy models performed well using the generalizations that were made to create the models. However, additional work should be performed to optimize the models:

1. The conversion function used to dichotomize the cue values in the metrics and mapping model requires more thought; alternative approaches should be evaluated.
2. The fuzzy model membership functions can be adjusted, possibly using a Delphi method or genetic algorithms as proposed by Mateou and Andreou (2008).
3. The exemplar model attention function that determines the distance weighting to apply to the exemplars can be explored and adjusted.
4. The multi-layer perceptron model with back propagation is only one of many possible artificial neural network instantiations; in addition to working with the current model's configuration parameters, other neural models are worth exploring further.

### Concluding Remarks

This study represents the culmination of several years of intensive work on the part of this writer, yet at the same time it is only a starting point. The original research questions have been answered yet new questions have appeared. Most notably, this writer would like to know what went wrong with the artificial neural network model that held so

much promise when this study was designed. The metrics and mapping model also presents itself as a fascinating concept worth further exploration.

As an exploratory study there are no claims to be made other than having achieved a better understanding of how we estimate reasonableness and how we can better model the judgment process. The irrational behavior detected in cost assessment has not gone unnoticed and it is recognized that simply reproducing the will of the general population won't lead to better decisions. The cognitive models evaluated in this study are only as good as the material used to train them, hence, to use them in a decision support system would require due diligence in creating a training data set. Group consensus building techniques such as Wideband Delphi (Stellman & Greene, 2006) are well suited for this type of development.

That the findings of this study can contribute to a more rational and equitable distribution of grant funds is encouraging. The social impact can be measured in the greater number of deserving programs that can be funded when bias and irrational decision making are effectively mitigated. It is therefore this writer's intention to pursue this work further with the ultimate goal of developing a viable decision support tool for the research community.

## REFERENCES

- Aiman-Smith, L., Scullen, S. E., & Barr, S. H. (2002). Conducting studies of decision making in organizational contexts: A tutorial for policy-capturing and other regression-based techniques. *Organizational Research Methods*, 5(4), 388-414. doi: 10.1177/109442802237117
- American Association for the Advancement of Science. (2008). R&D Budget and Policy Program, from <http://www.aaas.org/spp/rd/fy09.htm>
- Banghart, R. P. P. D. (2006). *An examination of the peer review process of a large research organization*. Retrieved from ProQuest Digital Dissertations. (UMI No. 3248514)
- Beckstead, J. W. (2007). A note on determining the number of cues used in judgment analysis studies: The issue of type II error. *Judgment and Decision Making*, 2(5), 317-325. Retrieved from <http://journal.sjdm.org/jdm7711.pdf>
- Bellingan, L. (2007). Reviewing peer review. *Biologist*, 54(1), 8-8. Retrieved from EBSCO Academic Search Premier (AN: 24599771)
- Benaroch, M., Jeffery, M., Kauffman, R. J., & Shah, S. (2007). Option-based risk management: a field study of sequential information technology investment decisions. *Journal of Management Information Systems*, 24(2), 103-140. doi: 10.2753/MIS0742-1222240205
- Bolton, A. E., Astwood, R. S., Jr., & Campbell, G. E. (2002). *Policy capturing and fuzzy logic: a better approach to representing judgment data*. . Paper presented at the Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Santa Monica, CA, USA.
- Bossé, É., Roy, J., & Wark, S. (2007). *Concepts, models, and tools for information fusion*. Norwood, MA: Artech House, Inc.
- Broome, J. (2008). Reasoning to be rational. In M. C. Galovatti, R. Scazierrri & P. Suppes (Eds.), *Reasoning, Rationality, and Probability* (pp. 119-138). Stanford, CA: CSLI Publications.
- Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review*, 100(3), 511-534. doi: 10.1037/0033-295X.100.3.511
- Brunswik, E., Hammond, K. R., & Stewart, T. R. (2001). *Essential Brunswik : Beginnings, Explications, Applications*. New York: Oxford University Press, Incorporated.

- Cenziper, D. (2009, December 13). Wasting away: The squandering of D.C.'s AIDS dollars, *The Washington Post*. Retrieved from <http://www.washingtonpost.com/wp-srv/special/metro/aids-funding/>
- Chin, K.-S., Yang, J.-B., Guo, M., & Lam, J. P.-K. (2009). An evidential-reasoning-interval-based method for new product design assessment. *IEEE Transactions on Engineering Management*, *56*(1), 142-156. doi: 10.1109/TEM.2008.2009792
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159. doi: 10.1037/0033-2909.112.1.155
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Cookson, C., & Jack, A. (2008). Science stifled? Why peer review is under pressure, *Financial Times* p. 11. Retrieved from Proquest ABI/INFORM Global database (Document ID: 1494273491)
- Cooper, R., Edgett, S., & Kleinschmidt, E. (2001). Portfolio management for new product development: results of an industry practices study. *R&D Management*, *31*(4), 361-380. doi: 10.1111/1467-9310.00225
- Cooper, W. S. (1987). Decision theory as a branch of evolutionary theory: A biological derivation of the savage axioms. *Psychological Review*, *94*(4), 395-411. doi: 10.1037/0033-295X.94.4.395
- Coppin, B. (2004). *Artificial intelligence illuminated* (1st ed.). Boston: Jones and Bartlett Publishers.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, *24*(1), 87. doi: 10.1017/S0140525X01003922
- De Piante Henriksen, A., & Palocsay, S. W. (2008). An Excel based decision support system for scoring and ranking proposed R&D projects. *International Journal of Information Technology & Decision Making* *7*(3), 529-546. doi: 10.1142/S0219622008003022
- De Piante Henriksen, A., & Traynor, A. J. (1999). A practical R&D project-selection scoring tool. *IEEE Transactions on Engineering Management*, *42*(2), 158-170. doi: 10.1109/17.759144

- Doherty, M. E. (2006). Lens model. In S. G. Rogelberg (Ed.), *Encyclopedia of Industrial and Organizational Psychology*. Thousand Oaks, CA: Sage Publications.  
Retrieved from [http://www.sage-reference.com/organizationalpsychology/Article\\_n188.html](http://www.sage-reference.com/organizationalpsychology/Article_n188.html).
- Domino, G., & Domino, M. (2002). *Psychological testing : An introduction*. New York: Cambridge University Press.
- Dorsey, D. W., & Coovert, M. D. (2003). Mathematical modeling of decision making: a soft and fuzzy approach to capturing hard decisions. . *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(1), 117-135 doi: 10.1518/hfes.45.1.117.27228
- Drolet, A., & Suppes, P. (2008). The good and the bad, the true and the false. In M. C. Galovatti, R. Scazierra & P. Suppes (Eds.), *Reasoning, Rationality, and Probability* (pp. 13-35). Stanford, CA: CSLI Publications.
- Ebersole, J. (2004). The future of graduate education: as graduate student demographics change, institutions also must change to keep pace. *University Business*, 2004(August). Retrieved from <http://www.universitybusiness.com/viewarticle.aspx?articleid=527>
- Ennis, T. R. P. D. (2007). *Assessing and enhancing critical thinking skills: Enhanced Peer Review (EPR) with the CAT instrument*. Retrieved from Proquest Digital Dissertations (UMI No. 3263266)
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191. doi: 10.3758/BRM.41.4.1149
- Faulkner, T. W. (1996). Applying options thinking to R&D valuation. *Research Technology Management*, 39(3), 50(57). Retrieved from EBSCO Business Source Premier (AN: 9605083470)
- Fenton, A. (2006). Weft QDA (Version 1.0.1) [Computer Software]. Retrieved from <http://www.pressure.to/qda/>
- Fox, J. (2009). Package 'car' (Version 1.2-16) [Computer Software]. Retrieved from <http://cran.r-project.org/web/packages/car/index.html>
- Galbraith, C. S., DeNoble, A. F., Ehrlich, S. B., & Kline, D. M. (2007). Can experts really assess future technology success? A neural network and Bayesian analysis of early stage technology proposals. *Journal of High Technology Management Research*, 17(2), 125-137. doi: 10.1016/j.hitech.2006.11.002

- Gehris, R. (2008). *A simulation study of the classical method of expert judgment combination: How many seeds and how many experts.* . Retrieved from ProQuest Digital Dissertations. (UMI No. 3297446)
- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62(4), 255-262. doi: 10.1037/h0046845
- Hammond, K. R. (1996). Upon reflection. *Thinking & Reasoning*, 2, 239-248. doi: 10.1080/135467896394537
- Hansson, S. O. (2005). *Decision Theory - A Brief Introduction*. Royal Institute of Technology. Stockholm. Retrieved from <http://www.infra.kth.se/~soh/decisiontheory.pdf>
- Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation: Letting the environment do the work. In G. Gigerenzer & P. M. Todd (Eds.), *Simple Heuristics that make us Smart*. (pp. 209-234). New York: Oxford University Press.
- Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 57(2), 116-131. doi: 10.1037/h0047807
- Hovanov, N., Yudaeva, M., & Hovanov, K. (2009). Multicriteria estimation of probabilities on basis of expert non-numeric, non-exact and non-complete knowledge. *European Journal of Operational Research*, 195(3), 857-863. doi: 10.1016/j.ejor.2007.11.018
- Huang, C.-C., Chu, P.-Y., & Chiang, Y.-H. (2008). A fuzzy AHP application in government-sponsored R&D project selection. [Article]. *Omega*, 36, 1038-1052. doi: 10.1016/j.omega.2006.05.003
- Huang, X., & White, R. W. (2005). Policy capturing models for multi-faceted relevance judgments. *Proceedings of the American Society for Information Science and Technology*, 42(1), NA. doi: 10.1002/meet.14504201253
- Inselberg, A. (2005). Visualization of concept formation and learning. *Kybernetes: The International Journal of Systems & Cybernetics*, 34(1/2), 151-166. doi: 10.1108/03684920510699972
- Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2003). A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. *Journal of the Royal Statistical Society: Series A Statistics in Society* 166(3), 279(222). doi: 10.1111/1467-985X.00278

- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26. doi: 10.3102/0013189X033007014
- Karren, R. J., & Barringer, M. W. (2002). A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods*, 5(4), 337-361. doi: 10.1177/109442802237115
- Kight, W. (2009). *KAM 7 Research Design*. Walden University. Minneapolis, MN.
- Laudel, G. (2006). Conclave in the Tower of Babel: how peers review interdisciplinary research proposals. *Research Evaluation*, 15(1), 57-68. doi: 10.3152/147154406781776048
- LaVoie, N. N., Bourne, L. E., Jr., & Healy, A. F. (2002). Memory seeding: Representations underlying quantitative estimations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6), 1137-1153. doi: 10.1037/0278-7393.28.6.1137
- Limas, M. C., Meré, J. B. O., González, E. P. V., Ascacibar, F. J. M. d. P., Espinoza, A. V. P., & Elías, F. A. (2007). AMORE flexible neural network package (Version 0.2-11) [Computer Software]. Retrieved from <http://cran.r-project.org/web/packages/AMORE/index.html>
- MacMillan, I. C., Van Putten, A. B., McGrath, R. G., & Thompson, J. D. (2006). Using real options discipline for highly uncertain technology investments. *Research Technology Management*, 49, 29-37. Retrieved from EBSCO Business Source Premier database (AN: 19962272)
- Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, 63(3), 160-168. doi: 10.1037/0003-066X.63.3.160
- Mateou, N. H., & Andreou, A. S. (2008). A framework for developing intelligent decision support systems using evolutionary fuzzy cognitive maps. *Journal of Intelligent & Fuzzy Systems*, 19, 151-170. Retrieved from EBSCO Business Source Premier database (AN: 31547102)
- Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, 5(4), 434-458. doi: 10.1037/1082-989X.5.4.434
- Meade, L. M., & Presley, A. (2002). R&D project selection using the Analytic Network Process. *IEEE Transactions on Engineering Management*, 49(1), 59. doi: 10.1109/17.985748

