A DELPHI POLICY STUDY ON THE FUTURES USE
OF DISTANCE EDUCATION
WITHIN THE UNIVERSITY OF HAWAI'I SYSTEM

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
COMMUNICATION AND INFORMATION SCIENCES
AUGUST 2011

By
Patrick Joseph Gilbert

Dissertation Committee:
Raymond Panko, Chairperson
Elizabeth Davidson
Reginald Worthley
Daniel Suthers
Kent W. Bridges

Keywords: Policy Delphi, Distance Learning
Abstract

Information and communications technology is reshaping nearly every societal organization and higher education is not exempted. What kind of change will the interaction bring? The serendipitous, opportunistic, and unguided path argues for a float all boats and see which boats sail. The measured, planned, and managed path argues for resources to be allocated toward organization goals that effectively and efficiently expend scare resources. The determination of those goals is not a linear extrapolation of the present because the future configuration of higher education as reshaped by information and communications technology could be radically new in form. Because the future is involved and what is known today must be balanced with what is unknown in ten to fifteen years, expertise should contribute significantly to shaping goals. Exercises might usually attempt to identify what might be in terms of technology, courses, or structure but this research instead asks what is the desirable state – what is wanted – what should be. As such, this research has a future-focused, policy orientation for which the most appropriate approach appears to be the Policy Delphi. The procedure itself is implemented in software developed for this specific use.

The results are reported and the policy recommendations are constructed. The intervention proved challenging. Extended sections of the results deal with the shortcomings and issues in a web-based Policy Delphi including mortality, participation, use of feedback, and suggestions concerning the literature driven construction of the software.
Dedication and Acknowledgments

First, the chair of the committee – Dr. Raymond Panko – is owed my gratitude for taking on this project. His guidance and counseling have been invaluable. The other members – Dr. Kim Bridges, Dr. Reginald Worthley, Dr. Daniel Suthers, and Dr. Elizabeth Davidson -- of the committee have contributed and given great guidance. They put up with over-zealous digging into sometimes off-tangent details but they encouraged the pursuit of ideas because as one of them said – this work is ultimately something for me and about me and thus should fulfill my need to understand the topic I have selected.

Dr. Murray Turoff, world-renowned Delphi expert and proponent provided key input to the final methodology and the design of the interface. His input provided guidance on some questions concerning the procedure and process of conducting a Policy Delphi.

The Communication and Information Sciences program gave the opportunity to interact with many colleagues who have completed their work and moved on to improved things. Among these are Ravi Vatrapu, Marc Lepape, David Pai, and Susan Johnson.

In a first and earliest attempt at doctorate work, I encountered a number of people in the College of Education who gave a framework in management concepts, finance, law, and decision-making. Professors John A. Thompson and Charles Araki gave early encouragement even though I failed to complete the dissertation. Professor Ann Kepple, the consummate scholar and friend, was always encouraging and provided so many keen insights into the diversity of the human spirit in which she was so richly endowed.

Dean David Heenan and Dean David Bess gave me the unique opportunity to serve with them at the College of Business where I gained a deeper insight into the workings of the collegiate environment. While at the College, Dean Chuck Gee gave me the world’s best opportunities to lecture in the travel and tourism circuit while encouraging me to learn the industry. So strong was the pull of the exposure he provided that I left academia to participate actively in the travel industry. At the same time, Professor Dexter Choy and Alwin Zecha allowed me the pleasure of interacting and teaching with them giving me yet broader exposure. Professor
Choy reminded me many times that if I was serious about the University that I should get the
degree finished and then enjoy life.

The College of Business period also gave me the opportunity to work with Professor
Ralph Sprague and Professor Bill Remus. Ever encouraging and ever helpful, these scholars
provided me the opportunity to dabble in my first published research. Bill pressed me to finish my
larger work but the siren call of the private sector was just too loud.

Many thanks to Raymond Miyashiro for all the years at TransHawai‘ian, Regal Travel,
and Rendezvous Tours. These unique opportunities to manage and create were invaluable in my
growth. Thanks to Manon Janssen of Hewlett-Packard Company for her bringing me into the
large corporate world and allowing me to continue to grow.

To my friends, my colleagues, my students, and so many others I owe thanks. To my
parents Charles and Martha who ever encouraged me in academic and personal pursuits.
Table of Contents

TITLE PAGE ........................................................................................................................................... I

ABSTRACT ............................................................................................................................................... II

DEDICATION .......................................................................................................................................... III

TABLE OF CONTENTS .......................................................................................................................... V

LIST OF TABLES ....................................................................................................................................... XIV

LIST OF FIGURES ...................................................................................................................................... XVI

CHAPTER 1. INTRODUCTION .................................................................................................................. 1

1.1 DISTANCE EDUCATION..................................................................................................................... 4

1.2 TRANSITION AND DELIMITATION .................................................................................................. 6

1.3 IMPORTANCE OF THE STUDY ......................................................................................................... 7

1.3.1 As the Study Adds to Methodology ............................................................................................. 8

1.4 RESEARCHER’S PERSPECTIVE ...................................................................................................... 9

CHAPTER 2. LITERATURE REVIEW ....................................................................................................... 11

2.1 SELECTING A FORESIGHTING METHODOLOGY .......................................................................... 11

2.2 DELPHI METHOD ............................................................................................................................ 14

2.2.1 Delphi Technique History .......................................................................................................... 15

2.2.2 The Delphi Procedure as the Object of Research ......................................................................... 17

2.2.2.1 Sackman Critique .................................................................................................................. 22

2.2.2.2 Sackman Critique in Summary ............................................................................................ 24

2.2.2.3 Response – Delphi as Experimental Research ..................................................................... 25

2.2.2.4 Countering Sackman Directly ............................................................................................... 26

2.2.2.5 Accuracy and Confidence ....................................................................................................... 30

2.2.2.6 Is Delphi Measureable? ........................................................................................................... 32

2.2.2.7 Is Delphi Replicable? .............................................................................................................. 33
CHAPTER 3. RESEARCH METHODOLOGY

3.1 INTERNAL AND EXTERNAL VALIDITY, CREDIBILITY, AND TRANSFERABILITY ............. 121

3.1.1 Credibility and Transferability ........................................................................ 122
3.1.2 Credibility ........................................................................................................ 122
3.1.3 Transferability .................................................................................................. 123
3.1.4 Evidence versus Methods ............................................................................... 124
3.1.5 Maxwell Validity Checklist ............................................................................. 124
3.1.6 Mitigation of Maxwell’s Checklist ................................................................... 125
  3.1.6.1 Intensive, long-term involvement .............................................................. 125
  3.1.6.2 Rich Data .................................................................................................. 126
  3.1.6.3 Respondent Validation: ............................................................................ 126
  3.1.6.4 Discrepant and Negative Cases: ............................................................... 127
  3.1.6.5 Triangulation ............................................................................................ 127
  3.1.6.6 Quasi-statistics ......................................................................................... 128
3.1.7 Mitigation of Validity Issues Specific to Delphi ................................................ 128
  3.1.7.1 Sackman Concerns .................................................................................. 128
  3.1.7.2 Epistemological Claims to Experimental Research .................................. 129
  3.1.7.3 Panel Selection ......................................................................................... 129
  3.1.7.4 Instrument Creation .................................................................................. 130
  3.1.7.5 Statistical Treatment ................................................................................ 131
  3.1.7.6 Efficacy of Group Processes and Alternatives ......................................... 131
  3.1.7.7 Feedback, Forcing Consensus, and Accuracy .......................................... 132
  3.1.7.8 Expert Prediction versus Prediction of Experts ......................................... 133
  3.1.7.9 Containment of Researcher Influence ...................................................... 133
  3.1.7.10 Bias ......................................................................................................... 134
  3.1.7.11 Delphi Bias Sources .............................................................................. 135
  3.1.7.12 Oversight and Review .............................................................................. 136
  3.1.7.13 Remaining Sources of Bias .................................................................... 137
  3.1.7.14 Panel Selection ....................................................................................... 137
3.1.7.15 Motivating Participation ................................................................. 139
3.1.7.16 Participant Identification and Solicitation ........................................ 140
3.1.7.17 Panel Size ....................................................................................... 140
3.1.7.18 Initial Instrument Creation ............................................................... 141
3.1.7.19 Item Construction ........................................................................... 142
3.1.7.20 Item Construction Guidelines ......................................................... 142
3.1.7.21 Initial Seeded Items ........................................................................ 144
3.1.7.22 Questions and Initiated Feed-in ...................................................... 144
3.1.7.23 Rounds, Consensus, and Calendar .................................................. 145
3.1.7.24 Computer Mediated Communication ............................................... 146
3.1.7.25 Attrition ......................................................................................... 146
3.1.7.26 Expertise Survey ............................................................................. 147
3.1.7.27 Results and Write Up ....................................................................... 148

3.2 INFORMING THE DESIGN ................................................................... 148

3.2.1 Fundamentals ...................................................................................... 148
3.2.2 Solicitation – Panel Selection ............................................................... 149
3.2.3 Accepting the Solicitation ................................................................... 149
3.2.4 Human Subjects Consideration ............................................................ 149
3.2.5 Validation of Expertise ......................................................................... 149
3.2.6 Anonymity .......................................................................................... 149
3.2.7 Data Retention ..................................................................................... 150
3.2.8 Calendar .............................................................................................. 150
3.2.9 Voting and Ranking – Display Options ............................................... 150

3.2.10 Feedback Options .............................................................................. 150

3.2.10.1 Notice of Directed Feedback ......................................................... 150
3.2.10.2 Option to read Feedback and its Display ........................................ 151
3.2.10.3 Option to check Background .......................................................... 151
3.2.10.4 Option to ask a question ................................................................. 151
3.2.10.5 Option to challenge Item wording .................................................. 151
5.2.12  Policy 12 – Skill Level and Target Audience .......................................... 243
5.2.13  Policy 13 – IT Infrastructure Expectations ................................................ 244

5.3  WEB-BASED POLICY DELPHI TOOL ......................................................... 245

1.1.3 Participation Challenge ............................................................................. 245
5.3.1  Item List Generation .................................................................................. 247
5.3.2  Construction of a Delphi Web Tool ........................................................... 248
5.3.3  Role of Feedback and Feedback ................................................................. 250
5.3.4  Consensus and Concluding the Delphi ....................................................... 253
5.3.5  Item Attributes and Cumulative Ranking ................................................... 254
5.3.6  Change during the Interaction .................................................................. 256
5.3.7  Requesting Research ................................................................................ 257

5.4  SIGNIFICANCE AND POTENTIAL CONTRIBUTIONS ............................... 258

5.4.1  Significance for Policy .............................................................................. 258
5.4.2  Significance for Practice .......................................................................... 258

5.5  LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH ............. 259

5.5.1  Future Technical web tool Design and Implementation ............................. 259
5.5.2  Richness of Possible Future Research ....................................................... 261

5.6  LESSONS LEARNED .................................................................................... 262

5.6.1  Distance Education .................................................................................... 262
5.6.1.1 Distance Education and Organizational Change ....................................... 263
5.6.1.2 Distance Education Research .................................................................. 265
5.6.2  Policy Delphi .............................................................................................. 266
5.6.2.1 Participant Expectations of Delphi as a Survey ....................................... 267
5.6.2.2 Item List Construction .......................................................................... 267
5.6.2.3 Creating the Initial Item List .................................................................... 268
5.6.2.4 Stakeholders, Experts, and Motivation .................................................... 269
5.6.2.5 Roundless Delphi .................................................................................. 270
5.6.2.6 New Participants, the Swarm, and New Items ........................................ 270
5.6.2.7 Feedback and Decision Making ................................................................. 271
5.6.2.8 Multiple Attributes per Item in Ranking .................................................. 272
5.6.2.9 Demographic and Expertise Survey ......................................................... 273
5.6.2.10 Software Complexity .............................................................................. 274
5.6.2.11 Writing the Conclusions ......................................................................... 274
5.6.3 In Conclusion .............................................................................................. 276

APPENDIX 1 USER COMMENTS .......................................................................... 278
APPENDIX 2 USER ITEM CONTRIBUTION .......................................................... 284
APPENDIX 3 ITEM RANKING .............................................................................. 286
BIBLIOGRAPHY .................................................................................................... 291
List of Tables