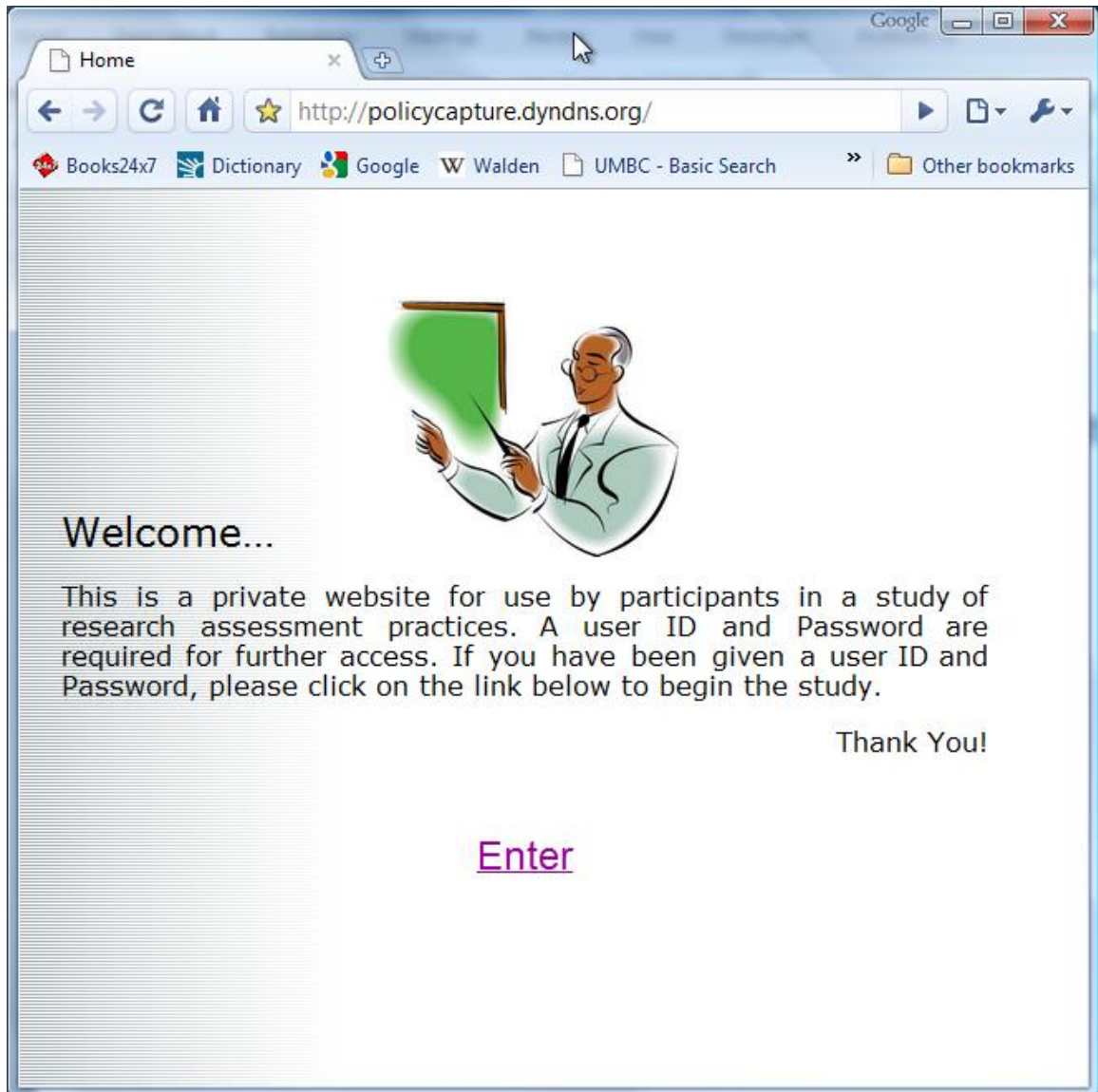


- National Science Foundation. (2009a, 09 Jun 2009). NSF - How We Work, from <http://www.nsf.gov/about/how.jsp>
- National Science Foundation. (2009b, 27 Mar 2009). NSF - Overview, from <http://www.nsf.gov/about/glance.jsp>
- Newell, B., & Broder, A. (2008). Cognitive processes, models and metaphors in decision research. *Judgment and Decision Making*, 3(3), 195-204. Retrieved from <http://journal.sjdm.org/bn1.pdf>
- Nissen, S. (2003). *Implementation of a Fast Artificial Neural Network library (FANN)*. Department of Computer Science, University of Copenhagen. Retrieved from <http://sourceforge.net/projects/fann/>
- Obrecht, M., Tibelius, K., & D'Aloisio, G. (2007). Examining the value added by committee discussion in the review of applications for research awards. *Research Evaluation*, 16, 79-91. doi: 10.3152/095820207X223785
- Politzer, G. (2004). Reasoning, judgment, and pragmatics. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 94-115). London: Palgrave Macmillan.
- Python Software Foundation. (2009a). Python (Version 2.6.2) [Programming Language]. Hampton, NH. Retrieved from [www.python.org](http://www.python.org)
- Python Software Foundation. (2009b). Python v2.6.2 documentation : The Python standard library : Numeric and mathematical modules, from <http://docs.python.org/library/random.html>
- Rakowsky, U. K. (2007). Fundamentals of the Dempster-Shafer theory and its applications to reliability modeling. *International Journal of Reliability, Quality & Safety Engineering*, 14, 579-601. doi: 10.1142/S0218539307002817
- Reed, R. D., & Marks, R. J. (1999). *Neural smithing : supervised learning in feedforward artificial neural networks*. Cambridge, Mass.: The MIT Press.
- Rolf, B. (2006). Decision support tools and two types of uncertainty reduction. In L. Emmelin (Ed.), *Effective Environmental Assessment Tools - Critical Reflections on Concepts and Practice - Blekinge Institute of Technology Research report No. 2006:03* (pp. 134-157). Karlskrona, Sweden: Kasernttryckeriet AB.
- Saha, A. (2002). *Introduction to artificial neural networks*. [PowerPoint slides]. Retrieved from <https://engineering.purdue.edu/people/mireille.boutin.1/ECE301kiwi/FeedForwardBackPropagationAlgorithm>

- Sahlin, N., & Vareman, N. (2008). Three types of decision theory. In M. C. Galovatti, R. Scazierra & P. Suppes (Eds.), *Reasoning, Rationality, and Probability* (pp. 37-59). Stanford, CA: CSLI Publications.
- Schwartz, D. A., Mastin, J. P., & Martin, M. (2006). Improving grant application peer review for the NIEHS. *Environmental Health Perspectives*, *114*(5), A270-A270. Retrieved from EBSCO Academic Search Premier database (AN: 20814569)
- Silberglitt, R., Sherry, L., Wong, C., Tseng, M., Ettetdgui, E., Watts, A., et al. (2004). *Portfolio analysis and management for Naval research and development*. Santa Monica, CA: RAND Corporation.
- Singleton, R., & Straits, B. (2005). *Approaches to social research* (4th ed.). New York: Oxford University Press.
- Srivastava, C. V., Towery, N. D., & Zuckerman, B. (2007). Challenges and opportunities for research portfolio analysis, management, and evaluation *Research Evaluation*, *16*(3), 152-156. doi: 10.3152/095820207X236385
- Srivastava, P. (2006). *The effect of new product development project portfolio mix on new product development program performance*. Retrieved from Proquest Digital Dissertations (UMI No. 3252778)
- Stellman, A., & Greene, J. (2006). *Applied software project management*. Sebastopol, CA: O'Reilly.
- Stewart, T., Ward, J., Naylor, J., & Cooksey, R. (1998). Comments on the distinction between "policy capturing" and "judgment analysis" Retrieved 19 Apr, 2009, from <http://www.albany.edu/cpr/brunswik/notes/japc.html>
- Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools*, *13*(1), 12-28. Retrieved from EBSCO PsycINFO database (AN: 2007-09345-002)
- The Higher Learning Commission. (2009, 4/15/2009). Affiliated Institutions Retrieved 12 June, 2009, from [http://www.ncahlc.org/index.php?option=com\\_directory&Itemid=192&Action=ShowBasic&instid=1574](http://www.ncahlc.org/index.php?option=com_directory&Itemid=192&Action=ShowBasic&instid=1574)
- Thomaidis, N. S., Nikitakos, N., & Dounias, G. D. (2006). The evaluation of information technology projects: A fuzzy multicriteria decision-making approach. *International Journal of Information Technology & Decision Making*, *5*, 89-122. doi: 10.1142/S0219622006001897

- Trochim, W. (2001). *The research methods knowledge base* (2nd ed.). Mason, OH: Thompson.
- Vasant, P. M. (2005). Fuzzy linear programming for decision making and planning under uncertainty. [Article]. *International Journal of Information Technology & Decision Making*, 4, 647-662. doi: 10.1142/S0219622005001751
- Von Altrock, C. (1997). *Fuzzy logic and neuroFuzzy applications in business and finance*. Upper Saddle River, N.J.: Prentice Hall PTR.
- von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, 137(1), 73-96. doi: 10.1037/0096-3445.137.1.73
- Yu, L., Wang, S., & Lai, K. K. (2009). An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring. *European Journal of Operational Research*, 195(3), 942-959. doi: 10.1016/j.ejor.2007.11.025

## APPENDIX A: POLICY-CAPTURING DISPLAYS



*Figure A1.* Eligible participants will be given a username and password to access the exercise.

Consent Form

Please read the following statement carefully...

**Please, don't use the back button.**

[Questions?](#)

[Exit](#)

You are invited to take part in a research study of the factors that influence an evaluator's assessment of a research proposal's reasonableness. You were chosen for the study because you are familiar with research proposals and have some experience evaluating them. This form is part of a process called "informed consent" to allow you to understand this study before deciding whether to take part.

This study is being conducted by a researcher named William Kight, who is a doctoral student at Walden University. Mr. Kight is also an employee of Verizon Business, Inc. which has kindly sponsored his participation in Walden University's doctoral program.

**Background Information:**

The purpose of this study is to conduct what is known as a policy-capturing exercise. Policy-capturing is a method of interviewing that helps researchers understand how people make decisions. The exercise will provide insight into the factors that influence an assessment of a proposal's reasonableness.

**Procedures:**

If you agree to be in this study, you will be asked to:

- o Answer a brief series of questions describing your background
- o Participate in the exercise where you will act in the role of a senior manager who must assess a collection of 50 proposals that have been scored in advance by your team on the basis of scope, resources, cost, and schedule.
- o Answer a brief series of questions about your opinions of the exercise and what factors you felt were most important.
- o The entire exercise will take about 30 minutes. Please ensure that you have enough time to complete the exercise because an intermediate survey cannot be saved.

**Voluntary Nature of the Study:**

Your participation in this study is voluntary. This means that everyone will respect your decision of whether or not you want to be in the study. No one will treat you differently if you decide not to be in the study. If you decide to join the study now, you can still change your mind during the study. If you feel stressed during the study you may stop at any time. You may skip any questions that you feel are too personal.

Figure A2. A standard consent form with an explanation of the exercise is presented to the participant. (continued next figure)

**Risks and Benefits of Being in the Study:**

Participating in this study will help to improve the proposal review process ensuring that the best proposals get selected fairly. There are no known risks to participating in this study.

**Compensation:**

Compensation for your participation in this study is limited to the sincerest appreciation on the part of the researcher. Your time and cooperation are gratefully accepted.

**Confidentiality:**

Any information you provide will be kept anonymous. No user identifiable information is collected or maintained in the course of this exercise. The researcher will not use your information for any purposes outside of this research project. Also, the researcher will not include your name or anything else that could identify you in any reports of the study.

**Contacts and Questions:**


You may ask any questions you have now. Or if you have questions later, you may contact the researcher:

- via phone at 410 309-7362
- via e-mail at William.Kight@waldenu.edu.

If you want to talk privately about your rights as a participant, you can call Dr. Leilani Endicott. She is the Walden University representative who can discuss this with you. Her phone number is 1-800-925-3368, extension 1210. Walden University's approval number for this study is **IRB approval number** and it expires on **IRB expiration date**.

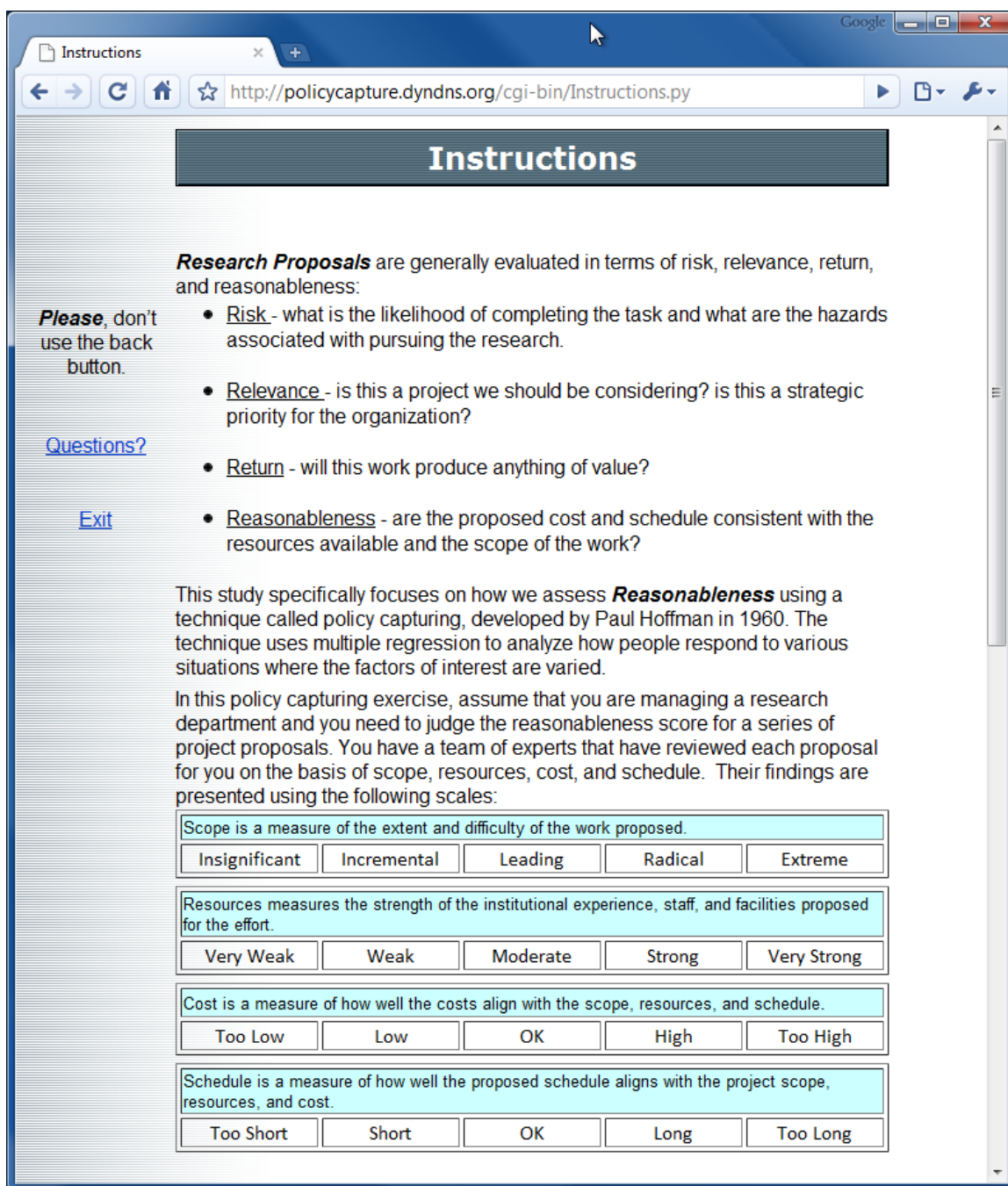
**Statement of Consent:**

I have read the above information and I feel I understand the study well enough to make a decision about my involvement. By clicking on the "I agree" link provided below, I am agreeing to the terms described above. I may close this browser window to exit the study now or at anytime during the course of the exercise. Please print a copy of this page for your records.

BUILT WITH  
NetObjects  
**FUSION**  
ESSENTIALS  python  
powered

The projects, names of individuals, and ratings are fictional. Any resemblance to or association with real projects or people is unintentional. Copyright 2009. All Rights Reserved.

Figure A3. The user must agree with the consent form in order to proceed with the exercise. (continued from previous figure)



**Instructions**

**Research Proposals** are generally evaluated in terms of risk, relevance, return, and reasonableness:

**Please**, don't use the back button.

[Questions?](#)

[Exit](#)

- **Risk** - what is the likelihood of completing the task and what are the hazards associated with pursuing the research.
- **Relevance** - is this a project we should be considering? is this a strategic priority for the organization?
- **Return** - will this work produce anything of value?
- **Reasonableness** - are the proposed cost and schedule consistent with the resources available and the scope of the work?

This study specifically focuses on how we assess **Reasonableness** using a technique called policy capturing, developed by Paul Hoffman in 1960. The technique uses multiple regression to analyze how people respond to various situations where the factors of interest are varied.

In this policy capturing exercise, assume that you are managing a research department and you need to judge the reasonableness score for a series of project proposals. You have a team of experts that have reviewed each proposal for you on the basis of scope, resources, cost, and schedule. Their findings are presented using the following scales:

Scope is a measure of the extent and difficulty of the work proposed.				
Insignificant	Incremental	Leading	Radical	Extreme

Resources measures the strength of the institutional experience, staff, and facilities proposed for the effort.				
Very Weak	Weak	Moderate	Strong	Very Strong

Cost is a measure of how well the costs align with the scope, resources, and schedule.				
Too Low	Low	OK	High	Too High

Schedule is a measure of how well the proposed schedule aligns with the project scope, resources, and cost.				
Too Short	Short	OK	Long	Too Long

Figure A4. The instructions page begins with background information for the exercise. (Continued next figure.)

Instructions

http://policycapture.dyndns.org/cgi-bin/Instructions.py

Each proposal is presented as a summary worksheet like the one shown below:

Preliminary Team Assessment Report						
Project # 1 of 50	Project ID # 85					
Principal Investigator	Page, Lyman					
Project Name	Inorganic Distributed Nanocrystal Heterojunction Solar Cells					

Criteria	Description					Rating
<b>Scope</b>	The scope of this project was found to be insignificant and does not meaningfully contribute to the discipline.					Insignificant
	insignificant	incremental	leading	radical	extreme	Score 1-5 pts
	X					1
<b>Resources</b>	The proposal has identified nominally sufficient resources to perform the work.					Moderate
	Very Weak	Weak	Moderate	Strong	Very Strong	Score 1-5 pts
			X			3
<b>Cost</b>	The costs are somewhat high for the scope of the project and the labor and expenses proposed.					High
	Very Low	Low	OK	High	Very High	Score 1-5 pts
				X		3
<b>Schedule</b>	The schedule is very short for the scope of the project and the labor and resources proposed.					Too Short
	Very Short	Short	OK	Long	Very Long	Score 1-5 pts
	X					1

With respect to **Reasonableness**, how would you rate this proposal?

Very Unreasonable	Unreasonable	Somewhat Unreasonable	Neutral	Somewhat Reasonable	Reasonable	Very Reasonable
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Review the next proposal >>

Your role is to judge how **reasonable** each proposal is based upon your team's assessment of the four critical factors. Select a value that best represents your judgment about the reasonableness of the proposal, then click on the Next button to proceed.

There are a total of 50 proposals to review; it should take less than 30 minutes to complete the exercise - you may exit the exercise at any time by closing your browser window. Setting your browser to full screen will make entering your responses easier. Click on the Begin button below to start the exercise!

Start Here ! >>>

BUILT WITH NetObjects FUSION ESSENTIALS python powered

The projects, names of individuals, and ratings are fictional. Any resemblance to or association with real projects or people is unintentional. Copyright 2009. All Rights Reserved.

Figure A5. The instructions then explain the participant's role and how to answer the questions. (Continued from previous figure.)



Scenario

http://policycapture.dyndns.org/cgi-bin/scenarios.py

### Preliminary Team Assessment Report

Project # 7 of 50	Project ID # 208
Principal Investigator	Harley, Ronald
Project Name	Biaxial Semiconductor Films on Amorphous Substrates for Solar Energy Conversion

**Please, don't use the back button.**

[Questions?](#)

[Exit](#)

Criteria	Description	Rating										
<b>Scope</b>	This project addresses a radically new and promising concept.	Radical										
	<table border="1"> <tr> <td>insignificant</td> <td>incremental</td> <td>leading</td> <td>radical</td> <td>extreme</td> </tr> <tr> <td></td> <td></td> <td></td> <td>X</td> <td></td> </tr> </table>	insignificant	incremental	leading	radical	extreme				X		Score 1-5 pts 3
insignificant	incremental	leading	radical	extreme								
			X									
<b>Resources</b>	The resources available to perform the work specified are marginally capable of performing the work proposed.	Weak										
	<table border="1"> <tr> <td>Very Weak</td> <td>Weak</td> <td>Moderate</td> <td>Strong</td> <td>Very Strong</td> </tr> <tr> <td></td> <td>X</td> <td></td> <td></td> <td></td> </tr> </table>	Very Weak	Weak	Moderate	Strong	Very Strong		X				Score 1-5 pts 2
Very Weak	Weak	Moderate	Strong	Very Strong								
	X											
<b>Cost</b>	The costs are somewhat high for the scope of the project and the labor and expenses proposed.	High										
	<table border="1"> <tr> <td>Very Low</td> <td>Low</td> <td>OK</td> <td>High</td> <td>Very High</td> </tr> <tr> <td></td> <td></td> <td></td> <td>X</td> <td></td> </tr> </table>	Very Low	Low	OK	High	Very High				X		Score 1-5 pts 3
Very Low	Low	OK	High	Very High								
			X									
<b>Schedule</b>	The schedule is somewhat short for the scope of the project and the labor and resources proposed.	Short										
	<table border="1"> <tr> <td>Very Short</td> <td>Short</td> <td>OK</td> <td>Long</td> <td>Very Long</td> </tr> <tr> <td></td> <td>X</td> <td></td> <td></td> <td></td> </tr> </table>	Very Short	Short	OK	Long	Very Long		X				Score 1-5 pts 3
Very Short	Short	OK	Long	Very Long								
	X											

With respect to **Reasonableness**, how would you rate this proposal?

Very Unreasonable	Unreasonable	Somewhat Unreasonable	Neutral	Somewhat Reasonable	Reasonable	Very Reasonable
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Review the next proposal >>

**Reasonableness:** An estimate of the degree of belief that a project can be completed within the constraints imposed by the scope, cost, schedule, and resources proposed without incurring unnecessary risk or expense.

BUILT WITH  
NetObjects  
**FUSION**  
ESSENTIALS

python  
powered

The projects, names of individuals, and ratings are fictional. Any resemblance to or association with real projects or people is unintentional. Copyright 2009. All Rights Reserved.

Figure A6. The participant is presented with a series of 50 randomly generated scenarios to evaluate.

Demographics

Please tell me a little about yourself. Your responses are completely confidential.

**Please**, don't use the back button.

[Questions?](#)

[Exit](#)

1. What is your age group?
 

< 25	25 - 45	> 45
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. How many years of experience do you have managing and planning projects?
 

< 5	5-15	> 15
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. How many years of experience do you have specifically with planning or evaluating research projects?
 

< 3	3 - 10	> 10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. What is your highest level of education achieved?
 

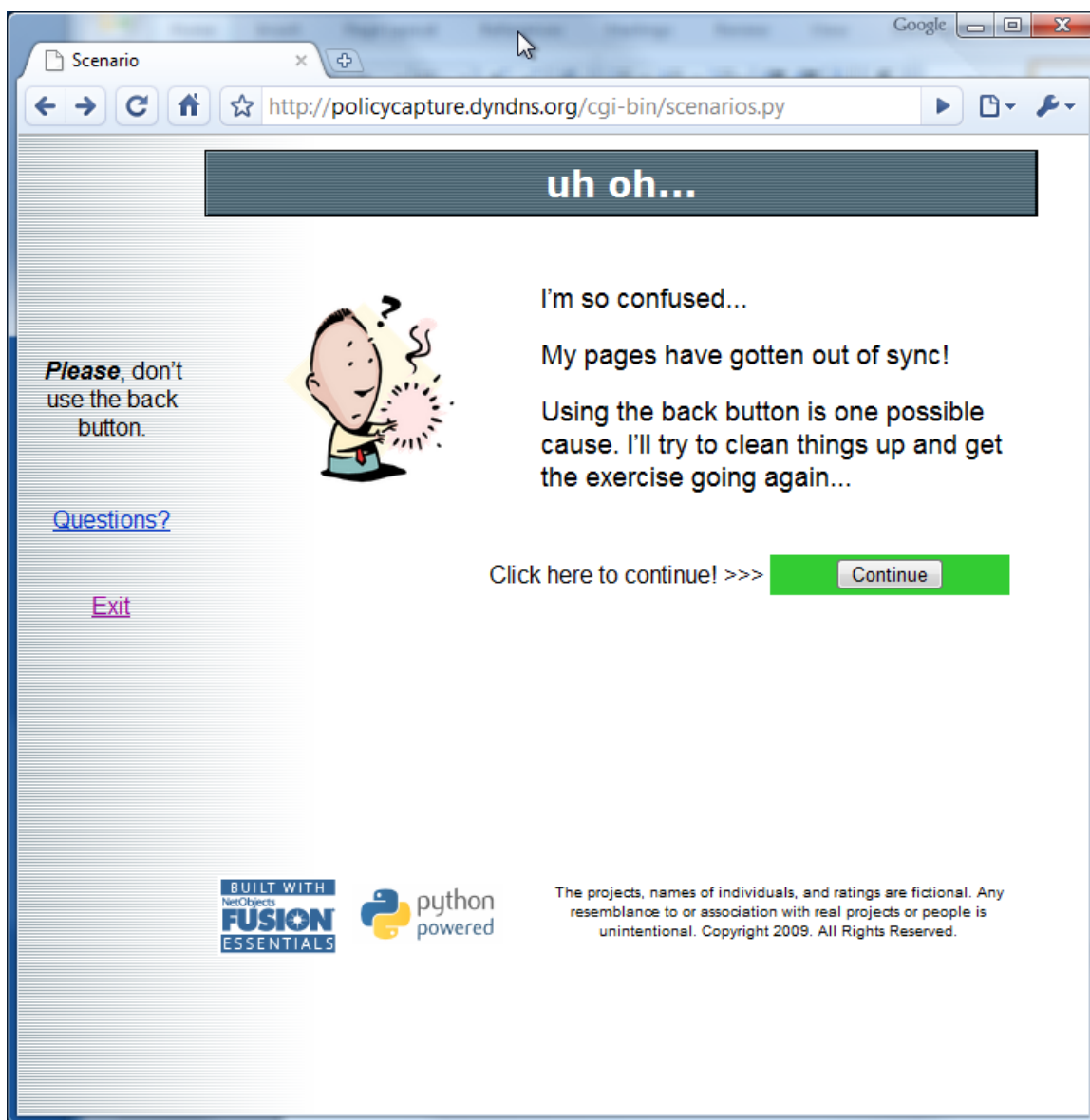
Bachelors	Masters	Ph.D.
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. What is your professional background?