Table 1 Frequency of Delphi Dissertations and Theses ................................................. 37
Table 2 Respondent Data Sampling versus Swarm ........................................................... 81
Table 3 Demographic Question 1 .................................................................................. 190
Table 4 Demographic Question 2 .................................................................................. 191
Table 5 Demographic Question 3 .................................................................................. 191
Table 6 Demographic Question 4 .................................................................................. 192
Table 7 Demographic Question 5 .................................................................................. 192
Table 8 Demographic Question 6 .................................................................................. 193
Table 9 Demographic Question 7 .................................................................................. 193
Table 10 Demographic Question 8 ............................................................................... 194
Table 11 Demographic Question 9 ............................................................................... 195
Table 12 Demographic Question 10 ............................................................................. 195
Table 13 Demographic Question 11 ............................................................................. 196
Table 14 Demographic Question 12 ............................................................................. 196
Table 15 Demographic Question 13 ............................................................................. 197
Table 16 Demographic Question 14 ............................................................................. 197
Table 17 Demographic Question 15 ............................................................................. 197
Table 18 Demographic Question 16 ............................................................................. 198
Table 19 Demographic Question 17 ............................................................................. 199
Table 20 Demographic Question 18 ............................................................................. 199
Table 21 Demographic Question 19 ............................................................................. 200
Table 22 Demographic Question 20 ............................................................................. 200
Table 23 Demographic Question 21 ............................................................................. 201
Table 24 Demographic Question 22 ............................................................................. 201
Table 25 Demographic Question 23 ............................................................................. 202
Table 26 Demographic Question 24 ............................................................................. 202
Table 27 Demographic Question 25 ................................................................. 203
Table 28 Illustrative Index Score Calculation .................................................. 207
Table 29 Wilcoxon Signed Rank Results ......................................................... 211
Table 30 Example Item and Rank Scores ......................................................... 212
Table 31 Student Preferred Items ................................................................... 213
Table 32 Faculty Preferred Items ..................................................................... 214
Table 33 Items with Student-Faculty Agreement ............................................ 216
List of Figures

Figure 1 Foresighting Methodology Tree ............................................................... 12
Figure 2 Delphi System Voting Results ................................................................. 104
Figure 3 Delphi System Comments ............................................................... 104
Figure 4 Original Conceptual Design ................................................................. 158
Figure 5 Panel Selection .................................................................................... 159
Figure 6 Original First Pass Sequence ............................................................. 160
Figure 7 Original Detailed Introduction to the Project ...................................... 161
Figure 8 System Flowchart ............................................................................... 162
Figure 9 Original Detailed Tutorial – Page 1 of 3 .............................................. 163
Figure 10 Human Subjects Informed Consent .................................................. 164
Figure 11 Password Screen ............................................................................... 165
Figure 12 Demographic Survey ....................................................................... 166
Figure 13 Nomination (Swarm) for other participants ....................................... 167
Figure 14 Item Contribution .............................................................................. 167
Figure 15 Original Second Pass Sequence ...................................................... 168
Figure 16 Video Replacing Text in Explanation ............................................... 169
Figure 17 Current Status .................................................................................... 169
Figure 18 Sequence Revision including Status Page ......................................... 170
Figure 19 Major Revision 2 Second Pass Sequence .......................................... 170
Figure 20 Summer 2010 Sequence ................................................................... 170
Figure 21 Revised Introduction to the Project .................................................... 171
Figure 22 Summer 2010 Full Final Sequence .................................................. 172
Figure 23 Summer 2010 Final Second Pass Sequence .................................... 172
Figure 24 Main Page Conceptual Diagram ....................................................... 173
Figure 25 Screen Shot of Ranking Page ........................................................... 174
Figure 26 Explanation of Ranking Page Major Area ......................................... 174
Figure 27 Items Displayed As List ................................................................. 176
Figure 28 Item Expanded in Accordion Control ........................................... 176
Figure 29 Screen Shot of Ranking Items ....................................................... 177
Figure 30 Screen Shot of Ranking Results .................................................... 178
Figure 31 Item Context Action Menu ............................................................. 179
Figure 32 Item Context Action Menu- Detailed View ................................. 179
Figure 33 Entering A Comment on an Item ................................................... 180
Figure 34 Reviewing Comments .................................................................. 181
Figure 35 Individual Detailed Demographic Responses ............................... 181
Figure 36 Adding a research request .............................................................. 182
Figure 37 Ranking Page Access to Previous Pages ....................................... 182
Figure 38 Example Attribute and Scale ......................................................... 206
Figure 39 Distribution of Ranking Differences ........................................... 220
Figure 40 Delphi Voting .............................................................................. 252
Figure 41 Summary Delphi Voting ............................................................... 252
Chapter 1. Introduction

Lord Crowther, the first chancellor of the United Kingdom’s Open University innovative, extensive experiment in higher education for the masses that uses an anywhere-anytime model, noted in 1969 that an information and communications technology (hereinafter ICT) revolution is underway with unknown future changes to higher education. Crowther’s prediction predates the massive change brought about by the near ubiquitous digital revolution just parts of which are the Internet and World Wide Web. ICT brings with it potentially significant changes for higher education. He argued that this revolution would cause us to rethink many traditional organizational structures. Crowther sees a revolution in the interaction between education and information and communications’ technology. Issroff and Scanlon quote Crowther as stating,

The world is caught in a communications revolution, the effects of which will go beyond those of the industrial revolution of two centuries ago. Then the great advance was the invention of machines to multiply the potency of men’s muscles. Now the great new advance is the invention of machines to multiply the potency of men’s minds. As the steam engine was to the first revolution, so the computer is to the second. It has been said that the addiction of the traditional university to the lecture room is a sign of its inability to adjust to the development of the printing press. That of course is unjust. But at least no such reproach will be leveled at the Open University in the communications revolution. Every new form of human communication will be examined to see how it can be used to raise and broaden the level of human understanding. There is no restriction on techniques. (as cited in Issroff & Scanlon, 2002, p. 2)

Peter Drucker, famed management author and lecturer on business processes, claims some extreme positions. First, Drucker predicts, “Traditional universities as we know them will become a big wasteland in the next 25 years.” (Dunn, 2001, p. 2) Second, rather than look at institutional change from new ways to deliver higher education to meet an ever growing demand, Drucker argues that financial constraints will compel restructuring that will go to the heart of what is today’s higher education. Lenzer and Johnson cite Drucker who state,

Thirty years from now, the big university campuses will be relics. Universities won’t survive. It’s as large a change as when we first got the printed book. Do you realize that the cost of higher education has risen as fast as the cost of health care? Such totally uncontrollable expenditures, without any visible improvement in either the content or the quality of education, means that the system is rapidly becoming untenable. Higher education is in deep crisis... Already we are beginning to deliver more lectures and classes off campus via satellite or two-way video at a fraction of the cost. The college won’t survive as a residential institution. (as cited in Lenzer & Johnson, 1997, p. para. 70)
Murray Turoff wonders whether tenure, believed by many to be instrumental to a quality university education, will exist as technology’s onslaught rips deeper into the heart of higher education. Like Drucker, he fears the increasing cost of higher as evidenced by the “actual income distribution in this country getting worse.” (Anonymous, 2008, para. 3) Education becomes affordable only to the rich. Those who cannot afford the more traditional higher education will increasingly find themselves in large, computer-mediated, distance education courses with pre-packaged content. Or worse, students will turn to “web based learning delivery resulting in numerous diploma mills that will make people think they are getting an education for low cost.” (Anonymous, 2008, para. 3)

Meyer sees change compelled by declining revenues to higher education, the recent rounds of tuition increases, and society’s pleading to higher education for “more efficiency, more productivity, more graduates, and more learning.” (K. Meyer, 2010, para. 1) Meyer cites Christensen’s work on disruptive technologies as “technological innovations, products, services, processes, or concepts that disrupt the status quo” (K. Meyer, 2010, para. 3) that initially are in the fringe or niche market. This technology “eventually exceeds the performance of prior products and improves to the extent that it satisfies the mainstream market. Firms that support the disruptive technology displace incumbent firms that supported the prior technology.” (K. Meyer, 2010, para. 3) Meyer eventually concludes that information technology will similarly transform higher education.

Tapscott and Williams go further to assert that the “transformation of the university is not just a good idea. It is an imperative, and evidence is mounting that the consequences of further delay may be dire.” (Tapscott & Williams, 2010, para. 1) They assert that “universities are losing their grip on higher learning as the Internet is, inexorably, becoming the dominant infrastructure for knowledge – both as a container and as a global platform for knowledge exchange – and as a new generation of students requires a very different model of higher education.” (Tapscott & Williams, 2010, in para. 1) They claim the elimination of campuses; higher education free to all; tenure’s elimination; and, departments morphing into “problem-focused disciplines.” Oblinger, editor of the journal in which the Tapscott and Williams article appears, concludes,
The economic downturn has caused many college and university leaders to question — even more frequently than usual — what the future of their institutions will look like. But the future challenges for higher education are not just financial. Information technology has catalyzed society-wide changes. Rethinking the future of higher education must take the innovations of the digital age into consideration. (Oblinger, 2010, para. 1)

Rashef takes the themes of Oblinger to argue that ICT “In an age of dropping technology costs, unprecedented access to the Internet, and the free exchange of information online” (Reshef, 2010, para. 1) must obliterate the have-have-nots division along social, financial, and cultural lines. He argues that the low “percentage of university-aged young people enrolled in postsecondary education doesn’t even reach double figures — a blatant sign of a failed education system.” (Reshef, 2010, para. 1)

Reshef not only writes about the potential changes but he has also worked to establish the University of the People with this mission,

University of the People (UoPeople) is the world’s first tuition free online university dedicated to the global advancement and democratization of higher education. This tuition free university embraces the worldwide presence of the Internet and dropping technology costs to bring tuition free and university-level studies within reach to millions of people across the world. (Anonymous, circa 2010)

The free university concept is not new but the concept combined with ICT changes the dynamic potentially altering many assumptions about the structure of higher education.

Change will come. Will the future for higher education institutions arrive serendipitously and haphazardly only having evolved with little guidance? Will the future arrive influenced by humankind who proactively attempted to shape that future?

This work steps into the maelstrom looking not so much to solve the problems raised but rather to look at a futures foresighting methodology. There is the potential to shed some light on the ongoing and evolving dynamic interaction of higher education and information and communications technology.
1.1 Distance Education

The future’s interaction of education with information and communications technology (ICT) is the focus of this dissertation. Past and current interaction weigh less importantly except as they provide a backdrop from which futures emerge. Instead of intense analysis of the past, this dissertation is a policy futures foresighting exercise. Many foresighting exercises attempt to predict dates when events might take place. In contrast, this exercise asks what thought leaders want results to be thus this exercise becomes a policy investigation. This dissertation charts strategic options – the policies of the here and now that might favor one futures over another.

The intersection of education and ICT in this dissertation is distance education. Many varied definitions of distance education exist due to differing motivations and sources.

Keegan recognized the difficulty in defining the term distance education in his 1980 article. Keegan wrote, "the forms of education that fall within the concept of distance education as outlined are considered from the point of view of choice of medium, institutional type and didactic model." He argued that the term distance education is a “satisfactory solution to the problem of terminology.” (Keegan, 1980).

As the University of Hawai’i is the target of this study, the Regent’s definition is included as,

*Distance learning* occurs when students are physically separated from their instructor but connected by technology. Students in multiple locations interact with the instructor and their fellow classmates using technology.

Distance learning refers to educational programs and activities designed to serve clientele other than regular on-campus students and/or deliver such courses and programs in ways and/or at times and locations suited to the needs and convenience of those to be served. Distance learning includes telecommunications-assisted and/or off-site instruction. Distance learning may involve credit courses at the graduate, post-baccalaureate, and undergraduate levels, and/or non-credit, professional, and community service courses. (University of Hawaii Board of Regents, 2002)

---

1 In this dissertation, the word “future” generally appears in a pluralized form. Correct English grammar might dictate the use of the singular form “future”. The use of the plural is a choice of emphasis – a choice to remind readers that there is no “a future” or “single future” or “the future”. The futures is always “plural” -- as in many futures, many alternatives, and many themes. Use of the singular implies too much certainty describing the futures as a single entity.
No matter which of the definitions selected from the literature, the definitions share the following general tenets in varying degrees of emphasis.

1. The teacher and the learner may be physically co-located or separated (location aligned) for some or all of the time of formal instruction. That is, there is no absolute requirement that there be significant separation at all times.

2. The teacher and learner can be synchronous or asynchronous (time aligned) for some or all of the time of formal instruction.

3. The delivering institution can have a physical presence with classroom, dorms, libraries, or be virtual for some or all time of formal instruction. There is no sense that all education should be absent the traditional physical plant of education—a school or campus.

4. A medium is involved for the transmission of the teaching and learning. The media is generally one from ICT. The nature of the ICT varies significantly in sophistication.