Begin reviewing the scenarios!>>>

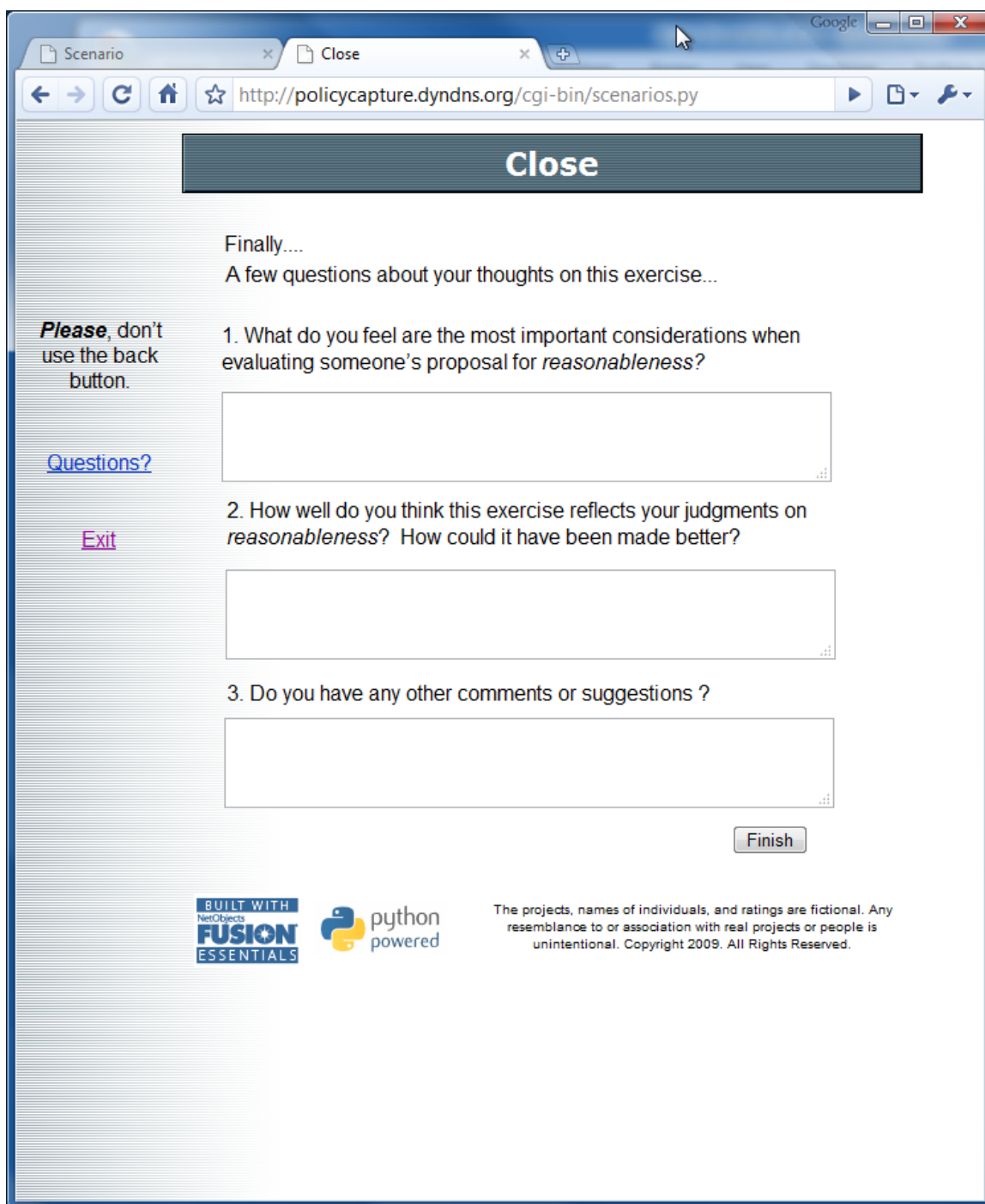
BUILT WITH NetObjects FUSION ESSENTIALS python powered

The projects, names of individuals, and ratings are fictional. Any resemblance to or association with real projects or people is unintentional. Copyright 2009. All Rights Reserved.

Figure A7. Basic demographics information relevant to the study is collected.



*Figure A8.* If a participant attempts to back up or enter the exercise out of sequence, the program tactfully reminds the participant not to back up.



The screenshot shows a web browser window with two tabs: 'Scenario' and 'Close'. The address bar contains the URL <http://policycapture.dyndns.org/cgi-bin/scenarios.py>. The page content is as follows:

## Close

Finally...  
A few questions about your thoughts on this exercise...

**Please**, don't use the back button.

[Questions?](#)

[Exit](#)

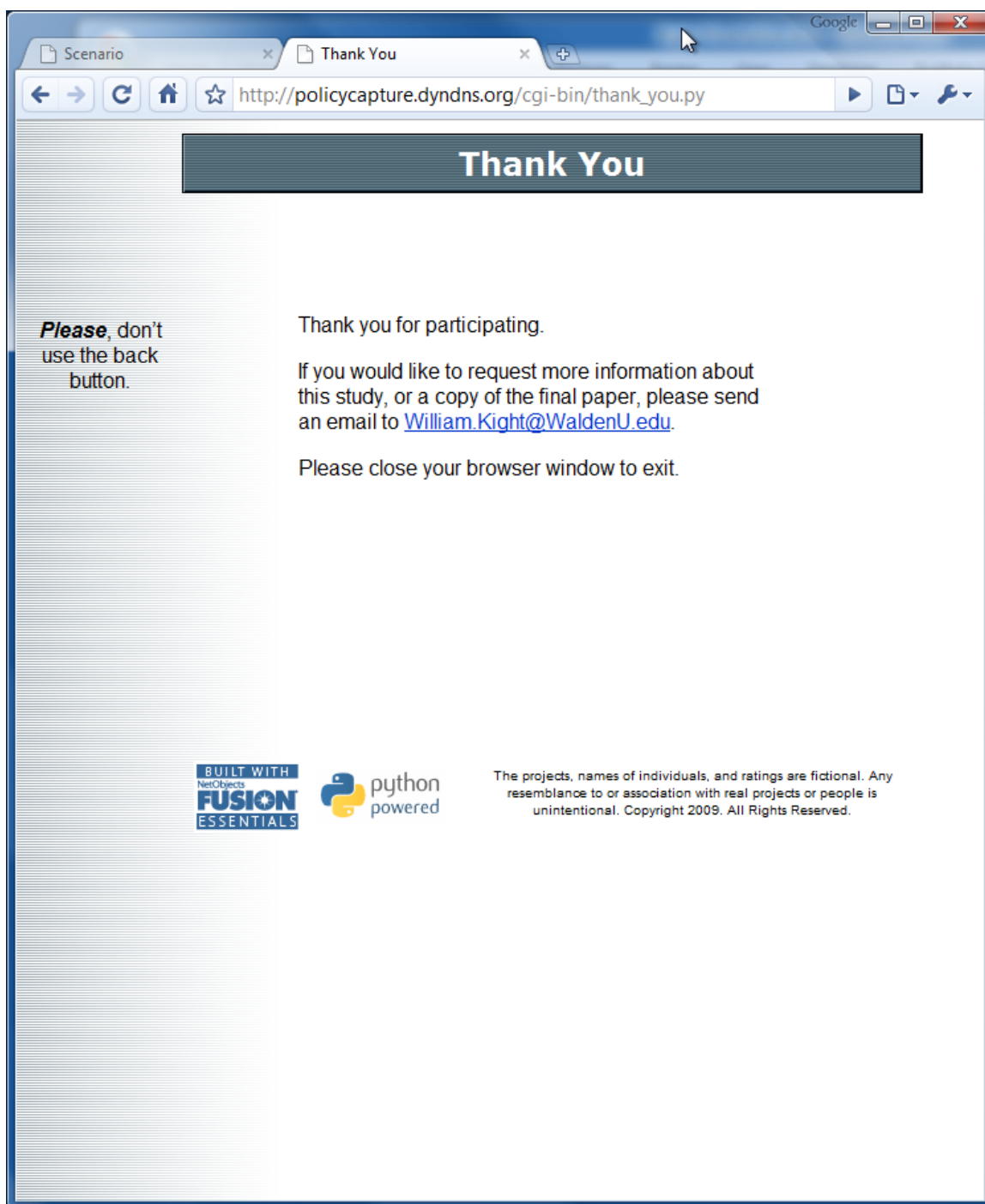
1. What do you feel are the most important considerations when evaluating someone's proposal for *reasonableness*?
2. How well do you think this exercise reflects your judgments on *reasonableness*? How could it have been made better?
3. Do you have any other comments or suggestions ?

BUILT WITH  
NetObjects  
**FUSION**  
ESSENTIALS

python  
powered

The projects, names of individuals, and ratings are fictional. Any resemblance to or association with real projects or people is unintentional. Copyright 2009. All Rights Reserved.

Figure A9. The participant is asked to comment on the exercise before leaving.



*Figure A10.* The participant is provided with a point of contact for more information at the conclusion of the exercise.

## APPENDIX B: WEB SITE PYTHON CODE

### Openconsent.py

```
#!/usr/bin/python
# -*- coding: cp1252 -*-
#
# open the consent form and pass along the entry point value
#
import sys, os, cgi, random, datetime, Cookie
#
# get the entry value
#
form = cgi.FieldStorage()
accesscode = form.getfirst("accesscode", "nocode")
#
# check the access code
#
if accesscode == "Walden":          # participant pool
    entrypointvalue = "1"
elif accesscode == "PolicyStudy": # personal invite
    entrypointvalue = "2"
elif accesscode == "RobertFrost": # testing purposes
    entrypointvalue = "3"
else:                                # set to "nocode"
    entrypointvalue = "0"
#
#
if entrypointvalue == "0":
    #
    #  paint the entry page with error message
    #
    filename="../html/entryerror.html"
    PAGE = open(filename,"r")
    for line in PAGE:
        print line,
    PAGE.close

else:

    #
    # Display the consent form and pass along the endpoint
    #
    htmlsubs=dict()
    htmlsubs["endpoint"]=str(entrypointvalue)

    print "Content-type: text/html"
    print "\n\n"

    #
    # load HTML
    #
    filename="../html/consentform.html"
```

```

PAGE = open(filename,"r")
for line in PAGE:
    print line.replace('%"', '%"%')%htmlsubs,
PAGE.close

```

### Instructions.py

```

#!/usr/bin/python
# -*- coding: cp1252 -*-
#
# Set up session id and collect environment variables
# display instructions page
#
import sys, os, cgi, random, datetime, Cookie
import cPickle as pickle
#
# get the entry value from the form
#
form = cgi.FieldStorage()
entrypoint = form.getfirst("entrypoint", "0")

# pick up environmental variables, set a sequence ID number,
# and write them to a comma separated datafile
sessionid = hex(random.getrandbits(64))
dtemp = datetime.datetime.now()
sessiontime = dtemp.ctime()
sessionip = os.environ.get("REMOTE_ADDR")
#
# write the session information to a file!
#
filename="../data/sessionid.txt"
FILE = open(filename,"a")
record = "%s,%s,%s,%s\n"%(sessionid,sessiontime,sessionip,entrypoint)
FILE.write(record)
FILE.close()
#
# create the scenario sequence and set the repeats
#
sequence=random.sample(xrange(625),60) # set higher than it needs to be
sequence[19]=sequence[9]
sequence[29]=sequence[10]
sequence[39]=sequence[11]
sequence[49]=sequence[12]
#
# pickle the sequence
scenariosave=pickle.dumps(sequence)

#
# Give a cookie to the client , initialize scenario counters
#

```

```

mycookie=Cookie.SimpleCookie()
mycookie["sessionid"] = sessionid           #session identifier
mycookie["scenariocount"] = "0"            #number of scenarios
completed
mycookie["scenarionumber"] = "0"          #current scenario number
mycookie["scenariosave"] = scenariosave    #scenario sequence
mycookie["scenariodone"] = "not yet"      # have we finished the
scenario's

print mycookie

#
# Display the instruction page
#

print "Content-type: text/html"
print "\n\n"

#
# load HTML
#
filename="../html/instructions.html"
PAGE = open(filename,"r")
for line in PAGE:
    print line,
PAGE.close

```

### Scenarios.py

```

#!/usr/bin/python
# -*- coding: cp1252 -*-
#
# Display the Scenarios
#
# William Kight
# Walden University
#
# Updated 7/9/09 - added code to process error handling and eliminated
errorpage.py
#
#
import sys, os, cgi, random, datetime, Cookie
import cPickle as pickle
#
countmax = 50 #number of scenarios to present (set to 50 for study)
#
#
# define description strings
#
scopstr = ["0",

```



```

        "The scope of this project was found to be insignificant and
does not appear to meaningfully contribute to the discipline.",
        "This project represents an incremental development that is
potentially worthwhile but not particularly challenging.",
        "This project is addressing a moderately challenging
problem.",
        "This project addresses a radically new and promising
concept.",
        "The scope of this project is extremely challenging and its
goals will be difficult to achieve."
    ]
#
rsrcstr = ["0",
        "The resources available to perform the work specified
appear to be insufficient.",
        "The resources available to perform the work specified are
marginally capable of performing the work proposed.",
        "The proposal has identified nominally sufficient resources
to perform the work.",
        "The resources available to perform the work specified are
more than capable of performing the work proposed.",
        "The resources available to perform the work specified far
exceed what is needed to perform the work proposed."
    ]
#
coststr = ["0",
        "The costs are very low for the scope of the project and the
labor and expenses proposed.",
        "The costs are somewhat low for the scope of the project and
the labor and expenses proposed.",
        "The costs are in line with the scope of the project and the
labor and expenses proposed.",
        "The costs are somewhat high for the scope of the project
and the labor and expenses proposed.",
        "The costs are very high for the scope of the project and
the labor and expenses proposed."
    ]
#
skedstr = ["0",
        "The schedule is very short for the scope of the project and
the labor and resources proposed.",
        "The schedule is somewhat short for the scope of the project
and the labor and resources proposed.",
        "The schedule is consistent with the scope of the project
and the resources available.",
        "The schedule is somewhat long for the scope of the project
and the labor and resources proposed.",
        "The schedule is very long for the scope of the project and
the labor and resources proposed."
    ]
#
#
# define ratings strings

```

```

#
scopstrv =
["0","Insignificant","Incremental","Leading","Radical","Extreme"]
rsrcstrv = ["0","Very Weak","Weak","Moderate","Strong","Very Strong"]
coststrv = ["0","Very Low","Low","Appropriate","High","Very High"]
skedstrv = ["0","Very Short","Short","Appropriate","Long","Very Long"]
#
# define the score values (fuzzy scores)
#
scopscor = [0,1,4,5,3,1]
rsrcscor = [0,1,2,3,4,5]
costscor = [0,1,3,5,3,1]
skedscor = [0,1,3,5,3,1]
#
# Read the program names
#
filename = "projnames.txt"
FILE = open(filename,"r")
projnames=[]
projnames=FILE.readlines()
FILE.close
#
# Read the investigator names
#
filename = "investigators.txt"
FILE = open(filename,"r")
invnames=[]
invnames=FILE.readlines()
FILE.close
#
# get the cookie
#
cookie_string = os.environ.get('HTTP_COOKIE')
#
if cookie_string: # if the cookie is present
    #
    # recover the cookie values
    #
    mycookie = Cookie.SimpleCookie()
    mycookie.load(cookie_string)
    sessionid = mycookie["sessionid"].value
    scenariocount = mycookie["scenariocount"].value #page number
    scenarionumber = mycookie["scenarionumber"].value #scenario id
    scenariosave = mycookie["scenariosave"].value #sequence string
    scenariodone = mycookie["scenariodone"].value
    sequence=pickle.loads(scenariosave)
    count = int(scenariocount)
#
# pull in form data
#
form = cgi.FieldStorage()
rscore = form.getfirst("rscore", "novalue")
dtemp = datetime.datetime.now()

```

```

rectime = dtemp.ctime()
pagenumber = form.getfirst("pagenumber","0")
scenariovalue = form.getfirst("scenariovalue","novalue")
errorflag = form.getfirst("errorflag","1")
#
#
if not cookie_string:      # make sure cookies are working
    #
    # cookies disabled - print error page
    #
    print "Content-type: text/html"
    print "\n\n"
    #
    filename="../html/nocookie.html"
    FILE = open(filename,"r")
    for line in FILE:
        print line,
    FILE.close
    #
else:      # process the scenario form submission
    #
    # first check to see if we are here by mistake
    #
    if scenariodone == "finished" :      # we've been here before - get
out!
        #
        # paint the thank you page
        #
        print "Content-type: text/html\n\n" #start the page
        #
        filename="../html/thank_you.html"
        FILE = open(filename,"r")
        for line in FILE:
            print line,
        FILE.close()
        #
        #
        # then check to see if the form matches the cookie
        #
        elif int(pagenumber) == count: # if so, then process form
            #
            # save the form data if this is for any page but 0
            # or if we're coming in from the error page
            #
            if count <> 0 and errorflag <> 1:
                #
                # save the scenario form values
                #
                record = "%s,%s,%s,%s,%s\n"
%(sessionid,rectime,pagenumber,scenariovalue,rscore)
                filename="../data/policycapture.txt"
                FILE = open(filename,"a")
                FILE.write(record)

```

```

        FILE.close()
#
# get the next scenario number (sequence index begins at 0)
#
scenario = sequence[count]
newpagenumber = count + 1
#
# calculate the cue values using modulo 5 arithmetic
#
snumber = scenario
nscope = snumber % 5 + 1
snumber = snumber // 5
nresource = snumber % 5 + 1
snumber = snumber // 5
ncost = snumber % 5 + 1
snumber = snumber // 5
nskded = snumber % 5 + 1
#
# calculate the total score
#
nscortotl =
scopscor[nscope]+rsrcscor[nresource]+costscor[ncost]+skedscor[nskded]
#
# set up the dictionary for string substitution
#
htmlsubs=dict()
htmlsubs["counter"] = str(newpagenumber) #page number to be
displayed
htmlsubs["scid"] = str(scenario) #the scenario value
displayed
htmlsubs["piname"] = invnames[count]
htmlsubs["projname"] = projnames[count]
htmlsubs["scopetxt"] = scopstr[nscope]
htmlsubs["rsrctxt"] = rsrcstr[nresource]
htmlsubs["costtxt"] = coststr[ncost]
htmlsubs["skedtxt"] = skedstr[nskded]
htmlsubs["scopeval"] = scopstrv[nscope]
htmlsubs["rsrcval"] = rsrcstrv[nresource]
htmlsubs["costval"] = coststrv[ncost]
htmlsubs["skedval"] = skedstrv[nskded]
htmlsubs["scopscor"] = str(scopscor[nscope])
htmlsubs["rsrcscor"] = str(rsrcscor[nresource])
htmlsubs["costscor"] = str(costscor[ncost])
htmlsubs["skedscor"] = str(skedscor[nskded])
htmlsubs["scortotl"] = str(nscortotl)
# for i in range(6):htmlsubs["scb"+str(i)]= " "
# for i in range(6):htmlsubs["rsb"+str(i)]= " "
# for i in range(6):htmlsubs["cob"+str(i)]= " "
# for i in range(6):htmlsubs["skb"+str(i)]= " "
# htmlsubs["scb"+str(nscope)]= "X"
# htmlsubs["rsb"+str(nresource)]= "X"
# htmlsubs["cob"+str(ncost)]= "X"
# htmlsubs["skb"+str(nskded)]= "X"

```

```

#
# Update the cookie counters
#
mycookie["scenariocount"] = str(newpagenumber) #page number
displayed
mycookie["scenarionumber"] = str(scenario) #scenario value
displayed
#
# check to see if we're finished with scenarios
#
if count < countmax: # not done yet
#
# paint the scenario and set the cookie
#
print mycookie.output()
#
print "Content-type: text/html\n\n" #
#
filename="../html/scenario.html"
FILE = open(filename,"r")
for line in FILE:
    print line.replace('%"', '%&quot;')%htmlsubs,
FILE.close()
#
else: # over the limit
#
# update the cookie and paint the demographics page
#
mycookie["scenariodone"] = "finished" # mark that we're
done
print mycookie.output() # set the cookie
#
print "Content-type: text/html\n\n" # start the page
#
filename="../html/demographics.html"
FILE = open(filename,"r")
for line in FILE:
    print line,
FILE.close()
#
else:
#
# If we land here it is because something is wrong
# so we will try to recover with the error page.
# Assume that the cookie has the correct page number
# of the form that should be up on the screen so we
# need to cause the page to be repainted without
# saving bad data.
#
# start by backing up cookie and form values to previous page
#
htmlsubs=dict()
countml = count-1

```

```

countm2 = count-2
htmlsubs["counter"] = str(countm1) #page number to be displayed
htmlsubs["scid"] = str(sequence[countm2]) #the scenario value
displayed
#
# reset the cookie
#
mycookie["scenariocount"] = str(countm1) #page number displayed
mycookie["scenarionumber"] = str(sequence[countm2])#scenario
value displayed
print mycookie.output() # set the cookie
#
# paint the error page
#
print "Content-type: text/html\n\n" #
#
filename="../html/error_page.html"
FILE = open(filename,"r")
for line in FILE:
    print line.replace('%','%%')%htmlsubs,
FILE.close()
#
#
# end of openconsent.py
#
#

```

### Demographics.py

```

#!/usr/bin/python
# -*- coding: cp1252 -*-
#
# this is called after the demographics form
#
#
import sys, os, cgi, random, datetime, Cookie
import cPickle as pickle
#
countmax = 50
#
#
# check for cookies and proceed
#
#
cookie_string = os.environ.get('HTTP_COOKIE')
mycookie = Cookie.SimpleCookie()
mycookie.load(cookie_string)
#
#
if not cookie_string:
    # cookies disabled - print error page
    #

```

```

print "Content-type: text/html"
print "\n\n"
#
filename="../html/nocookie.html"
FILE = open(filename,"r")
for line in FILE:
    print line,
FILE.close
#
else:
#
# recover the cookie values
#
sessionid = mycookie["sessionid"].value
scenariocount = mycookie["scenariocount"].value
scenarionumber = mycookie["scenarionumber"].value
scenariosave = mycookie["scenariosave"].value
scenariodone = mycookie["scenariodone"].value
#
# convert scenario strings to numbers
#
count = int(scenariocount)    # number of scenarios completed
according to the cookie
    if count > countmax : count = countmax    # don't allow the counter
to overflow
#
# recover the current scenario number from the pickle
#
sequence=pickle.loads(scenariosave)
scenario = sequence[count]
#
# pick off the demographic form values
#
form = cgi.FieldStorage()
usage = form.getfirst("usage", "novalue")
pmyears = form.getfirst("pmyears", "novalue")
researchyears = form.getfirst("researchyears", "novalue")
educationlevel = form.getfirst("educationlevel", "novalue")
profbkgnd = form.getfirst("profbkgnd", "novalue")
surveysource = form.getfirst("surveysource", "novalue")
dtemp = datetime.datetime.now()
demotime = dtemp.ctime()
#
record =
"%s,%s,%s,%s,%s,%s,%s,%s\n"%(sessionid,demotime,usage,pmyears,research
years,educationlevel,profbkgnd,surveysource)
#
# save the demographic data to file
#
filename="../data/demographics.txt"
FILE = open(filename,"a")
FILE.write(record)
FILE.close()