The potential interaction of education and ICT is not passive and many see significant change motivated by the intersection or as a response to external forces. Wilson makes very strong claims about distance education and the change he anticipates,

Distance education has been described as a disruptive technology—an innovation that, while initially posing no threat to established institutions, over time challenges conventional practices and contributes to new ways of thinking (Archer, Garrison, & Anderson, 1999). I agree with this assessment. Distance education has already become more than an alternative form of delivery. It has shaken up the educational establishment, especially at higher education and corporate levels. Precisely because of this success, though, assessing potential at midstream can be a challenge. (B. Wilson, 2001)

Cannell claims that budgetary constraints will compel increased use of distance education,

Distance education, and its accompanying technology, is attractive to higher education because it seems to address the challenges of declining enrollments, increasing costs, the potential market of adult professionals, pressure from corporations, institutional competition for faculty or increased sharing of faculty, and increasing global access to technology [emphasis added] (Cannell, 1999)

The economic argument aside, other advocates claim that the technology increases effectiveness while others argue for a more measured assessment. Ringstaff writes,

Debates aside, there is a substantial body of research that suggests that technology can have a positive effect on student achievement under certain circumstances and when used for certain purposes. However, there is no magic formula that educators and policymakers can use to determine if this “return” is actually worth the “investment.” Perhaps, rather than asking, “Is technology worth the cost?” the more important question is, “Under what conditions does technology have the most benefits for students?” The research presented in this paper seeks to answer this question, and offers some suggestions — related to
issues such as teacher training, access to technology, and long-term planning — that policymakers should seriously consider as they seek to enhance student learning through technology use. (Ringstaff & Kelley, 2002, p. 23)

Common sense might lead to the assumption that with all the sophistication in social science research today and with billions of dollars expended that ICT and education interactions that a list of effective best practices would exist. However, as Ringstaff and Kelley state, “measuring the impact of technology use on student achievement is fraught with difficulties.” (Ringstaff & Kelley, 2002, p. 23). ICT infatuates with implied promises and potentialities but “we have learned from prior experiences in the field of educational technology: There is no one best medium; the medium is the means, not the end; and the medium is not the message.” [emphasis added] (Earle, 2002, para. 1). Earle cautions us that “technology enthusiasts continue to forget a basic fact: Machines are tools, valuable only when a human intelligence organizes their use in a productive way.” (Earle, 2002, para. 1).

For this very reason — that technology is but a tool — does this dissertation claim value in its foresighting exercise to outline a desired future state of distance education.

1.2 Transition and Delimitation

The interaction of higher education and ICT – like all interactions of any organization and technology – is complex, multifaceted, dynamic, and hard to reduce to simple statements. ICT’s interaction with education compels change just as ICT’s interaction has done to other organizations. The nature and rapidity of the change is yet to be determined. Educational institutions might evolve. Education institutions might face revolutionary change. Education will not disappear though it may be structured and delivered differently.

Each independent subject – higher education or ICT -- encompass topics far too broad for a single dissertation. The vector cross product of these topics creates a yet larger topic of study. Constraints are required. This dissertation constrains itself to a future state of pedagogy of distance education. Even the limitation to distance education is far too broad a topic for a dissertation. More limitations are required.

This dissertation is not concerned with the past evolution of distance education. This dissertation is not concerned with the details of the efficacy of distance education. As will be
repeated here and elsewhere, this work is not an evaluation of distance education even given that in the literature review and as part of the Delphi item list evaluative statements appear. This dissertation is not a media comparison study contrasting alternative delivery of ICT delivered instruction with the classic, traditional, face-to-face classroom.

The intersection of education and information and communications technology is the construct distance education.

Distance education exists. Distance education will not disappear. Distance education will change as institutions change. The fundamental question of this dissertation is not who, what, where, or when. The fundamental question is how an organization wants to transform itself. This dissertation focuses on policy and strategy that might influence the configuration of distance education some years from now.

This study limits itself to those actively participating in formulating policy for distance education at the University of Hawai‘i. The Delphi exercise will take place in a semester during which external events are at best challenging. The institution remains challenged to deliver quality education to growing numbers of students while simultaneously experiencing resource restrictions from budget cuts. One expects that economic concerns, security, and long-term concerns will reflect a bit more pessimism than might normally be the case.

In the appropriate sections, considerable space is devoted to sources of bias. Qualitative research remains particularly subject to claims that the intimacy of the researcher in conducting the research greatly precludes the objective examination of the report. The responses to this claim are many faceted and discussed in the bias section. There is no way to contain bias except to recognize its existence and create a record that allows others to review. Bias is an accepted reality of this research.

1.3 Importance of the Study

This desired end of this study is the codification of a Policy Delphi application focused on distance education within the University of Hawai‘i System. The results of the study are expected to outline a desired future state of distance education some 15 years into the future. As with any consideration of the future, the desired states become more and more difficult to predict with
certainty and great care must be taken to present multiple outcomes for any one theme
developed.

A number of researchers cited in the literature review argue that one will not know the
future with certainty but the simple fact of not knowing should not preclude discussing it. The
very acts of communicating and of thinking about the future shapes what the future might be
because resources may be devoted in the present. However, the future that might deliver ICT
mediated higher education is something that an institution can decide now and observe the
outcome in the future. The institution could decide to change its reward structure to encourage
distance education delivery or do nothing to discourage it. The institution could improve faculty
acceptance or do nothing. There are choices that can be made to influence the future and
painting a picture of the most desired outcomes is an important activity.

Waig, Willging, and Wentling summarize the central concern of this study in the
paragraph the follows,

Banas and Emory, (1998) stated, “while there is a growing recognition of and
attention to distance learning in higher education, its growing inclusion in
academia significantly changes the educational environment and experience.
These changes need to be acknowledged and discussed by all stakeholders.”
(p.5) This study opens the door to the second phase of this research which is
getting the detail on how technology will bring a new way of learning, how
technology will improve the quality of learning, how e-Learning will sustain and
maintain the intellectual capital and address learner diversity. (Waig, Willging,
& Wentling, 2002, Conclusions Section)

1.3.1 As the Study Adds to Methodology
This study is not a methodological investigation. Thus, outcomes related to methodology
are serendipitous. This study does not investigate either Delphi or distance education
methodology, efficacy, or efficiency. True, the literature review provides insights into these
concerns but the literature review is not converted into the focus of the research. There is no
intent to draw specific conclusions.

The methodology is realized in a web-based implementation of the Policy Delphi. The
software is discussed in detail later in this document.

The methodology value expected may come from the selection of Delphi literature that
informs the software design. The software design directly affects the Policy Delphi exercise. For
example, the literature suggests that anonymity of the participants is not absolute and even in the strictest sense in an asynchronous web-based exchange that the participants construct social identity. Many researchers suggest that this implies that anonymous participants be provided a “handle” or “pen name” for identification so that others might trace an individual’s argument and potentially even counter it. The actual identity of the participant behind the pen name remains anonymous. The software might include a calendar to provide a timing sense. The software uses a snowball or modified snowball selection of participants instead of a direct selection by the investigator.

1.4 Researcher’s Perspective

Education, learning, and teaching are the fundamentally human activities. Each generation advances understanding of these processes to improve the processes themselves but also to improve the result of these processes.

This researcher believes that information and communications technology will play an ever-increasing role in teaching, learning, and education. This researcher is not convinced that we will replace the human component of education with an ICT intervention. However, it is highly likely – just from the economics and world egalitarian viewpoints and the current insatiable demand for more education – that ICT will play an increasing role. Our current institutions will change as they have done before. Some believe the change will be radical and some believe less than moderate. Our institutions will adapt to insure the most effective, efficient, and quality driven use of ICT. This researcher believes that significant change is likely just as it has in business and the general society.

This researcher is not an advocate of speedy implementation because fads in the use of ICT have come and gone. The human organization side warns against belief that radical, immediate change will take place. Rogers (Rogers, 2003) and his concepts of diffusion of innovation along with the wide variety of material read make this researcher realize that good management, good plans, solid theory, good technology, and involvement of the organization will be keys to success.
This researcher embarks upon this investigation with the hope of learning where others who have key roles in higher education in Hawai‘i want the education, learning, teaching, and ICT future to evolve.
Chapter 2. Literature Review

The Delphi methodology exists within the larger framework of futures research as a foresighting tool to consolidate expert opinion. Many other tools exist and the literature review will justify the selection of the Delphi technique. This literature review will investigate the Delphi methodology, what is known about it, and the issues in its use.

The more important use of the Delphi literature review is extracting procedural findings from the literature. The procedural findings become the traditional user requirements that inform software design.

Delphi is one of member in the class of structured communication methodologies. Rules, procedures, and protocols characterize structured communication. Groups typically identify an external enforcer who is knowledgeable and skilled in the structured communication to be used. With the procedures encapsulated in a web-based software application, the application itself becomes the enforcer. The set of rules are the user requirements previously identified and these rules are an amalgam of many current practices and research findings.

2.1 Selecting a Foresighting Methodology

Numerous foresighting tools exist. Even though this chapter began by indicating that Delphi is the preferred methodology, the decision to use this technique followed from analysis of the problem space where "no known right answer exists; people disagree about the nature of the problem; the investigation is cross disciplinary." (Morrison, Renfro, & Boucher, 2005, in Stages).

As discussed in the introduction to this dissertation, the investigation and results are judgmental, by others, and with feedback. Beginning with the first selection of “judgmental” versus “statistical” in reference to the figure above entitled “Foresighting Methodology Tree”, this research is not statistical, not model driven. Considering the self-other choice, the knowledge source is not self. Continuing down the tree, the communication is structured. The Delphi is then one of the potential choices. The Delphi, as previously discussed, is the methodology best approximating the available resources and the required outcome. Rowe writes,

Delphi is not a procedure intended to challenge statistical or model-based procedures, against which human judgment is generally shown to be inferior; it is intended for use in judgment and forecasting situations in which pure model-based statistical methods are not practical or possible because of the lack of appropriate historical/economic/technical data and thus where some form of human judgmental input is necessary. Such input needs to be used as efficiently as possible and for this purpose the Delphi technique might serve a role.” (Rowe & Wright, 1999, p. 354)

According to the chart, the two most reasonable approaches are “prediction markets” and “Delphi.” Prediction markets apply market theory and valuation to the expectation of a number of outcomes (Berg & Rietz, 2003; Leigh & Wolters, 2006; Ray, 2006; Servan-Schreiber, Wolters,
Pennock, & Galebach, 2004; Surowiecki, 2005; Wolfers & Zitzewitz, 2004). Again, from the introduction, the participants in the process all have significant expertise within the context of the question.

What characteristics influence the choice of Delphi? Turoff provides characteristics of an investigation for which Delphi is appropriate as below,

1. The problem does not lend itself to precise analytical techniques but can benefit from subjective judgments on a collective basis. The individuals needed to contribute to the examination of a broad or complex problem have no history of adequate communication and may represent diverse backgrounds with respect to experience or expertise
2. More individuals are needed than can effectively interact in a face-to-face exchange
3. Time and cost make frequent group meetings infeasible
4. The efficiency of face-to-face meetings can be increased by a supplemental group communication process
5. Disagreements among individuals are so severe or politically unpalatable that the communication process must be refereed and/or anonymity assured
6. The heterogeneity of the participants must be preserved to assure validity of the results, i.e., avoidance of domination by quantity or by strength of personality ("bandwagon effect") (Linstone & Turoff, 2002, p. 4)

The central tenant of Delphi is the belief that more heads are better than one for forecasting. Of course, there is no measure to indicate that an empanelled group at the same time with the same background might not have prognosticated similarly.

Delphi studies do result in “extremely helpful strategic planning; a few virtually decided the future of the sponsoring organization. But most had little or no effect, apart from providing general background information or satisfying a momentary curiosity about this novel method of forecasting.” (Morrison, et al., 2005, in Stages).

Delphi has an advantage for many decision-makers because “It is not difficult to explain the essence of the method to potential participants or to one's superiors. Though there are trappings of quantitative results, the hearing of every participant and the gathering of their feedback makes determining uncertainties of participants and reviewers easier to determine. With careful panel selection and attention to procedures the Delphi produces a full spectrum of hopes, fears, and other expectations” (Morrison, et al., 2005, in Stages) of the participants.
2.2 Delphi Method

Delphi fills the methodological gap for a technique to provide insight for complex problems. As will be outlined in a brief history to follow, the Delphi originally limited itself to a panel of experts asked to forecast highly technical outcomes far into the future for which no other methodology existed. Though consensus results as Delphi proceeds, consensus is not the central goal. Often divergent opinions remain thus requiring continued consideration in the final report. Delphi is not about the answer but rather it may result in many varied options with some relative weighting assigned to each.

Rowe, Wright, and Bolger recognize “The main criterion for Delphi’s employment is indispensability of judgmental information, which may arise in cases (such as forecasting) where no historical data exist, or when such data are inappropriate (that is, new influencing factors are expected that are not incorporated in the past data).” (Rowe, Wright, & Bolger, 1991, p. 236). They reviewed the literature to find Delphi used in “policy formation, where ‘validity’ of judgment is a less salient concept.” (Rowe, et al., 1991, p. 239). However, they expressed concern that Delphi for prediction should be “a technique of ‘last resort’, to be used when no adequate models exist upon which statistical prediction or judgment can be based.” (Rowe, et al., 1991, p. 239).