```

```

#
# print the closing form
#
print "Content-type: text/html"
print "\n\n"
#
filename="../html/close.html"
FILE = open(filename,"r")
for line in FILE:
    print line,
FILE.close
#
#

```

### Closing.py

```

#!/usr/bin/python
# -*- coding: cp1252 -*-
#
# Collect final questions
# display thankyou page
#
import sys, os, cgi, datetime, Cookie
import urllib
#
#
# get the cookie
#
cookie_string = os.environ.get("HTTP_COOKIE")
mycookie = Cookie.SimpleCookie()
mycookie.load(cookie_string)
#
if not cookie_string:
    #
    # cookies disabled - print error page
    #
    print "Content-type: text/html"
    print "\n\n"
    #
    filename="../html/nocookie.html"
    PAGE = open(filename,"r")
    for line in PAGE:
        print line,
    PAGE.close
    #
else:
    #
    # recover the cookie values
    #
    sessionid = mycookie["sessionid"].value
    scenariocount = mycookie["scariocount"].value
    scenarionumber = mycookie["scenarionumber"].value

```



```

#
#
#
# pick off the closing form values
#
form = cgi.FieldStorage()
factors = form.getfirst("factors", "novalue")
madebetter = form.getfirst("madebetter", "novalue")
suggestions = form.getfirst("suggestions", "novalue")
dtemp = datetime.datetime.now()
closetime = dtemp.ctime()
#
# clean up the closing data in case there is some trash in it
#
#
pfactors=urllib.quote_plus(factors)
pmadebetter=urllib.quote_plus(madebetter)
psuggestions=urllib.quote_plus(suggestions)
#
precord="%s , %s , %s , %s ,
%s\n"%(sessionid,closetime,pfactors,pmadebetter,psuggestions)
#
# save the closing data to file
#
filename="../data/closing.txt"
FILE = open(filename,"a")
FILE.write(precord)
FILE.close()
#
# paint the thank you page
#
print "Content-type: text/html"
print "\n\n"
#
filename="../html/thank_you.html"
FILE = open(filename,"r")
for line in FILE:
    print line,
FILE.close()
#

```

## APPENDIX C: R PROGRAMMING CODE

### Main Program

```
#
# Analysis of Reasonableness Estimates
# William Kight
# Walden University
# November 2009
#
# open the data base and initialize the main tables
# perform some basic stats on the raw data
#

# clean house
rm(list=ls(all=TRUE))

# load libraries
library(AMORE) # neural processing
library(lattice) # graphics
library(RODBC) # database access
library(gdata) # for random sampling
library(gtools) # to permute vectors
library(car) # vif function by John Fox

starttime <- date()
DirPath <-
"C:\\Users\\DPC2\\Documents\\Walden\\Dissertation\\Data\\Analysis\\"
reportfile=paste("C:\\RPlot\\Report",format(Sys.time(), "%j-%H%M"),
"-model.txt", sep="")
logfile=paste("C:\\RPlot\\LogFile",format(Sys.time(), "%j-%H%M"),
".txt", sep="")

# set up analysis parameters

nTrials <- 1000
nTraining <- 250
nTest <- 250
nAccumulate <- 25
neuralconfig <-c(4,8,8,1)
neurallearningrate <- 0.001
neuralmomentum <- 0.9
neuralshows <- 100

# *****
# define a new-plot function to open png files with
# date time and figure number in the filename
openpng <- function(figurenumber,height=480,width=480) {
```

```

        plotfile=paste("C:\\RPlot\\Report",format(Sys.time()), "%j-
%H%M"), "-",
        figurenumber, ".png", sep="")
    png(filename=plotfile,height=height,width=width)
  }
# *****

# open the captured data scores database
capturechannel <- odbcConnect("policycapture")

# initialize the scenario table
# could calculate this table, but easier to just read it in
ScenarioTable <- sqlQuery(capturechannel,"select * from
FixedScenarioValues")

# pickup the raw scores -

Scoresdataframe <- sqlQuery(capturechannel,"SELECT * FROM
MainDataTable")

# get the average score (Ybar), standard deviation, and total across
all entries

Ybar <- mean(Scoresdataframe$Response)
Ysd <- sd(Scoresdataframe$Response)
Ytot <- sum(Scoresdataframe$Response)
Ycount <- length(Scoresdataframe$Response)
Yindex <- seq(1,Ycount,by=1)      # this is used as the index vector

# initialize the main data collection vectors

# mean squared error
FuzzyMSEresults <- NULL
ExemplarMSEresults <- NULL
QuickEstMSEresults <- NULL
MappingMSEresults <- NULL
LinearMSEresults <- NULL
NeuralMSEresults <- NULL

# residuals
FuzzyERRresults <- NULL
ExemplarERRresults <- NULL
QuickEstERRresults <- NULL
MappingERRresults <- NULL
LinearERRresults <- NULL
NeuralERRresults <- NULL

# neural model convergence

```

```

NeuralMerror <- NULL

# create the training and testing dataframes by prepopulating them with
scores
Trainingdataframe <- Scoresdataframe[1:nTraining,]
Testingdataframe <- Scoresdataframe[1:nTest,]

#
# process the selected number of trials to Train/Evaluate
# the models

#
# Begin Trials
#
columns=TRUE
plot.new()
for (trialnumber in 1:nTrials) {      #top of the trial loop

  # create a random index list
  Yindex <- permute(Yindex)

  # Start by picking nTraining records at random to use for training
  for (k in 1:nTraining) {
    scrsindex <- Yindex[k]
    Trainingdataframe[k,] <- Scoresdataframe[scrsindex,]
  }

  # Then pick up nTest records at random for prediction test
  # These are offset by nTraining so as not to test with training
records
  for (k in 1:nTest) {
    scrsindex <- Yindex[k+nTraining]
    Testingdataframe[k,] <- Scoresdataframe[scrsindex,]
  }

  # Fuzzy Model (does not require training)
  # returns FuzzyMSE and FuzzySquaredError
  source(paste(DirPath, "Fuzzy Model.r", sep=""), echo=TRUE)

  # Exemplar Model
  # returns ExemplarMSE and ExemplarSquaredError
  source(paste(DirPath, "Exemplar.r", sep=""), echo=TRUE)

  # QuickEst
  # returns QuickEstMSE and QuickEstSquaredError
  source(paste(DirPath, "QuickEst.r", sep=""), echo=TRUE)

  # Mapping
  # returns MappingMSE and MappingSquaredError
  source(paste(DirPath, "Mapping.r", sep=""), echo=TRUE)

  # Regression

```

```

# returns LinearMSE and LinearSquaredError
source(paste(DirPath, "Regression Model.r", sep=""), echo=TRUE)

# Neural
# returns NeuralMSE and NeuralSquaredError
source(paste(DirPath, "Neural.r", sep=""), echo=TRUE)

# save trial MSE results
FuzzyMSEresults <- c(FuzzyMSEresults, FuzzyMSE)
ExemplarMSEresults <- c(ExemplarMSEresults, ExemplarMSE)
QuickEstMSEresults <- c(QuickEstMSEresults, QuickEstMSE)
MappingMSEresults <- c(MappingMSEresults, MappingMSE)
LinearMSEresults <- c(LinearMSEresults, LinearMSE)
NeuralMSEresults <- c(NeuralMSEresults, NeuralMSE)

# only collect residuals for a few of the runs
if(trialnumber <= nAccumulate) {

  # save trial residuals
  FuzzyERRresults <- c(FuzzyERRresults, FuzzyError)
  ExemplarERRresults <- c(ExemplarERRresults, ExemplarError)
  QuickEstERRresults <- c(QuickEstERRresults, QuickEstError)
  MappingERRresults <- c(MappingERRresults, MappingError)
  LinearERRresults <- c(LinearERRresults, LinearError)
  NeuralERRresults <- c(NeuralERRresults, NeuralError)

  # save the Neural Convergence data
  NeuralMerror <- c(NeuralMerror, netresult$Merror)
}

# try to catch the outlier with the neural model
if ((max(NeuralError) > 1.0) | (NeuralMSE > 0.1)){

  logrecord=data.frame(trial=trialnumber,
                      residual=NeuralError,

  testrecord=Testingdataframe$RecordNumber,

  trngrecord=Trainingdataframe$RecordNumber)
  write.table(logrecord, file=logfile, append=TRUE, sep=" ",
             col.names=columns, row.names=FALSE)

  columns=FALSE # only print the column names once

}

# display the iteration number
plot.new()
text(0, 0.8, labels=trialnumber, pos=4)
print(trialnumber)
} ## end of trial loop

```

```

# now analyze the results of the trials

# short names used in the dataframe to facilitate plotting Tukey

MSEresults <-data.frame(Fuzzy=FuzzyMSEresults,
                       Exemplar=ExemplarMSEresults,
                       QuickEst=QuickEstMSEresults,
                       Mapping=MappingMSEresults,
                       Linear=LinearMSEresults,
                       Neural=NeuralMSEresults)

MSEstack <- stack(MSEresults)

# run a box plot of the MSE results
# first set up the plotting vectors
parameterstring = paste("nTrials:",trialnumber,
                        "/" nTraining:",nTraining,
                        "/" nTest:",nTest,
                        "/" ",starttime, sep=" ")

neuralstring <- paste("Neural config:", neuralconfig[1],
                    "/" ",neuralconfig[2],
                    "/" ",neuralconfig[3],
                    "/" ",neuralconfig[4],
                    " learning rate:",neurallearningrate,
                    " momentum:",neuralmomentum,
                    " shows:",neuralshows)

openpng(3)
plot.new()
bwplot( ind ~ values, data=MSEstack,
        horozontil=TRUE,
        main="MSE Over Multiple Random Trials",
        xlab="Mean Squared Error",
        notch=FALSE
        )
# display the trial parameters
mtext(parameterstring,side=3,line=2)
dev.off()

# Run an TukeyHSD to see if the means are different

MSEanova<-aov(values~ind,data=MSEstack)
TukeyResults<-TukeyHSD(MSEanova,"ind",ordered=TRUE)

TukeyResults          #prints out the Tukey results
openpng(5)           #plot to png file
plot(TukeyResults)   #need to work on the graphic
dev.off()

# run some histograms of the residuals
ErrorVector <- c(FuzzyERRresults,

```

```

        ExemplarERRresults,
        QuickEstERRresults,
        MappingERRresults,
        LinearERRresults,
        NeuralERRresults)

ErrorType <- factor(rep(c("Fuzzy", "Exemplar", "QuickEst", "Mapping",
                        "Linear", "Neural"),
                      each=length(FuzzyERRresults)))

ErrorData=data.frame(Error=ErrorVector, Type=ErrorType)

openpng(4, width=480, height=720)
histogram(~Error|Type, data=ErrorData, layout=c(1, 6),
          main="Histograms of Residuals vs. Model Type",
          xlim=c(-1, 1),
          nint=100)

dev.off()

# run the normal probability plots
openpng("5-fuzzy")
qqnorm(FuzzyERRresults, main="Fuzzy Normal Probability Plot")
qqline(FuzzyERRresults)
dev.off()
openpng("5-exemplar")
qqnorm(ExemplarERRresults, main="Exemplar Normal Probability Plot")
qqline(ExemplarERRresults)
dev.off()
openpng("5-quickest")
qqnorm(QuickEstERRresults, main="QuickEst Normal Probability Plot")
qqline(QuickEstERRresults)
dev.off()
openpng("5-mapping")
qqnorm(MappingERRresults, main="Mapping Normal Probability Plot")
qqline(MappingERRresults)
dev.off()
openpng("5-linear")
qqnorm(LinearERRresults, main="Linear Normal Probability Plot")
qqline(LinearERRresults)
dev.off()
openpng("5-neural")
qqnorm(NeuralERRresults, main="Neural Normal Probability Plot")
qqline(NeuralERRresults)
dev.off()

# take a look at the neural convergence properties

openpng(6)
tlength<-length(NeuralMerror)/neuralshows
plot(NeuralMerror[1:neuralshows], type="l",
     ylim=c(0, 2.5), main="Neural Convergence",
     xlab="iteration", ylab="error")