Delphi procedures today apply to far more varied situations than originally imagined. Mullen identified at least 23 named variations of Delphi including such titles as Delphi, classical Delphi, real-time Delphi, Policy Delphi, fuzzy Delphi, quantitative Delphi, exploratory Delphi, and so on. (Mullen, 2003, pp. 38-39). Each of these variations modified one or more of the protocols and/or assumptions of the original Delphi. As will be discussed, the variations in the technique -- generically captured under the Delphi banner -- contribute to the contradictory efficacy findings in research literature. Published research seldom provides self-categorization against a matrix of documented variations – partially because such authoritative categorizations do not exist.

Dissertations typically explore theoretical issues and judge inquiry methodology like Delphi based on some abstract construction. The long history of mathematical based forecasting in fields like economics, management science, demography, and stock markets takes on the
appearance of validity yet "In each of these the ability to predict is overstated, often with forecasters only doing as well as chance." (Sutherland, 2006, p. 600).

Sutherland, writing about the need for prediction in ecology, claims a lack of prediction of outcomes due to input changes leads to ecology being taken less seriously. Sutherland sees the challenge as "extending prediction while maintaining credibility" with the value accrued as "stimulating thought and further empirical exploration." Sutherland writes,

There is an important difference between predicting the consequences of routine changes, such as habitat loss and change in exploitation, compared with predicting the consequences of novel conditions, such as climate change and new agricultural systems. Predicting the consequences of novel conditions is particularly difficult. (Sutherland, 2006, p. 600)

Sutherland argues, that the approach traditionally used to make assumptions, derive models, and use these models "to make predictions about the consequences of environmental change" (Sutherland, 2006, p. 610) is less than satisfactory for complex problems. He believes, "future work on such problems will require greater exploration of techniques such as expert opinion, output–driven modeling and scenarios." (Sutherland, 2006, p. 610).

2.2.1 Delphi Technique History

The close of WWII generated a realization that future conflict would rely increasingly upon scientific and technological advantages. The recognition of the need appears in a report to the Secretary of War:

During this war the Army, Army Air Forces, and the Navy have made unprecedented use of scientific and industrial resources. The conclusion is inescapable that we have not yet established the balance necessary to insure the continuance of teamwork among the military, other government agencies, industry, and the universities. Scientific planning must be years in advance of the actual research and development work.” (RAND, 2007, in History page)

US Air Force General Arnold ordered the creation, in 1944, of a report on "the future technological capabilities that might be used by the military." (RAND, 2007, in History page). The contract to produce the report was given to Douglas Aircraft Company. Douglas Aircraft in turn established Project RAND. The RAND project – the acronym standing for “research and development” – eventually was spun off from Douglas to become an independent, non-profit organization. RAND's self-history reports,
It was on May 14, 1948, that Project RAND — an outgrowth of World War II — separated from the Douglas Aircraft Company of Santa Monica, California, and became an independent, nonprofit organization. (RAND, 2007, in History page)

As RAND began the investigation, tools for foresighting were few. RAND needed a new approach. Each method RAND investigated appeared to be inappropriate or weak. Delphi evolved from this investigation and need. “Project Delphi” originated in the early 1950’s as a classified research project “to obtain the most reliable consensus of opinion of a group of experts [using] a series of intensive questionnaires interspersed with controlled opinion feedback.” (Dalkey & Helmer, 1963, p. 458).

The task facing the original expert panel was a forecast of the number of A-Bombs the US would target at the Soviet Union to reduce capabilities of the Soviets by a prescribed amount.

The original paper by Norman Dalkey and Olaf Helmer on the Delphi methodology laid out the principles that continue to guide the use of the methodology today as below,

The technique employed involves the repeated individual questioning of the experts and avoids direct confrontation of the experts with one another.

The questions, which are all centered around some central problem, are designed to bring out the respondent’s reasoning that went into his reply to the primary question, the factors he considers relevant to the problem, his own estimate of these factors and, thereby a more confident answer to the primary question. The information fed to the experts between rounds of questioning is generally of two kinds: It consists either of available data previously requested by some experts (e.g., output statistics for steel mills, or of factors and considerations suggested as potentially relevant by one or another respondent. With respect to the later type of information, an attempt was made to conceal the actual opinion of other respondents and merely to present the factor for consideration without introducing unnecessary bias.” (Dalkey & Helmer, 1963, p. 458)

Dalkey and Helmer believed that classic decision-making approaches placed limits on participants being not “conducive to independent thought on the part of the experts.” (Dalkey & Helmer, 1963). Classic decision-making procedures are confrontational thus inducing “the hasty formulation of preconceived notions, an inclination to close one’s mind to novel ideas, a tendency to defend a stand once taken or, alternatively and sometimes alternately, a predisposition to be swayed by persuasively state opinions of others.” (Dalkey & Helmer, 1963). A better process would explore and clarify ideas of the individual without the concomitant negative effects originating from group dynamics. During the exploration of ideas, experts may require data input.
Dalkey and Helmer warn against researcher bias in the provided data. Any input data must be the request of at least one participant and be “obtainable from reliable sources, and to suggest only such theoretical assumptions as seem to represent a consensus of a majority of respondents.” (Dalkey & Helmer, 1963, p. 459).

Because the Delphi appears to deal with facts and expert opinion, the feedback mechanism reduces divergence and increases consensus among the panel members. There is no expectation of a final, true, singular answer. The authors state,

> It cannot be ideally expected that the final responses will coincide, since the uncertainties of the future call for intuitive probability estimates on the part of each respondent.” (Dalkey & Helmer, 1963, p. 459)

Dalkey and Helmer believe the consensus reached by the expert panel remains subject to further correction upon completion. Thus, the panel itself is not the final arbiter of its own findings. The authors argue that final clarifications are an important and justifiable consequence.

Final clarifications and interpretations require great care,

> Such corrections are in fact an integral part of the procedure; they must however, be based on a careful analysis of the responses, taking into account what can be learned regarding (i) a consensus as to basic assumptions, (ii) the sensitivity of the individuals’ responses to changes in these basic assumptions, and (iii) their estimates of functional dependencies rather than mere point estimates. Essentially, the resulting corrections amount to a replacement of the individual expert’s estimates concerning some of the components of the main problem by a consensus of estimates by all the experts.” (Dalkey & Helmer, 1963, p. 459)

Dalkey and Helmer provide the original definition of the Delphi process, the problems in communication and forecasting that the process was to address, and the expectations of the outcome. The original writings make very clear the foundations and the process. With time, the actual process varied sometimes being the subject and object of research. The technique enjoys wide and continued application in diverse environments.

2.2.2 The Delphi Procedure as the Object of Research

The Delphi method has grown considerably in use in the broad community of group dynamics and decision-making. The continued frequency of use makes Delphi a good candidate for academic research. Researchers investigate the basic structure of the procedure. Researchers investigate the efficacy of component parts. In addition, researchers also use the tool to conduct foresighting exercises.
Landeta reviewed the use of Delphi in the Social Sciences. He writes,

However, the method is not exempt from significant methodological weaknesses, highlighted in numerous works, such as its basic source of information (who is expert, what biases each expert has, etc.), the use of consensus as a way to approach the truth, the limitation of the interaction involved in written and controlled feedback, the restriction to the possibility of social compensation for individual contribution to the group (the reinforcement and motivation normally provided by the support and social approval of the other expert group members are removed), the impunity conferred by the anonymity with respect to irresponsible actions on the part of the experts, the ease inherent in the methodology of interested manipulation by the person running the study, the difficulty of checking the method’s accuracy and reliability, the time required to carry it out, the effort required on the part of the participants and the non-consideration of possible interrelations between the forecast incidents.

In addition to these methodological weaknesses, the Delphi has also received criticism that is not due to the technique itself but to its deficient application, such as the not very rigorous selection of experts, the lack of explanation concerning its evolution and dropout, questions and problems that are badly formulated, insufficiently analyzed results, etc. (Landeta, 2006, p. 469)

Landeta also comments on the literature concerning the efficacy of Delphi by noting that researchers have often varied the Delphi process itself or failed to follow requirements. Thus, he would assert that the lack of control by researchers investigating the efficacy of Delphi itself is a major contributor to variation in results reported. He states that “this sensation is due principally to the contrast existing between the apparent simplicity of the technique, which motivates neophytes to use it the first time, compared with the work and difficulty involved in its execution.” (Landeta, 2006, p. 469).

In other articles that review Delphi efficacy, the phrase “classical Delphi” as opposed to “Delphi” appears to indicate the procedure adheres reasonably well to the original conception of the process versus some significant variation. Rowe and Wright did an exhaustive review of Delphi studies in which they applied filters relatively closely aligned with “classical Delphi.” Their review focused on experimental evaluations. They further filtered their review by eliminating articles that simply reported the use of Delphi in some application. If methodology was not clearly stated in the article or if the results or process were so “highly complex, difficult to encode and, indeed, difficult to interpret the study” (Rowe & Wright, 1999, p. 356), then the article was eliminated. They also eliminated articles that claimed universal validity or reliability without clear statement of control conditions. They cite this example,
For example, the study of Ono and Wedemeyer reported that a Delphi panel produced forecasts that were over 50% accurate, and used this as ‘evidence’ that the technique is somehow a valid predictor of the future. However, the validity of the technique, in this sense, will depend as much on the nature of the panelists and the task as on the technique itself, and it is not sensible to suggest that a percentage accuracy achieved here is in any way general stable or fundamental (i.e., to suggest that Delphi will generally result in 50% correct predictions.) (Rowe & Wright, 1999, p. 356)

Rowe and Wright believe Delphi procedures reduce variance in nearly all cases making the study of consensus and convergence no longer a research interest. The question thus is not the study of consensus. The more interesting research focus is the relationship between consensus on Delphi outcome statements and the accuracy of the outcomes. For example, does consensus on a list of events related to the economy mean that the predictions are accurate? Is accuracy a function of consensus? If not, then what is the role of consensus?

Delphi proponents believe that panelists draw toward consensus because of reasoned feedback. Consensus is not due to increased pressure to conform to group norms. Not everyone agrees that Delphi panel’s consensus convergence is the result of reasoned feedback. Some research suggests “panelists were simply altering their estimates to conform to the group without actually changing their opinions.” (Rowe & Wright, 1999, p. 363). Other research indicates more acceptance of the role of feedback in consensus building. Others report that those with the extreme positions are highly likely to drop out, as the norms are not representative of their views. Dropping extremes creates false reduction in variance and contrived consensus. The loss of extremes raises the importance of post-Delphi attrition consideration. Delphi creators argue that extreme views gain more respect in the process. Panelists with extreme views are encouraged to continue participation.

The interaction of feedback and consensus remains central to Delphi. Rowe and Wright openly worry, “if the panelists are being drawn towards a central value for reasons other than a genuine acceptance of the rationale behind that position, then inefficient process-loss factors are still present in the technique.” (Rowe & Wright, 1999, p. 369). In earlier writings, they cite the work of Hogarth who added additional experts to a panel to find that more convergence to the mean appeared. However, they found that “the greatest improvement in group accuracy might actually be obtained by adding a less valid expert who acted to decrease inter individual
correlation within the group." (Rowe, et al., 1991, p. 369). Though Hogarth was looking at statistical group processes – the feedback being only mean scores on the feedback instrument -- the implication develops that Delphi may act like a statistical group for the reason that after the feedback is considered there is no difference between a statistical group and Delphi. The implication for Delphi is "the effectiveness of any individual Delphi study is going to be influenced by aspects such as the number of panelists and the relative validity or expertise" (Rowe, et al., 1991, p. 369) rather than the believed feedback effect resulting in considered choices. The challenge to Delphi that reasoned feedback may not be the dominant factor creating group consensus building remains unanswered.

Delphi’s long-range predictions challenge determination of the accuracy of Delphi. Rowe and Wright write, "long-range forecasting and policy formation items that are typical in Delphi are seldom used." (Rowe & Wright, 1999, p. 364). To counter this limitation, efficacy experiments use short-range tasks or tasks that estimate well-known, well-established almanac data. Groups with no-feedback compared to Delphi with feedback generally did less well. There is also evidence that Delphi processes made no difference. Additional work used alternative structured communication techniques like nominal grouping, but no clear advantage exists.

Though Delphi does not unquestionably establish efficacious behavior, there are fundamental issues in the design of the experiments that make claims to be investigations of Delphi. First, most experiments did not use experts. Secondly, most did not do foresighting but attempted to equate reaching agreement on known almanac data as tantamount to assertions concerning an unknown future. For example, "It is difficult to imagine how Delphi might conceivably benefit a group of unknowledgeable subjects—who are largely guessing the order of magnitude of some quantitatively large and unfamiliar statistic, such as the tonnage of a certain material shipped from New York in a certain year." (Rowe & Wright, 1999, p. 368). Sensible questions to a panel must "relate to the domain of knowledge of specific panelists." (Rowe & Wright, 1999, p. 368). Rowe and Wright drew no conclusion.