```

```

for (k in 2:tlength) {
  indxlow <- k*neuralshows+1
  indxhigh <- indxlow+neuralshows-1
  lines(NeuralMerror[indxlow:indxhigh])
}
dev.off()

#
# write model data to a file
#

sink(reportfile, append=FALSE, split=FALSE)

c("Summary Model Report - ",reportfile)
"starttime"
starttime
"stoptime"
stoptime <- date()
stoptime
"      "
parameterstring
"      "
neuralstring
"      "
"number of successful trials"
trialnumber

# Trial MSE results"
"      "
"Fuzzy"
summary(FuzzyMSEresults)
"      "
"Exemplar"
summary(ExemplarMSEresults)
"      "
"QuickEst"
summary(QuickEstMSEresults)
"      "
"Mapping"
summary(MappingMSEresults)
"      "
"Linear"
summary(LinearMSEresults)
"      "
"Neural"
summary(NeuralMSEresults)

# Standard Deviations
paste("FuzzyMSE std.dev. : ",sd(FuzzyMSEresults))
paste("ExemplarMSE std.dev. : ",sd(ExemplarMSEresults))
paste("QuickEstMSE std.dev. : ",sd(QuickEstMSEresults))
paste("MappingMSE std.dev. : ",sd(MappingMSEresults))
paste("LinearMSE std.dev. : ",sd(LinearMSEresults))

```



```

paste("NeuralMSE std.dev. : ",sd(NeuralMSEresults))

# Tukey difference test

TukeyResults
sink()

stoptime
#
# done for now
#

```

### Exemplar Model

```

#
# Exemplar Model
#

# define the attenuation factors
dweights<-c(1.00,0.50,0.25,0.10,0.01)

exemplarscore <- function(scenarionumber) {

  tableindex <- scenarionumber+1
  scopeindex <- ScenarioTable$Scope[tableindex]
  resourceindex <- ScenarioTable$Resource[tableindex]
  costindex <-ScenarioTable$Cost[tableindex]
  scheduleindex <- ScenarioTable$Schedule[tableindex]

  # calculate the distance weights against the exemplars
  ntraining <- length(Trainingdataframe$ScenarioID)

  wproduct<-matrix(nrow=ntraining,ncol=1)
  wscore<-matrix(nrow=ntraining,ncol=1)

  for (itrain in 1:ntraining) {
    wscope <- dweights[abs(scopeindex -
Trainingdataframe$Scope[itrain])+1]
    wresource <- dweights[abs(resourceindex -
Trainingdataframe$Resource[itrain])+1]
    wcost <- dweights[abs(costindex -
Trainingdataframe$Cost[itrain])+1]
    wschedule <- dweights[abs(scheduleindex -
Trainingdataframe$Schedule[itrain])+1]

    wproduct[itrain] <-wscope*wresource*wcost*wschedule
    wscore[itrain]<-
wproduct[itrain]*Trainingdataframe$Response[itrain]
  }
}

```

```

productsum <- sum(wproduct)
scoresum <- sum(wscore)

exemplarscore <- (scoresum/productsum)/7      #normalize the output

}

# tally up the residuals for the trial
# uses the Testingdataframe records selected in Trials.r

# Normalize the test data

NormalResponse <- Testingdataframe$Response/7

# Calculate the predicted response
ExemplarPredictedResponse <-
  apply(Testingdataframe$ScenarioID,exemplarscore)

# determine the squared deviation
ExemplarError <- (NormalResponse - ExemplarPredictedResponse)
ExemplarSquaredError <- (ExemplarError)^2

# calculate the mean squared error for the trial
ExemplarMSE <- mean(ExemplarSquaredError)

#
# done for now
#

```

## Fuzzy Model

```

#
# fuzzy model
#

# define the fuzzy membership functions

fuzzyscope      = c(0.25,1.00,0.75,0.50,0.00)
fuzzyresource   = c(0.00,0.25,0.50,0.75,1.00)
fuzzycost       = c(0.25,0.50,1.00,0.50,0.25)
fuzzyschedule   = c(0.00,0.50,1.00,0.50,0.25)

# uses the ScenarioTable imported by OpenDataBase.r
# function returns the normalized fuzzy score given scenario ID
# output is normalized to 1.0

fuzzyscore <- function(scenarionumber){
  tableindex <- scenarionumber+1
  scopeindex <- ScenarioTable$scope[tableindex]
  resourceindex <- ScenarioTable$Resource[tableindex]
  costindex <- ScenarioTable$Cost[tableindex]

```

```

scheduleindex <- ScenarioTable$Schedule[tableindex]
# calculate the fuzzy score from the membership functions
fuzzyscore <- fuzzyscope[scopeindex]+
              fuzzyresource[resourceindex]+
              fuzzycost[costindex]+
              fuzzyschedule[scheduleindex]
fuzzyscore <- fuzzyscore/4      #normalized to 1.0
}

# tally up the residuals for the trial
# uses the Testingdataframe records selected in Trials.r

# Normalize the test data

NormalResponse <- Testingdataframe$Response/7

# Calculate the predicted response
FuzzyPredictedResponse <-
sapply(Testingdataframe$ScenarioID, fuzzyscore)

# determine the squared deviation
FuzzyError <- NormalResponse - FuzzyPredictedResponse
FuzzySquaredError <- (FuzzyError)^2

# calculate the mean squared error for the trial
FuzzyMSE <- mean(FuzzySquaredError)

#
# done for now
#

```

## Mapping Model

```

#
# Mapping Model
#

# set up the function to convert to dichotomous cues
# needs the ScenarioTable from OpenDataBase.r

Bscope <- c(0,1,1,1,0)
Bresource <- c(0,0,1,1,1)
Bcost <- matrix(nrow=5,ncol=5)
Bcost[1,] <- c(0,1,1,0,0)
Bcost[2,] <- c(0,1,1,0,0)
Bcost[3,] <- c(0,0,1,0,0)
Bcost[4,] <- c(0,0,1,1,0)
Bcost[5,] <- c(0,0,1,1,0)
Bschedule <- Bcost

cuetotal <- function(scenarionumber) {
  tableindex <- scenarionumber+1

```

```

scopeindex <- ScenarioTable$Scope[tableindex]
resourceindex <- ScenarioTable$Resource[tableindex]
costindex <- ScenarioTable$Cost[tableindex]
scheduleindex <- ScenarioTable$Schedule[tableindex]

Q1 <- Bscope[scopeindex]
Q2 <- Bresource[resourceindex]
Q3 <- Bcost[scopeindex, costindex]
Q4 <- Bschedule[scopeindex, scheduleindex]

cuetotal <- Q1+Q2+Q3+Q4
}

# calculate the training values based upon Trainingdataframe

TrainingCueTotals <- sapply(Trainingdataframe$ScenarioID, cuetotal)
m0 <- mean(Trainingdataframe$Response[TrainingCueTotals==0])
m1 <- mean(Trainingdataframe$Response[TrainingCueTotals==1])
m2 <- mean(Trainingdataframe$Response[TrainingCueTotals==2])
m3 <- mean(Trainingdataframe$Response[TrainingCueTotals==3])
m4 <- mean(Trainingdataframe$Response[TrainingCueTotals==4])

#
# need to figure out how to handle a mean with no data
# for now just pull another random sample
#

MappingMeans <- c(m0, m1, m2, m3, m4)

mappingscore <- function(scenarionumber) {
  tableindex <- scenarionumber+1
  scopeindex <- ScenarioTable$Scope[tableindex]
  resourceindex <- ScenarioTable$Resource[tableindex]
  costindex <- ScenarioTable$Cost[tableindex]
  scheduleindex <- ScenarioTable$Schedule[tableindex]

  Q1 <- Bscope[scopeindex]
  Q2 <- Bresource[resourceindex]
  Q3 <- Bcost[scopeindex, costindex]
  Q4 <- Bschedule[scopeindex, scheduleindex]

  cueindex <- Q1+Q2+Q3+Q4+1

  # normalize the output
  mappingscore <- MappingMeans[cueindex]/7
}

# tally up the residuals for the trial
# uses the Testingdataframe records selected in Trials.r

# Normalize the test data

NormalResponse <- Testingdataframe$Response/7

```

```

# Calculate the predicted response
#
MappingPredictedResponse <-
sapply(Testingdataframe$ScenarioID,mappingscore)

# determine the squared deviation
MappingError <- NormalResponse - MappingPredictedResponse
MappingSquaredError <- (MappingError)^2

# calculate the mean squared error for the trial
MappingMSE <- mean(MappingSquaredError)

#
# done for now
#

```

## Neural Model

```

#
# Runs the Neural Model (Neural1 - Amore)
library(AMORE) # neural processing
# uses the training data set loaded in Trials.r

# define the neural network
net.start <-
newff(n.neurons=neuralconfig,learning.rate.global=neuralllearningrate,
      momentum.global=neuralmomentum,
error.criterion="LMS", Stao=NA,
      hidden.layer="sigmoid", output.layer="purelin",
method="ADAPTgdwm")

# train the neural model
xv1 <- as.numeric(Trainingdataframe[,"Scope"])
xv2 <- as.numeric(Trainingdataframe[,"Resource"])
xv3 <- as.numeric(Trainingdataframe[,"Cost"])
xv4 <- as.numeric(Trainingdataframe[,"Schedule"])
xtrain <- cbind(xv1,xv2,xv3,xv4)

ytrain <- as.numeric(Trainingdataframe[,"Response"])

netresult <- train(net.start, xtrain, ytrain, error.criterion="LMS",
                 report=TRUE, show.step=100, n.shows=neuralshows )

# calculate test values
xt1 <- as.numeric(Testingdataframe[,"Scope"])
xt2 <- as.numeric(Testingdataframe[,"Resource"])
xt3 <- as.numeric(Testingdataframe[,"Cost"])
xt4 <- as.numeric(Testingdataframe[,"Schedule"])
xtest <- cbind(xt1,xt2,xt3,xt4)

ytest <- sim.MLPnet(netresult$net,xtest)

```

```

# tally up the residuals for the trial
# uses the Testingdataframe records selected in Trials.r

# Normalize the test data
NormalResponse <- Testingdataframe$Response/7

# Calculate the predicted response (normalize mlp output)
NeuralPredictedResponse <- ytest/7

# determine the squared deviation
NeuralError <- NormalResponse - NeuralPredictedResponse
NeuralSquaredError <- (NeuralError)^2

# calculate the mean squared error for the trial
NeuralMSE <- mean(NeuralSquaredError)

#
# done for now
#

```

## QuickEst Model

```

#
# QuickEst Model
#

# set up the function to convert to dichotomous cues
# needs the ScenarioTable from OpenDataBase.r

Bscope <- c(0,1,1,1,0)
Bresource <- c(0,0,1,1,1)
Bcost <- matrix(nrow=5,ncol=5)
Bcost[1,] <- c(0,1,1,0,0)
Bcost[2,] <- c(0,1,1,0,0)
Bcost[3,] <- c(0,0,1,0,0)
Bcost[4,] <- c(0,0,1,1,0)
Bcost[5,] <- c(0,0,1,1,0)
Bschedule <- Bcost

getcues <- function(scenarionumber) {
  tableindex <- scenarionumber+1
  scopeindex <- ScenarioTable$Scope[tableindex]
  resourceindex <- ScenarioTable$Resource[tableindex]
  costindex <- ScenarioTable$Cost[tableindex]
  scheduleindex <- ScenarioTable$Schedule[tableindex]

  Q1 <- Bscope[scopeindex]
  Q2 <- Bresource[resourceindex]
  Q3 <- Bcost[scopeindex,costindex]
  Q4 <- Bschedule[scopeindex,scheduleindex]

 getcues <- c(Q1,Q2,Q3,Q4)
}