Rowe and Wright make direct comment about "technique comparison studies" as "such studies have used versions of Delphi somewhat removed from the ‘classical’ archetype of the
technique." (Rowe & Wright, 1999, p. 367). The assumption of Delphi remains the interaction of experts who gain deeper or clearer understanding by interacting with other experts. The laboratory, experimental simplification that usually uses homogenous samples of students “requires [them] to make judgments about scenarios in which they can by no means be considered expert, and about which they are liable to possess similar knowledge.” (Rowe & Wright, 1999, p. 368). They make this point: “if varied information is not available to be shared, then what could possibly be the benefit of any group-like aggregation over and above a statistical aggregation procedure?.” (Rowe & Wright, 1999, p. 368). As expected, research reports little improvement in Delphi results but “the lack of concern of experimenters in following even the fairly loose definitional guidelines that do exist suggests that there is a general lack of appreciation of the factors like the nature of feedback in contingently influencing the success of procedures like Delphi.” (Rowe & Wright, 1999, p. 369).

Rowe and Wright make clear that experimenters must improve their design. Experiments clarify the “type of feedback that constitutes Delphi feedback; the criteria that should be met for convergence of opinion to signal the cessation of polling; the selection procedures that should be used to determine number and type of panelists; and so on.” (Rowe & Wright, 1999, p. 369). Far too many have underestimated the need for definitional precision. They also state a need for “a much greater understanding of the factors that might influence Delphi effectiveness in order to prevent the technique’s potential utility from being underestimated through accurate representations use in unsympathetic experimental scenarios.” (Rowe & Wright, 1999, p. 353).

Woudenberg provides this summary,

The main claim of Delphi – to remove the negative effects of unstructured, direct interaction – cannot be sustained. In many Delphi a slight increase in accuracy over rounds is found. But the increase can be ascribed partly to mere repetition of judgment (possibly giving judges opportunity to contemplate their judgments) and partly to an artificial consequence of group pressure to conformity. A Delphi is extremely efficient in obtaining consensus, but this consensus is not based on genuine agreement; rather, it is the result of some strong group pressure to conformity. (Woudenberg, 1991, p. 145)

The comments from Woudenberg and others continue to generate evaluative research.
2.2.2.1 Sackman Critique

No single critique of Delphi appears stronger than Sackman’s. He reviews “the methodological principles and key assumptions associated with Delphi.” (Sackman, 1974, p. 2). He correctly reports that “the elusiveness of a fixed, universally agreed upon, working definition of Delphi” (Sackman, 1974, p. 3) limit a critique construction. He characterizes Delphi as assuming “participating panelists are experts in their subject area and that reported consensus was obtained through reliable and valid procedures.” (Sackman, 1974, p. 5).


The following paragraphs provide a high level summary of many of Sackman’s key arguments. Readers requiring complete details are encouraged to read his original work.

Sackman raises the obvious issue related to the Delphi panel. He notes the lack of an operational definition of “expert” and the non-random selection of the panel. Strict external validity -- gained by pure experimental design’s random selection and assignment to control and experimental groups with clearly operationalized variables -- is a Delphi weakness. Sackman raises concerns that the group procedures of Delphi allow attrition that is frequently uncontrolled and/or unreported.

Most Delphi use a form of questionnaire to gather expert input. Sackman aims critique at the reliability of the instrument. The use of Lickert scale instrument appears to be instrumented research but Sackman notes that none of the controls on instrumentation appears – construct validity, test-retest reliability, and so on. Even if one accepts the instrument, Sackman questions feedback mechanism whereby the normal procedure is to modify or retain items between rounds removing the consistency of the instrument. He asks, “What kind of feedback, how much feedback and in what form, should be presented to panelists? When is the point of diminishing
return reached? How long should intervals be between successive rounds, and how can participants be encouraged to respond quickly between rounds?" (Sackman, 1974, p. 6).

Critical to Delphi is the movement toward consensus. Sackman asks, what measure confirms consensus? He asks about the prominent role of the researcher: "Does the director reinforce and encourage conformist or dissenting behavior in successive rounds?" (Sackman, 1974, p. 6).

Even if all the methodological limitations outlined were countered, Sackman challenges the "write up and dissemination of results." (Sackman, 1974, p. 7). He posits the question relative to statistical results asking, "Should all statistics be accompanied by standard errors of estimate"? (Sackman, 1974, p. 7). He states, "Statistical significance is rarely reported in Delphi studies, either for precision of estimates or for tests of significance of mean or median differences between two or more forecasts. Consensus and precision are implied from suggestive graphs, not from standard errors of estimates." (Sackman, 1974, p. 14). Delphi reports fail to report main and interaction effects between rounds. He continues, "Nor are items compared for homogeneity of variance, linearity, and type of empirical frequency distributions." (Sackman, 1974, p. 15).

Sackman challenges the "claim that the end result of a Delphi study is a series of expert forecasts of future events." He writes,

Delphi items are typically broad, amorphous classes of events, not precisely defined empirical occurrences. Delphi forecasts are options about such broad classes of events, not systemic, documented predictions of such events. These opinions are typically snap judgments frequently based on free-association stereotypes. Consensus for such opinion tends to be manipulated consensus to minimize dispersion of opinion. (Sackman, 1974, p. 59)

Sackman summarizes the belief of Delphi proponents that the procedure "produces reasonable and useful forecasts of object events" by asserting the "goal is not attained" (Sackman, 1974, p. 59) given the small samples and the manipulated convergence. More so, "The epistemological confusion arises from focusing on Delphi results and taking them at face value as expert predictions of the future, rather than looking at the underpinning method which reveals Delphi as an attitude polling technique deal in snap judgments of ill-defined issues." (Sackman, 1974, p. 59). Sackman attacks the use of panel anonymity wondering if the "blandishments of joining the inner circle of eminent experts" comes with "the price of such
inducements is the abandonment of accountability, and may promote elitist vested interests.” (Sackman, 1974, p. 63). The director of the Delphi reports “group opinion following an objective ritual” from faceless panelists who can “blame nameless others for findings he dislikes.” (Sackman, 1974, p. 63). Directors of Delphi panels should be accountable for “flaws in the method and implicitly reporting and explicitly overstating the value and significance of potentially misleading final results.” (Sackman, 1974, p. 63). Sackman directly criticizes expert participants who “should be accountable for unwittingly lending the authority of their reputations and their support to demonstrably unreliable and invalid short-cuts to the future.” (Sackman, 1974, p. 63).

Sackman challenges the very process of decision-making within a Delphi panel. Sackman worries, “Delphi deliberately factors out face-to-face confrontation, and the adversary process associated with it, as one of its prime philosophical tenets justifying efficient consensus. Arguments are filtered, buffered, and effectively neutralized in Delphi.” (Sackman, 1974, p. 65). Panelists can participate in the entire process having never been countered nor justifying their vote. Sackman believes that knowledge is validated when cross examined.

2.2.2.2 Sackman Critique in Summary

Sackman’s epilogue summarizes Delphi issues (Sackman, 1974, p. 69),

1. Often characterized by crude questionnaire design
2. Lacking in minimal professional standards for opinion item analyses and pilot testing.
3. High vulnerable on its concept of “expert” with unaccountable sampling, selection of panelists, expert or otherwise.
4. Abdicating responsibility for item population sample in relation to theoretical constructs for the object area of inquiry.
5. Virtually oblivious to reliability measurement and scientific validation of findings.
6. Capitalizing on the fallacy of the expert halo effect.
7. Typically generating snap answers to ambiguous questions representing inkblots of the future.
8. Seriously confusing aggregations of raw opinion with systematic prediction.
9. Capitalizing on forced consensus based on group suggestion.
10. Unwittingly inhibiting individuality and any adversary process by overtly and covertly encouraging conformity and penalizing the dissident.
11. Reinforcing and institutionalizing premature closure, using a highly questionable ritual for conducting opinion studies that tends to inhibit more scientific approaches.
12. Giving an exaggerated illusion of precision, misleading uniformed users of results.
13. Indifferent to and unaware of related techniques and findings in behavioral science
14. Producing virtually no serious critical literature to test basic assumptions and alternative hypotheses.
15. Denigrating group and face-to-face discussion, and claiming superiority of anonymous group opinion over competing approaches without sufficient proof.
16. Encouraging a short-cut social science method that is lacking in minimum standards of professional accountability.

2.2.2.3 Response – Delphi as Experimental Research

Delphi originated in the period where social science sought credibility by mimicking physical science epistemology. The dominant form of research as cited by Sackman was experimental. No persuasive alternatives existed even though Rowe makes clear the intent of Delphi was other than experimental research as the unitary model for explaining the world. Rowe writes,

Delphi is not a procedure intended to challenge statistical or model-based procedures, against which judgment is generally shown to be inferior: it is intended for use in judgment and forecasting situations in which pure model-based statistical methods are not practical or possible because of the lack of appropriate historical/economic/technical data, and thus where some form of human judgmental input is necessary (e.g., Wright, Lawrence & Collopy, 1996). Such input needs to be used as efficiently as possible, and for this purpose the Delphi technique might serve a role. (Rowe & Wright, 1999)

When Sackman wrote there were no methodological standards other than experimental that were acceptable in quality social science literature. Because Delphi used “numbers”, “instruments”, and “averages”, the method appeared to be subject to measure against experimental methodology. Yet, as indicated above by Rowe, Delphi is not part of the experimental genre. The use of some tools, the use of numbers and graphs, and guilt by association without denial by the experimenters and originators of Delphi allowed for appearances not justified. Many of Sackman’s specific criticisms, for example the focus on the instrument used in feedback, remain valid. The casual use of mean scores on items unvalidated could be deceiving and misleading toward a consensus incorrectly manufactured.

Delphi is not an experimental methodology. Delphi is a qualitative methodology. When Sackman wrote his critique, qualitative research was not an acceptable methodology in most prestigious social science journals. The literature is strewn with multiple articles arguing for the
acceptance of the qualitative approach in Sackman’s timeframe. Only recently, APA included reference to qualitative research. The proponents of qualitative research, like Lee (A. Lee, 1989), make clear that the argument for qualitative research’s credibility and inclusion in journals remains difficult though more acceptable today.

Rowe & Wright argue that Delphi is a quantitative technique but within the qualitative research model. Skulmoski reminds, “Qualitative research is interpretivist in the sense that the researcher is interested in how the social world is interpreted, understood and experienced; the researcher is flexible and sensitive to the social context within which the data was collected; and qualitative research is about producing holistic understandings of rich, contextual and detailed data.” (Skulmoski & Hartman, 2007, p. 9). The data normally collected in a qualitative study is non-quantitative yet there is no prohibition against the collection of quantitative data. When qualitative research speaks of the “rich, contextual, and detailed data”, the Delphi stands available to deliver for researchers as Skulmoski explains,

The Delphi method is well suited to rigorously capture qualitative data. It may be seen as a structured process within which one uses qualitative, quantitative or mixed research methods. Such flexibility not only affords the ability of the method to answer many research questions, but also can be well matched to the abilities and aptitudes of the graduate student. (Skulmoski & Hartman, 2007, p. 9)

2.2.2.4 Countering Sackman Directly

Goldschmidt provides a frontal assault on Sackman’s analysis concluding “Delphi Assessment2 is not a systematic or objective evaluation of either Delphi method or Delphi applications.” (Goldschmidt, 1975, p. 213). He constructs a thorough questioning of the structure, content, and foundation of Sackman’s critique by stating, “As one might suspect by now, the author [Sackman] places heavy emphasis on the principles of scientific inquiry. This might lead one to conclude that Delphi Assessment is a systematic study of Delphi applications and an evaluation of Delphi effectiveness, but this would be wrong.” (Goldschmidt, 1975, p. 195).

---

2 Delphi Assessment is Goldschmidt's shorthand reference to the title of Sackman's monograph.
Goldschmidt notes many instances where Sackman fails to define terms and uses broad brush, common sense, non-operationalized definitions like “the fundamental mistake of equating the multitudinous applications of the Delphi technique with technique itself.” (Goldschmidt, 1975, p. 196). Ethically and professionally, he claims that Sackman misrepresents the research as the “report is replete with quotations taken out of context, incorrect reference, and on occasion ‘findings’ that are not borne out by the citation.” (Goldschmidt, 1975, p. 197).

A specific weakness in Sackman’s review stems from a failure to make clear which Delphi is the subject of his study – a qualification that scientific researchers understand as a prerequisite. Careful construct definition is paramount but Goldschmidt believes the proliferation of variants of Delphi confound Sackman’s use of “conventional Delphi” and that it is unclear the boundaries Sackman applied. Goldschmidt does agree that a new technique does require review because “it is not unusual for a new and promising technique to be hailed as a panacea, particularly by those who can see a way to make money from its application or those who do not understand the technique's limitations.” (Goldschmidt, 1975, p. 197).

Goldschmidt makes a distinction between an investigation of the Delphi technique and the Delphi application. A technique “represents the set of procedures that are to be applied in a prescribed manner” whereas an application “represents the way in which a technique is actually applied by different practitioners.” (Goldschmidt, 1975, p. 198). Only applications can be measured and applications must be congruent with a prescribed technique or excluded from review. Goldschmidt claims Sackman fails to distinguish which he includes and he fails “because of the absence of a report on the methodology used” (Goldschmidt, 1975, p. 198) to select the applications included in his review. One major theme of Sackman’s critique originates from his review of “150 Delphi studies conducted at RAND and elsewhere” (Goldschmidt, 1975, p. 198) but there is no discussion of the methodology Sackman used to select the studies nor is a discussion of the methodology applied to assess a study explicitly stated. Goldschmidt argues that Sackman fails to demonstrate “the studies Sackman cites match the eleven characteristics he lists for the ‘conventional Delphi’ technique.” (Goldschmidt, 1975, p. 198).
Goldschmidt is very critical of Sackman by calling his “strategy outlandishly misleading” in his contrast of Delphi with the standards of the American Psychological Association. At the heart of this challenge is Goldschmidt’s belief that Sackman misunderstands and misapplies the standards that explicitly state that the standards “are not meant to be standards ‘for social science experimentation and for opinion questionnaires’.” (Goldschmidt, 1975, p. 199). Moreover, the focus of instruments within scope of the document are “standardized tests that are used to make predictions or statements about the individuals tested on the basis of test items” whereas Delphi questionnaires “make predictions or statements about events or outcomes, and not to make predictions or statements about the members of the Delphi panel based on statistical inference with regard to their responses to Delphi questionnaire items.” (Goldschmidt, 1975, p. 199).

Goldschmidt identifies the next attack by Sackman as focusing on “methodological principles and key assumptions associated with Delphi.” (Goldschmidt, 1975, p. 198).

Goldschmidt makes this elegant statement relative to Sackman’s attack on the epistemological foundations of Delphi that futurists would support,

The future, being a hypothetical system, cannot be observed, and cannot, therefore, be the subject of empirical experimentation, that basis of classical science. (Goldschmidt, 1975, p. 200)

Sackman cites a number of authors – among them Pill. However, Goldschmidt remains concerned that Sackman selectively quotes out of context Pill’s work and he provides the full quotation to demonstrate the selective quoting. In specific, Sackman infers Pill as stating “So perhaps the Delphi technique should be less allied with science than with metaphysics.” (Goldschmidt, 1975, p. 200). Unfortunately, as Pill is quoted by Goldschmidt, there is significant pretext to the quote extracted above that is not presented nor is the text that follows that states “It would seem that Delphi would be most useful in cases where there is no way of immediately confirming the results, for instance, in long-range forecasting; even as an academic exercise, it
does not seem reasonable to test the method with almanac data, because these questions are too concrete. It is in the questions of intuitive judgment.. that Delphi is useful and as such, one cannot judge it on the same basis as concrete measurement.” (as cited in Goldschmidt, 1975, p. 299). Yet further, Pill as writing “Delphi should be used at high levels of uncertainty and one must accept the difficulty of its usefulness; its eventual usefulness will be judged by its performance [by evidence]; [and] research in Delphi should stress its psychological aspects in terms of communication rather than in mathematical terms.” (as cited in Goldschmidt, 1975, p. 201).

Sackman asserts that experts and any collection of participants in a Delphi are about equal in their ability to forecast future events but Goldschmidt again attacks Sackman on his lack of definition of “expert.” Further, Goldschmidt reviews the research used by Sackman to reach his conclusion and finds it to incorrectly summarize results and that the research itself may not have been applicable or related to the determination of the question at hand. Because Sackman is asserting that group forecasting fails “one would expect Sackman to cite a substantial body of empirical findings to support the claim. None is offered.” (Goldschmidt, 1975, p. 205).

Goldschmidt continues his review of Sackman on consensus countering some of the claims. Continuing to the Delphi questions [items in the Delphi instrument], Goldschmidt agrees that “Delphi practitioners have been lax in writing professionally acceptable items for their questionnaires, but does it necessarily follow that this is an inevitable characteristic of the technique itself?” or of sloppy execution (Goldschmidt, 1975, p. 207).

Another challenge by Sackman is Delphi’s ability to foresight – to state unambiguously a future event. Even Sackman agrees that accuracy presupposes “a scientifically replicable calendar/stopwatch concept of forecasting validity where an impartial observer with a stopwatch waits for the objective event to happen, clocks it, and records the date and time of occurrence.” (Goldschmidt, 1975, p. 211). Sackman, in agreeing to the difficulty of such a measurement, still

---

3 This early statement suggesting that research investigations into Delphi efficacy that would compare the results of a Delphi panel’s ability to “predict” what almanac data – supposedly true data – might say on a topic is not valid nor of interest due to the kind of problems Delphi is designed for and yet later work ignored this an proceeded to compare Delphi exercise results to almanac data.
asks “how can anyone validate the truth or falsity of an inkblot of the future?.” (Goldschmidt, 1975, p. 211) as cited in Sackman is arguing that accepting ambiguity in foresighting makes the validity impossible to determine. Goldschmidt asserts such a determination could be made as to “whether or not it aided decision-makers more than other methods [and] not that it offered better predictions than other methods.” (Goldschmidt, 1975, p. 211). On this very fundamental of the value of Delphi, Goldschmidt writes elegantly and is quoted verbatim and in full,

The problem with the concept of prediction validity is not that the event is ambiguous, but that the concept of prediction validity may be meaningless. The future can be conceived of as a hypothetical dynamic system subject to the influence of past events, random occurrences, and purpose-creating systems. This is in contrast to the static view of the future, such as divine preordination, where the forecaster’s role is to search for ways of unveiling future events that have already been determined.

Social events are subject to the influence of policy, and policies often represent the governing forces that produce social events. Individual forecasts made by different persons are bound to be based on diverse policy assumptions. Rather than being interested in a specific event as the outcome of unspecified policies, decision-makers are often more interested in the policies that would produce the event, or in the exploration of the relationship between specified policies and various events. Delphi studies that do not explore the assumptions underlying a forecast are, therefore, of limited ability to decision-makers.

If the decision-maker is to make an informed resource allocation, he must explore strategies that will lead to an event and understand the consequences of each, not only in their association with the event, but also their terminal value and the implications to his organization or society of committing resources to one strategy as opposed to another. (Goldschmidt, 1975, p. 211)

Concluding Goldschmidt's review of Sackman, he believes that if Sackman “stimulates practitioners to use better Delphi technique or to be more cautious in the use of experts, or if it simply encourages Delphi users [consumers] to be more critical of methods, results, and conclusions, then it [Sackman's article] will have achieved a great deal.” (Goldschmidt, 1975, p. 213). Ultimately, Goldschmidt believes that the promise encapsulated above of Sackman's intent “delivers so little” (Goldschmidt, 1975, p. 213) in relationship to this goal.

2.2.2.5 Accuracy and Confidence

Sniezek and Henry write, “Accurate human judgment of unknown quantities has great utility in many organizational contexts, but can be difficult to achieve. A common practice is to assign multiple persons to a judgment task. But while the use of multiple judges may increase
By the nature of the task, judgments of "unknown quantities will always have a degree of uncertainty." Judgment accuracy – the difference of the judged outcome and the actual outcome – may not be knowable. In this case, "the quality of the judgment must itself be estimated – a judgment of a judgment called confidence assessment." (Sniezek & Henry, 1989, p. 2). Their literature review suggests that "individuals are not well calibrated in the assessment of their responses in non-trivial tasks but are prone to overconfidence." (Sniezek & Henry, 1989, p. 3). In a similar fashion, group decision-making confidence assessments are required. However, should the group be overconfident then "groupthink" may result though they stated "it seems unlikely that groups will be subject to the same degree of overconfidence bias due to anchoring with insufficient judgment." (Sniezek & Henry, 1989, p. 20).

Sniezek and Henry demonstrated that group judgments were "significantly more accurate than mean or median individual judgments" and "group judgments were more accurate than the most accurate individual judgment." (Sniezek & Henry, 1989, p. 20). Their experiment concludes, "The overestimation and underestimation biases observed with individual judgments are reduced considerably through group interaction." (Sniezek & Henry, 1989, p. 20). Just using group decision-making is not in-and-of-itself sufficient cause but rather "groups with relatively larger variances in their distribution of individual judgments show more improvement over average individual accuracy" as "without disagreement, it is difficult for group members to distinguish among their judgments to find a basis for variable weighting." (Sniezek & Henry, 1989, p. 22).

Most of these experiments conducted used general information (almanac) questions, where the subjects did not know the answers to the questions, but the answers were available from some reference work. Of course, in Delphi applications, unknown answers are the focus. The best information available is the judgment of knowledgeable individuals.

In general, the outcome of the experiment was that the Delphi procedures were at least as effective with short-range forecasts as with almanac material. The proportion of cases in which median estimated changed as a consequence of feedback was somewhat lower for the forecast questions; but for medians that did change the proportion of case in which the estimates improved was somewhat higher.
Perhaps more significant, the correlations between standard deviation and accuracy and between self-rating index and accuracy and between a group self-rating index and accuracy were distinctly higher for the prediction questions. The experiment gives no basis for expecting that questions involving "objective uncertainty" are inappropriate for Delphi treatment. (Dalkey & Brown, 1971, p. 3)

Martino reviewed "over forty published and unpublished Delphi" to examine the potential significance of the dispersion from trend line forecasts concluding

The remoteness of the forecast date and the degree of dispersion are definitely related. The regression coefficient is in nearly all cases highly significant for a single panel addressing a related set of events. However, there is no consistent relation among different panels or within a panel when addressing unrelated events. (Brockhoff, 2002, p. 223)

2.2.2.6 Is Delphi Measureable?

Classic Delphi produces both qualitative and quantitative results. The existence of these results invites those who would evaluate Delphi efficacy to "measure" the outcomes. For most Delphi that result in predictive outcomes, measurement oriented researchers believe that it should be possible to determine the method's usefulness by determining the difference between the prediction and the actual event. Woudenberg argues that the "accuracy and reliability of a judgment method are difficult to evaluate [as they] cannot be equated to measurements [which] can be partitioned into a true score and an error component." (Woudenberg, 1991, p. 134). The error component is composed of a number of random variables that "cancel each other out in the long run, giving the error component an expectation of zero." (Woudenberg, 1991, p. 134). A judgment appears similarly composed of true score and error except that the error "cannot be regarded as consisting of random variables." (Woudenberg, 1991, p. 134). The error in a judgment is "influenced by person and situation specific factors [allowing] ample opportunity for bias, and this bias can vary from application to application of the method." (Woudenberg, 1991, p. 134). Each Delphi is a reapplication of the Delphi process not just a reassessment of the same object. Thus, accuracy and reliability are "hampered by the possible influence of person and situation specific biases." (Woudenberg, 1991, p. 135).

Regardless of the concern relative to the ability to "measure" a Delphi, Woudenberg expresses deep concern about Delphi efficacy. After a deep review of Delphi research, he writes,
“the reviewed studies do not offer easily interpretable conclusions or unequivocal outcomes of comparisons.” (Woudenberg, 1991, p. 135).

After having expressed his concerns, he returns to the experiments that attempt to assess Delphi efficacy. He makes clear that all studied experiments to determine efficacy are not actual Delphi but laboratory experiments. Of course, laboratory experiments allow the isolation of the Delphi from the confounding influences of external activity but the experiments were not real. He writes, "in almost no study were expert participants used." (Woudenberg, 1991, p. 136). The failure to use experts is a damning limitation and a fundamental violation of Delphi methodology. A second major violation in all the studies appears as “no arguments of deviating judges were used for and fed back to the entire group.” (Woudenberg, 1991, p. 136). Delphi lives on feedback but the experimenters removed it from the experiment.