```

```

# convert to dichotomous cues
Qvector <- t(sapply(Trainingdataframe$ScenarioID, getcues))

# calculate the nil mean and positive mean values

scopenilmean <- mean(Trainingdataframe$Response[Qvector[,1]==0])
scopeposmean <- mean(Trainingdataframe$Response[Qvector[,1]==1])

resourcenilmean <- mean(Trainingdataframe$Response[Qvector[,2]==0])
resourceposmean <- mean(Trainingdataframe$Response[Qvector[,2]==1])

costnilmean <- mean(Trainingdataframe$Response[Qvector[,3]==0])
costposmean <- mean(Trainingdataframe$Response[Qvector[,3]==1])

schedulenilmean <- mean(Trainingdataframe$Response[Qvector[,4]==0])
scheduleposmean <- mean(Trainingdataframe$Response[Qvector[,4]==1])

# determine the rank ordering of the nil means
nilmeans <- c(scopenilmean, resourcenilmean, costnilmean, schedulenilmean)
posmeans <- c(scopeposmean, resourceposmean, costposmean, scheduleposmean)

rankordermeans <- rank(nilmeans, ties.method="first")

qvalues <- c(0,0,0,0,0) # set up a vector
qvalues[1]<-nilmeans[rankordermeans==1]
qvalues[2]<-nilmeans[rankordermeans==2]
qvalues[3]<-nilmeans[rankordermeans==3]
qvalues[4]<-nilmeans[rankordermeans==4]
qvalues[5]<-posmeans[rankordermeans==4]

# function to determine the rank

quickscore <- function(scenarionumber){
  cues <- getcues(scenarionumber)
  testcues <- c(0,0,0,0)
  testcues[1]<-cues[rankordermeans==1]
  testcues[2]<-cues[rankordermeans==2]
  testcues[3]<-cues[rankordermeans==3]
  testcues[4]<-cues[rankordermeans==4]

  #progress through the cues in increasing nil mean order
  #look for the first negative "0" cue value
  if(testcues[1]==0) {
    quickscore <- qvalues[1]
  } else {
    if(testcues[2]==0) {
      quickscore <- qvalues[2]
    } else {
      if (testcues[3]==0) {
        quickscore <- qvalues[3]
      } else {
        if (testcues[4]==0) {

```

```

        quickscore <- qvalues[4]
      } else {
        quickscore <- qvalues[5]
      }
    }
  }

  # now normalize the result
  quickscore <- quickscore/7
}

# tally up the residuals for the trial
# uses the Testingdataframe records selected in Trials.r

# Normalize the test data

NormalResponse <- Testingdataframe$Response/7

# Calculate the predicted response
QuickEstPredictedResponse <-
sapply(Testingdataframe$ScenarioID, quickscore)

# determine the squared deviation
QuickEstError <- NormalResponse - QuickEstPredictedResponse
QuickEstSquaredError <- (QuickEstError)^2

# calculate the mean squared error for the trial
QuickEstMSE <- mean(QuickEstSquaredError)

#
# done for now
#

```

## Regression Model

```

#
# simple linear regression model
#

# uses the random training set opened in Trials.r

LinearModel <- lm( Response ~ Scope+Resource+Cost+Schedule,
                 data=Trainingdataframe )

# predict values using the Test data
# can't get predict.lm to work so we have to brute force it
# output is normalized to 1.0

linearscore <- function(scenarionumber){
  tableindex <- scenarionumber+1
  X1 <- ScenarioTable$Scope[tableindex]

```



```
X2 <- ScenarioTable$Resource[tableindex]
X3 <- ScenarioTable$Cost[tableindex]
X4 <- ScenarioTable$Schedule[tableindex]

B0 <- LinearModel$coefficients[1]
B1 <- LinearModel$coefficients[2]
B2 <- LinearModel$coefficients[3]
B3 <- LinearModel$coefficients[4]
B4 <- LinearModel$coefficients[5]

y <- B0 + B1*X1 + B2*X2 + B3*X3 + B4*X4

linearscore <- y / 7
}

# tally up the residuals for the trial
# uses the Testingdataframe records selected in Trials.r

# Normalize the test data

NormalResponse <- Testingdataframe$Response/7

# Calculate the predicted response
LinearPredictedResponse <-
sapply(Testingdataframe$ScenarioID,linearscore)

# determine the squared deviation
LinearError <- NormalResponse - LinearPredictedResponse
LinearSquaredError <- (LinearError)^2

# calculate the mean squared error for the trial
LinearMSE <- mean(LinearSquaredError)


#
# done for now
#
```

## APPENDIX D: COPYRIGHT PERMISSIONS

### APA Publications

American Psychological Association (APA) copyrighted materials are reprinted or adapted in accordance with part 3 of the APA guidelines shown in Figure D1:

4/10/2009 APA Copyright and Permissions Informa...

 **copyright & permissions** *American Psychological Association*

[HOME](#) [SITE MAP](#) [CONTACT](#)

Association Info	Students	Psychology Topics	Publications	Careers	Press/News	Join APA
------------------	----------	-------------------	--------------	---------	------------	----------

**APA Permissions Policy**

APA supports the dissemination of information to aid in the development of science and scholarly research. APA also values and respects its own intellectual property as well as the intellectual property of others. As a result, APA believes it is essential for publishers of scholarly and other proprietary material to develop an efficient and consistent system, based on mutual trust, for granting permissions for both electronic and print publication of proprietary works. Therefore, APA adopts the following guidelines for the use of APA copyrighted content.

**1. Permission is Required for the following:**

- A measure, scale, or instrument
- A video
- Full articles or book chapters
- Single text extracts of more than 400 words
- Series of text extracts that total more than 800 words
- More than three figures or tables from any one journal article
- More than three figures or tables from any one book chapter
- Placement of an abstract of a journal article in a database for subsequent redistribution
- Reuse of content from the public APA web site unless there is a copyright notice on that material stating otherwise
- Content essential to the character of the previously published book or article, when reuse could compromise the sale of the APA publication. Examples include complex illustrations, cartoons, maps, works of art, creative photographs.

This list is illustrative rather than exhaustive.

**2. Permissions Not Granted**

APA will not grant permission for use of the *Publication Manual of the American Psychological Association* or any portion of the Manual with the exception of the "Guidelines to Reduce Bias in Language" (pages 61–76).

**3. Permission is Not Required for the following:**

- a maximum of three figures or tables from a journal article or book chapter
- single text extracts of less than 400 words
- series of text extracts that total less than 800 words

No formal requests to APA or the author are required for the items in this clause.

**4. Requirement for Attribution and Credit**

APA requires that full credit be given to the author(s) and APA of the material(s) reused. That credit should include the following:

- For material republished from books: author, title, edition, publisher, city, copyright year, and "reprinted [or adapted] with permission."

For material republished from journal articles: author, title of article, title of journal, volume number, issue number (if relevant), page range (or first page if this is the only information available), date, APA as publisher, and "reprinted [or adapted] with permission."

**5. Permission from Authors**

- When formal APA permission is required to reprint, APA requires that author permission also be obtained. APA does not assist in this effort.
- In the case of multiple authorship, only one author's permission is required.

In most cases, if a good faith effort is demonstrated and the author cannot be contacted, APA will grant permission on its sole authority.

Figure D1. APA copyright and permission guidelines.

## Wiley-Blackwell Publications

Permission to use material from “Portfolio management for new product development: results of an industry practices study” by R. Cooper, S. Edgett, and E. Kleinschmidt, 2001, *R&D Management*, was obtained from the publisher as shown in Figure D2.

**William Kight** [Back to Message View](#) [Send To Printer](#)

**Date :** Tue, Apr 14, 2009 04:22 AM CDT  
**From :** Permission Requests - UK <permissionsuk@wiley.com>  
**To :** William Kight <william.kight@waldenu.edu>  
**Reply To :** Permission Requests - UK <permissionsuk@wiley.com>  
**Subject :** RE: Permission Request for the journal R&D Management

Dear William Kight,

Thank you for your email request. Permission is granted for you to use the material below for your thesis/dissertation subject to the usual acknowledgements and on the understanding that you will reapply for permission if you wish to distribute or publish your thesis/dissertation commercially.

Best wishes,

**Laura Mothersole**  
 Permissions Assistant  
 Wiley-Blackwell  
 9600 Garsington Road  
 Oxford OX4 2DQ  
 UK  
 Tel: +44 (0) 1865 476160  
 Fax: +44 (0) 1865 471158  
 Email: [laura.mothersole@wiley.com](mailto:laura.mothersole@wiley.com)

---

**From:** William Kight [mailto:william.kight@waldenu.edu]  
**Sent:** 12 April 2009 20:33  
**To:** Permission Requests - UK  
**Subject:** Permission Request for the journal R&D Management

To whom it may concern,

With respect to the article:

*Portfolio management for new product development: results of an industry practices study.*

by Robert Cooper, Scott Edgett, and Elko Kleinschmidt,  
 Appearing in the 2001 journal R&D Management, Volume 31 Issue (4), pp. 361-380,

I would like to request permission to incorporate "Figure 8. Typical risk-reward bubble diagram", pg. 369, in my dissertation. I would like to reference the figure as an example of the ways that people represent estimations of research and development projects. My dissertation is tentatively entitled "A mixed methods analysis of estimation best practices for research programs" and is being submitted to Walden University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Applied Management and Decision Sciences.

If you require any additional information, I may be contacted via e-mail at [William.Kight@WaldenU.edu](mailto:William.Kight@WaldenU.edu) or via postal service to the attention of:

William Kight

Figure D2. Permission obtained from Wiley-Blackwell.

## CURRICULUM VITAE

### **William D. Kight**

13885 Kennard Drive

Glenelg, MD 21737

(410) 309-7362

[wkight@ieee.org](mailto:wkight@ieee.org)

### **Education:**

Doctor of Philosophy – Applied Management and Decision Sciences, 2010  
Walden University, Minneapolis, Minnesota

Master of Science – Electrical Engineering, 1984  
University of Maryland, College Park, MD

Bachelor of Science – Physics, 1980  
Rensselaer Polytechnic Institute, Troy, NY

### **Professional Experience:**

**Director, Network Engineering** 1999-Present  
Verizon, Columbia, MD

Oversee the process improvement initiative towards a CMMI Level 3 appraisal. Provide engineering oversight on all Columbia engineering programs. Created a new business area for Columbia using Verizon FiOS Technology to support high performance secure communications within the intelligence community. Provide development support for facilities and infrastructure components for specialized network operations centers. Previously developed a line of products supporting ATM and SONET fiber optic communications and performance monitoring. Also developed products for high-speed network security.

**President** 1997-1999  
NxGen Technologies, Columbia, MD

Responsible for the overall management and operation of a small business centered on development of custom hardware and embedded systems. Product focus was on SONET fiber optic communications and performance monitoring. Developed a high-performance SONET Signal Processor supporting SONET fiber optic communications and performance monitoring for the Department of Defense (in conjunction with Advanced Concepts, Inc.)

**Director, Engineering** 1993-1997  
GRC International, Columbia, MD

Responsible for the overall engineering specification, design and development of an intelligent fiber optic demarcation unit used in SONET based fiber optic networks. Managed a staff of approximately 70 engineers and support staff. Also responsible for

ensuring compliance with all applicable Bellcore and Federal specifications concerning deployment of transmission equipment within the public switched telephone network. Acted as corporate liaison to the SONET Interoperability Forum. Developed and implemented an ISO 9000 based quality program for the engineering group.

**Vice President, Engineering** 1992-1993

PKE, Inc., Columbia, MD

Primary engineering responsibility was in support of design and development engineering including detailed product surveys and technical evaluation of all major SONET wideband digital cross-connect systems including AT&T, Tellabs, Alcatel, and DSC. Also responsible for design, development, integration, installation, and test of a Video analysis system for the U.S.A.F. supporting accurate field synchronized playback of multiple video tape sources for post mission analysis. Required an in-depth understanding of Video control, synchronization, recording, time code, and transmission issues.

**Sr. Staff Engineer** 1986-1992

CSTI, Columbia, MD

Primarily responsible for supporting new business development efforts. Designed a large scale range communications system for the U.S. Air Force using commercial cross-connect systems and related telecommunications products. Consistently won highest score on proposals for technical approach and overall quality of proposals.

**Sr. Engineer** 1984-1986

California Microwave, Inc., Columbia, MD

Responsible for the design, development, and installation of a 13 meter Ku-Band satellite earth station. Required extensive coordination with civil works contractors preparing the site, including road work, electrical services, water, and foundations. Also supported the feed retrofit operation on a 150 Ft. C-Band system.

**Sr. Engineer** 1982-1984

HRB Systems, Inc., Lanham, MD

Responsible for the design, development, integration, installation, and test of a 15 channel receiving system front end consisting of 7 racks of receivers, demodulators, and bit synchronizers with RF distribution subsystem and controllers. Also responsible for the design and coding of the user interface software for the AN/GSQ-206 control console based upon the PDP-11/23. Project was coded in assembly language.

**Research Physicist** 1980-1982

Naval Research Laboratory, Washington, DC

Performed basic research on the performance of large aperture surveillance arrays. Required in-depth understanding of ocean acoustic propagation. Required extensive use of real-time signal processing techniques to process and analyze array data.

**Publications:**

Kight, B., & Swanson, B. (1994). Making the Sonet connection. (Cover story).  
*Telephony*, 227(24), 36.

**Professional Presentations and Papers:**

Kight, W., & Pfister, R. (1993). *Application of Emerging Technologies to the Creation of a Virtual Range*. Paper presented at the International Telemetry Conference.