2.2.2.7 Is Delphi Replicable?

If an object is to be measured, then measuring the same object must yield the same result within the bounds of measurement error. Many assume that a “Delphi on subject X” is the same object and thus should result in the same “measurement.” However, each individual Delphi is a collection of communication acts. Each collection constructs/reconstructs its external reality – the understanding of the task, the task itself, and the outcome – as a unique set of understandings. Mitroff argues,

Since this article is about Delphi design, the important thing is not how many different realities we each have, but that one important product of each Delphi panel is the reality that is defined through its interaction. (Mitroff & Turoff, 2002, p. 35)

Measuring different panels on the same content, measuring the same panel on the same content, or measuring the same panel on different content is a completely different set of communication acts. It is entirely possible that the measuring protocol is in fact reporting truly the measurement, but the reality is that the measuring protocol never measures the same set of communication acts. Again, this summary,

Realities [as constructed by a single Delphi panel] can be described as presumed agreements which give meaning to our thoughts and make reasonable our actions in each setting. Most of these agreements about reality are implicit, and are merely confirmed and elaborated by our acts and conversations.
Sometimes our interactions subtly modify these realities. Occasionally, a group’s reality is actively renegotiated or even constructed de novo for a new situation. Delphi inquiries might produce any of these results. (Mitroff & Turoff, 2002, p. 35)

### 2.2.2.8 Results of Long-Term Delphi Forecasts

Ono and Wedemeyer (Ono & Wedemeyer, 1994) noted, “The accuracy of the Delphi technique in short-range forecasting has been proved empirically” but raised the question of the accuracy of Delphi for long-range forecasting. Having done an extensive Delphi in 1976 focused on telecommunication predictions, they revisited the outcomes in 1994 to find five of the nine events accurately forecasted. These results “lend support to the hypothesis that the Delphi is a valid technique for long-range forecasting.” They provide a number of caveats and limitations. Of interest is the importance of the items in the Delphi. They state, “The split of answers in the event questionnaire shows how difficult it was to phrase future events clearly so that the phrasing of the questions would still have meaning in subsequent eras.” (Ono & Wedemeyer, 1994). This comment is of particular insight when one considers the Sackman critique. Survey instruments demand construct validity. So, what kind of instrument claims content validity when the constructs against which it is to be measured are yet to be known?

Unlike the Ono and Wedemeyer results, Utgikar and Scott concluded nearly the opposite. A Policy Delphi on energy futures for the U.S. conducted in 1974 was revisited in 1985 and then 2000. To be generous, the results indicate that the process was unable to come reasonably close to outcomes as below,

The overall forecasts of the group can be termed as highly optimistic. The group anticipated a spectacular growth in nuclear power, predicting its share in electricity generation to be anywhere between 50% and 75% for individual nations. They predicted this growth not only through the increased capacity of traditional plants (boiling water reactor (BWR), pressurized water reactor (PWR)) but also via maturation of fast-breeder and high-temperature gas reactors. The study foresaw tremendous strides in the development fusion energy and thermonuclear power. Similar impressive growth was anticipated in power transmission technologies. Several technologies which appeared promising at the time of study (fuel cells, shale oil recovery, coal conversion) were forecast to be matured and implemented in practice in target years. The study was comprehensive and cognizant of the environmental problems including air pollution and oil spills. It also foresaw no technological or economic barriers to the conversion of waste flue gases. As discussed in the following section, the current status of the energy and environment presents a stark contrast to these prediction. (Utgikar & Scott, 2006, p. 3089)
The predictions of the Policy Delphi have not materialized and “they are not likely to materialize in the next two or three decades.” (Utgikar & Scott, 2006, p. 3090). The researchers were unable to separate underlying causes for the study error. They report that the “underlying causes of the uncertainty displayed by energy forecasts identified in this paper are: possible use of improper techniques, technological barriers to the development of technologies, socio-political considerations and economic considerations” with “economic considerations may be the single most dominant factor in determining the accuracy of energy forecasts.” (Utgikar & Scott, 2006, p. 3091). The group did a poor job in recognizing technological barriers. The role of government “regulations as highly influential in focusing or restricting technological innovations” received insufficient weight and such decisions “stifle innovative activity in a field.” (Utgikar & Scott, 2006, p. 3090). The weight of public opinion reflected through government was underestimated. For example, “Nuclear energy seems to have suffered from these factors, not achieving the potential at levels forecast by the study.” (Utgikar & Scott, 2006, p. 3090).

No obvious methodological problems with the conduct of the Delphi were reported leaving open the possible conclusion that the process does not work, or, at least it failed to identify future energy policy outcomes in this specific instance.

As noted, research assessing accuracy of long-range Delphi predictions is scant. As an example of a long-range Policy Delphi, a 1986 Canadian study focused on technology and education. The results demonstrate some accuracy when viewed in hindsight of 2008. The major findings of the study included both positive and negative potential outcomes. The study reports,

Developments which were identified as probable and undesirable included: underutilization of equipment, problems related to allocation of equipment, lack of appropriate courseware/software, piracy, confidentiality of files and information, equity, funding, resistance to change, software and hardware inertia restricting upward capability of hardware, migration of human resources, and increased and conflicting demands on educational organizations. Developments identified as desirable with little agreement regarding probability included: development of local software industry tailored to provincial curriculum, a strong emphasis on computer related in-service, an awareness of the social impact of computers, and increased involvement of parents in school matters. Major potential patterns for achieving a normative future included: increased courseware/software development and evaluation, increased levels of funding, in service, a recognition of the need for policies related to educational computing, curricular development,
preservice, planning, security procedures, and communication amongst organizations and stakeholders. (Steier, 1986, p. 1)

Considering the 22-year span of the forecast, the accuracy rate of the policy recommendations and concerns result in good predictive statements. However, keeping Sackman’s criticism of Delphi conclusions in mind relative to the blandness and superficiality of Delphi results, the above policy recommendations just as well apply to introduce previous educational technology. In fact, one could replace “computer” in this context with filmstrip, film, or educational TV and reach the same conclusions. So, as rich and interesting as this exercise shows itself to be, the Sackman concern remains standing.

2.2.3 Status and Use of the Technique.

Repeated use of an experimental or investigational technique does not establish validity. However, the nonuse strongly indicates its failure. A number of well-presented literature reviews demonstrate that the Delphi technique has increased in use from its inception and continues. The literature approaches Delphi in two ways – as the object of the research and as the technique or application. As the object of the research, the papers concentrate on the success of the technique and on what changes make the technique more or less effective. Reported use of the technique in studies demonstrates a broadening out of the fields of applicability.

In a review of the validity of use in the social sciences, Landeta notes earlier work by Gupta and Clarke:

“who carried out work between 1975 and 1994, distinguishing methodological articles from those of application, and those in which the Delphi method appeared as the main subject of the article and where it was a secondary element. The result was 463 articles, 254 of which dealt with Delphi as the main theme and the remaining 209 as a secondary element. Of the main, 75 were methodological and 179 applied”. (Landeta, 2006, p. 470)
Landeta’s own review in the period after Gupta and Clarke demonstrates continued use. Landeta reports on general and dissertation research. He provides the following on dissertation research use of Delphi (Landeta, 2006, p. 471).

<table>
<thead>
<tr>
<th>Years</th>
<th>Delphi dissertations and theses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965–69</td>
<td>0</td>
</tr>
<tr>
<td>1970–74</td>
<td>44</td>
</tr>
<tr>
<td>1975–79</td>
<td>99</td>
</tr>
<tr>
<td>1980–84</td>
<td>335</td>
</tr>
<tr>
<td>1985–89</td>
<td>335</td>
</tr>
<tr>
<td>1990–94</td>
<td>281</td>
</tr>
<tr>
<td>1995–99</td>
<td>276</td>
</tr>
<tr>
<td>2000–04</td>
<td></td>
</tr>
</tbody>
</table>

Landeta’s review of the use of Delphi in dissertations and theses work found that “the data show that, in the area of doctoral research, the technique reached maximum interest during the period of the 80s, but currently maintains a notable level of employment, relatively stable, seeming to indicate that, once the effect of fashion or novelty had been overcome, the scientific community has accepted this technique as another research technique, with present-day validity and use.” (Landeta, 2006, p. 471)

Delphi is not acceptable to the experimental purist. Delphi may also be less likely the choice of qualitative researchers who insist that no number shall disgrace their interpretivist results. Bowles reports, “The Delphi technique possesses many attributes that make it an attractive research technique. Its ability to develop qualitative and quantitative data, provide controlled anonymous feedback and its goal of consensus make it flexible, tolerant of panelist diversity and forward-looking (Bowles, 1999, p. 36).

2.2.3.1 Graduate Research

Graduate research continues using Delphi. Bowles argues “Few methods allow the researcher to use qualitative and quantitative analysis within the same study.” (Bowles, 1999, p. 33). The appeal of both qualitative and quantitative data entices graduate researchers. Skulmoski’s investigation of Delphi in graduate work notes that its “flexible research technique well suited when there is incomplete knowledge about phenomena” (Skulmoski & Hartman, 2007,
in Summary) – a common occurrence in advanced, social science research. Delphi studies in graduate research range from “qualitative to quantitative, to mixed-method.” (Skulmoski & Hartman, 2007, in Summary).

Review of graduate Delphi applications does evidence some quality improvement but there is a worry Delphi is treated as "a proven heuristic device much in the same way as Student’s t and chi-square tests represent accepted procedures in the field of statistics." (Rieger, 1986, p. 201). The assumption of the technique’s foundational validity "may be the beginning of many ‘sloppy’ Delphi applications".(Rieger, 1986, p. 201).

2.2.4 Experts, Expertise, and Panel Selection

The original RAND Delphi studies did not provide an operational definition of “expert” or “expertise.” Baker states, “Since inception [of Delphi] the reliance on experts within consensus research has been controversial. In the original studies there is no account of how or why experts were chosen, or the specific standards for selection of panelists (Dalkey and Helmer 1963).” (Baker, Lovell, & Harris, 2006, p. 60). Dalkey appears to have taken experts or expertise as an undefined term.

The Merriam-Webster dictionary defines expert as “having, involving, or displaying special skill or knowledge derived from training or experience” and expertise is the “skill of an expert.” (“Expert,” 2007). There is no reason to believe that Dalkey’s experts failed to meet this definition. Thus, at least in the original work, the concern over panelist expertise is prima facie met.

Delphi scrutinizers question the relationship of the panel of experts to the results of the Delphi process. Is the selection of an expert panel a fundamental attribute in the Delphi that determines the quality of the outcome? Is it possible that a non-expert panel will produce the same results as an expert panel? Baker writes,

Sackman’s (1975) major critique of Delphi panels maintains that expert and non-expert panels make little difference to outcomes, especially in relation to forecasting or evaluating social phenomena. It may be proposed that outcomes would be similar regardless of panel make-up. Two pieces of research have been identified that evaluate this claim. Walker (1994) made a direct comparison of two panels. One panel consisted of physiotherapist researchers and the other newly of qualified physiotherapists. Similar findings were reported between the two groups and the researcher concluded that the level of expertness required
was uncertain. Secondly Duffield (1993) explored the responses of two comparable expert panels; 93 per cent were accepted or rejected by both panels. This, it is proposed, was indicative of the reliability of a panel regardless of participants. (Baker, et al., 2006, p. 61)

The concern above is clear but the ability variations in the Delphi muddy responses. Where classical Delphi concerned itself with the assessments of future events that ostensibly demanded current, deep knowledge, the modification to other forms change the panel qualification characteristics. For example, the Policy Delphi supplants the classic Delphi with participants who are involved, who are stakeholders, who are participants, and for whom the results have direct and meaningful outcomes and not just some “guess” about future events.

The original focus of Delphi studies concerned military, science, and technology. Only a very limited number of people had any knowledge of the subject area. Delineating the attributes of “expert” was unnecessary. Dalkey had access to an excellent list and possibly the entire population of people who had requisite knowledge. The issue of identification and selection is moot if the known population of experts comprises the panel. Justification of a “representative selection” is only needed when a subset of a population stands in place of the population.

The construct of expertise and those who possess expertise demands clarity. Failure to establish clarity leaves Delphi open to those who advance the proposition that the “question as to whether responses from experts will be significantly better than those from non-experts who are ‘informed’.” (Mullen, 2003, p. 41). The literature is replete with examples of experts who are homemakers, students, nurses, doctors, police officers, firefighter, teachers, professors, and nearly every other category of person. When does a person transform from acting in a category to becoming an expert?

Suppose people exist who programmed the C-language like no others. No doubt, such people are experts. Does the deep and practiced knowledge about one language automatically extend to other languages? Suppose that the identified programmers commonly wrote business applications. Would the knowledge gained from business application programming seamlessly transfer to rendering real-time objects from laser feeds using the parallelism of a supercomputer? Is the expertise bounded by not only the language but also the use to which the language is put? More abstractly, when one knows the current C-language very well, is that same knowledge a
qualification to expound upon the future development of the C-language or other programming languages.

Does knowledge of today then qualify a panelist to make assertions of what might be? Alternatively, is the ability to make a prediction a skill and thinking process separate from having current knowledge?

Expertise is context specific but expertise may be multifaceted and matrixed. Equating expertise to certification remains simple but potentially wrong.

Baker reviews expertise and Delphi. She argues that as much as expertise remains a focal point of criticism of Delphi that little consensus emerges. She reports that Parente concludes, “there were no guidelines for defining an expert or evidence that using experts increased the accuracy of a Delphi study.” (Baker, et al., 2006, p. 62). She counters believing that “experts can be defined in a number of ways, such as their position in a hierarchy, public acknowledgement or as recommended by other participants in a study” or that “informed advocates be used.” (Baker, et al., 2006, p. 62). Operationalizing the expertise attributes of knowledge, experience, and policy influence that may initially appear as intuitively obvious – draw many challenges. For example, knowledge defined as “achieving a certain predefined knowledge and experience base” is challenged by the example that a “registered nurse would be unable to define research priorities but would be able to identify areas where they have practical difficulties”. (Baker, et al., 2006, p. 62).

Delphi’s original experts sought consensus estimates. Modified applications of Delphi seek decisions. In these cases, expertise is not the subject matter of the participant but the ability to make a decision and implement it to guide future development.

The debate over the use of experts extends to their behavior as part of a panel. Woudenberg claims that experts are “more strongly influenced by the desirability of an answer.” (Woudenberg, 1991, p. 145). Expertise and Delphi participation interact in ways researchers do not understand.
2.2.5 Selection, Expertise and Conflict of Interest

If an expert is willing to engage in a Delphi panel there must be some commitment or attachment to the outcome of the process. The interest and involvement is a source of bias. Keeney claims “respondents must be relatively impartial so that the information obtained reflects current knowledge or perceptions [but] this balance is difficult to achieve and justify to the consumers of the finished research.” (Keeney, Hasson, & McKenna, 2001, p. 196).

The selection of a panel usually falls to the Delphi coordinator. This selection process is a major source of bias. Baker reports further from Murphy, that selection should exclude “experts who are known personally to the researcher are not invited” (Baker, et al., 2006, p. 63).

Following on similar logic, Baker summarizes some of the key concerns of bias in panel selection,

The recruitment of experts is a potential for bias: participants with specific and cutting-edge knowledge in an area may have a vested interest in preventing research taking place or in manipulating the results.

The expression of conflict of interest should be a requirement of any potential panelists.

Another danger in utilizing experts in the development of clinical protocols is that those participating may possess knowledge but be very distant from clinical practice and therefore unable to articulate to practice their theory and/or knowledge.

Caution is, however, required as experts should not be judged on their representativeness but on their quality. (Baker, et al., 2006, pp. 62-63)

2.2.5.1 Random selection

Sackman’s critique of Delphi appears justified if one believes that Delphi is an experimental procedure. External validity concerns reduce when random selection from a population with random assignment to control and experimental group are applied. Mullen notes Reid’s comment that the “criticism [of Delphi] that does stand scrutiny is the danger inherent in the selection of the panel.” (Mullen, 2003, p. 41). Others cited believe that the “introduction of some basic sampling techniques with follow-ups of non-respondents would be a worthwhile innovation where the population of experts is genuinely large.” (Mullen, 2003, p. 41).
2.2.5.2 Structured Selection Panels

Delphi panel selection is a purposeful activity. Goodman suggests that content validity of a Delphi follows from a representative sample of experts in the area. Clearly, a random sample might be used to select panel members but “the premise that they will yield significantly better and substantially different responses from non-experts.” (Goodman, 1987, p. 729). Goodman warns that the intentional selection lays a “responsibility to prove and justify the selection procedures used” (Goodman, 1987, p. 730) have not introduced significant bias or limited the representativeness of the panel.

Some research indicates that panel coordinators intentionally seek heterogeneity with the belief the consensus achieved by originally highly divergent groups has more value and is more worthwhile (Baker, et al., 2006).

The structuring of the selection is heavily influenced by the type of Delphi to be conducted. As an example, the “Policy Delphi” seeks the participation of “stakeholders.” Stakeholders and experts are not synonymous constructs. The problem with stakeholders, especially where the stakeholders are members of the interested or motivated public, is that substantial education may be required to increase their fact-based participation. The changed model strives to retain expertise by insisting on participants having facts not just opinions. The introduction of education can taint the results but is also necessary to establish a baseline of “expertise.” A recent Policy Delphi conducted in Michigan related to energy choices reported difficulty with participants who had active interest but not necessarily the subject matter expertise.

Meaningful discourse (i.e. discourse that adds to participants’ understanding of an issue) is an essential process element for achieving the goals of incorporating public values in decisions, raising the substantive quality of decisions, and resolving conflicts. However, participants with insufficient experience in the subject matter were unable to engage in meaningful discourse and were therefore unable to participate effectively in the decision making process.

Participants with greater experience were able to engage in meaningful discourse, and were both able and willing to exert some control over the participation process; leading to more successful participation. Therefore, the participants’ prior experience with an issue became the critical factor in the success or failure of their participation in Michigan’s wind energy policy development. (Alberts, 2007, p. 2346)
As experts are human, they do not participate in any activity as value-neutral. The claim of being an expert does not negate agendas. Krishna and Morgan state that “experts dispensing advice are by no means disinterested” and “differing objectives among the parties may lead the experts to attempt to influence the decision maker in ways that are not necessarily in the decision maker’s interest.” They further state that “experts may strategically tailor their advice to counter that offered by another rival expert”. (Krishna & Morgan, 1999, p. 2).

2.2.5.3 Task and Expertise Interaction

Original Delphi had estimation of outcomes as the central task. Other Delphi exercises are not estimation oriented. Some research indicates that the nature of the task contained within the Delphi interacts with expertise. Rowe writes,

The utility of expertise has been reviewed elsewhere, with evidence suggesting that there is an interaction between expertise and the nature of the task, so that expertise is only helpful up to a certain level for forecasting tasks, but of greater importance for estimating tasks. (Rowe & Wright, 1999, p. 372)

Not everyone agrees with Rowe’s conclusions. Larreche and Moinpour reviewed decision-making panels in marketing. The findings indicate that “such decisions are based at least partially on managerial judgments stemming from past experience and supplemented by market research data in specific areas where empirical information is perceived to be inadequate.” (Larreche & Moinpour, 1983, p. 11). Even in the Delphi context, the results may be inconclusive requiring “judgmental decisions must be made on the basis of little or no empirical data within the prevailing time or budget constraints” (Larreche & Moinpour, 1983, p. 110).

They cite other work that indicates bias develops in the estimation of parameters and that “if the decision environment is complex, manager may have more difficulty providing the necessary estimates.” (Larreche & Moinpour, 1983, p. 119). The more amorphous the task the more assessment must rely on the collective experience of the expert.

Larreche and Moinpour assert, “selecting experts on the basis of a measure that is simple, area-specific, and externally qualified can be an effective means of identifying the most accurate judgments. However, some difficulties and limitations, both conceptual and methodological, remain.” (Larreche & Moinpour, 1983, p. 119). The difficulties Larreche and
Moinpour discuss relate to his use of MBA students as stand-ins for true experts. Again, the limitation of not using actual "experts" limits the results.

A literature search to locate a general, validated instrument to determine a priori expertise failed to produce any articles. A number of review articles point out that expert participation in processes like Delphi are time consuming for these busy, engaged people. The likelihood of cooperation to measure expertise is unexpected. Researchers have instead used simple, substitute self-reports of confidence of selected experts in their expertise as related to the focus of the Delphi. None of the literature review articles nor any of the substantive process review articles devoted significant emphasis on characterizing actual procedures used to determine "experts" and "expertise".

2.2.5.4 Identifying an Expert

Bowles asserts, "Expertise is a valid construct, but it is not easy to identify who possesses it." (Bowles, 1999, p. 32). Skulmoski advises thesis researchers who are generally inexperienced with Delphi that there are “four requirements for ‘expertise (i) knowledge and experience with the issues under investigation; ii) capacity and willingness to participate; iii) sufficient time to participate in the Delphi; and, iv) effective communication skills” (Skulmoski & Hartman, 2007, p. 4).

Larreche and Moinpour suggest,

Self-ratings and qualifying tasks are methods cited most often in the literature for separating experts from non-experts.

In some marketing decision-making situations, the expertise of the respondents has been verified by asking qualified questions with known values.

A reasonable definition of expertise involves recognition by others in the field in terms of some measurable criteria. (Larreche & Moinpour, 1983, p. 112)

For a Policy Delphi, Preble investigated the perceived need to use intra-company experts as opposed to extra-company experts in a socio-political foresighting exercise. He cites Martino as stating, “If however the forecast does not depend on knowledge of the organization but depends more on familiarity with some area of technology, then it is probably better to obtain the best people available, and, in general, these will come from outside of the organization.” (Preble,
Preble reviewed a number of other studies containing the same tacit assumption of extra-company expertise use. Preble concludes, "There was no evidence presented in these studies to indicate that any determination of inside or outside ... were like to be more qualified or expert." (Preble, 1984) In his investigation, Preble determined that the intra-company and extra-company experts develop approximately the same results. He states, "The assumption that planners should use outside experts to generate complex Delphi forecasts has been refuted for socio-political forecasting." (Preble, 1984, p. 167).

2.2.5.5 Law and Expertise

Though expert and expertise are not well qualified in Delphi studies, the construct is more clearly and operationally defined in Law. Gillotte writes,

In order to qualify as an expert, a witness must demonstrate sufficient knowledge of the subject matter through specialized training or practical experience to be able to render an opinion upon which the court can reasonably rely. Qualification of an expert is within the sound discretion of the court. Once relevancy has been determined and an expert's qualifications have been properly established, subsequent challenges to the expert's techniques and the basis for his or her opinion will ordinarily focus on how much weight the expert's opinion should be given. (Gillotte, , para. 2)

Experts at trial enjoy a unique position. Experts are the only testifiers whose opinions -- as related to their expertise -- may become part of the record. Experts express opinions on facts not in evidence but based upon their expertise. All other parties must testify as to facts and events and their opinions are usually struck from the record. Experts are allowed to utilize material "they would normally consult outside of the courtroom" and as they would reasonably rely upon that information in their field of expertise.” (Murphy, 2000, in Rule 704). An expert is allowed the unique privilege of introducing hearsay evidence in that “the expert may give an opinion, utilizing his or her research and clinical case experience, as to how individuals in his or her field would view the relevant facts in issue.” (Murphy, 2000, in Rule 702). Rules contemplate that the trial Judge may intercede to assess “whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.” (Murphy, 2000, in Rule 702). Such decision may be informed by
normal processes including peer-reviewed material, appearance in credible journals, associated error rates, and applicability of techniques to the matter before the Court.

The use of experts in Law and Science differ. Some even question the appropriateness of the use of science in the courtroom. Smith writes “law appreciates, and even venerates, finality. In fact, the entire post-conviction structure is designed to uphold an original conviction, refusing except in the most egregious circumstances to second-guess that decision.” (Smith, p. 1). Smith goes on to state that that difference between the finality of law and the indefiniteness of science would matter less if science was “getting it right.” Probably the single most quoted sentence highlighting the divergent interests of Law and Science comes from Daubert v. Merrell Dow Pharmaceuticals: “there are important differences between the quest for truth in the courtroom and the quest for truth in the laboratory. Scientific conclusions are subject to perpetual revision. Law, on the other hand, must resolve disputes finally. . .” (Smith, p. 1).

Judicial proceedings remain adversarial. Thus, the interpretation, credence, and weight of expert testimony will be challenged and interpreted in conjunction with all other rules of evidence (Murphy, 2000).

The special relationship between the Court and the expert engendered Litigation. This litigation sought to establish who is an expert and the manner in which expertise is established. The Courts face many complex issues related to experts and their testimony. Remembering that once a testifier is accepted as an expert, the testimony allowed is broad including opinions. For the Court, allowing expert’s testimony potentially allows advocacy positions of the testifier to become part of the record. The Courts usually assume experts to provide a more neutral presentation of fact and theory as related to their expertise. Rymer writes,

As judges increasingly deal with the challenges of litigation that present complex scientific questions, many express concern regarding the objectivity and reliability of parties’ experts. In a 1998 survey of federal judges conducted by the Federal Judicial Center (FJC), the education and research arm of the federal judiciary, judges were asked about problems associated with experts. “Experts abandon objectivity and become advocates for the side that hired them” was the most frequently cited problem. (Rymer, 2002, p. 6)

In Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993), the US Supreme Court “mandated that the trial court play the role of ‘gatekeeper’ as to the admissibility of expert
testimony." (Klein, Lott, & Rojas, 2007, p. 5). In order to aid the Courts in effectively playing the role of gatekeeper on expert testimony, four factors are to be considered:

“ (1) whether the theory or technique the expert employs is generally accepted in the scientific community;

(2) whether the theory has been subject to peer review and publication;

(3) whether the theory can and has been tested; and

(4) whether the known or potential rate of error is acceptable”. (Klein, et al., 2007, p. 5)

In Kumho Tire Co. v. Carmichael 526 U.S. 137, 119 S.Ct. 1167, 143 L.Ed.2d 238 (U.S.Al., Mar 23, 1999) (NO. 97-1709), the Supreme Court extended the gatekeeper requirements to “any expert testimony which contains ‘elements of science, technology, and methodology' is subject to scrutiny pursuant to Daubert”. (Klein, et al., 2007, p. 5).

Experts are not necessarily scientists and the judgment as to expertise and the testimony of the expert remain delegated to the trial court judge.

Having founded the concept of expertise from the Court’s perspective, the key element for the trial is determination of the most probable scenario by the un-expert jurors. Goodfrey writes, “Jurors are sometimes said to find or decide facts, but they actually decide probabilities. Cases are tried because there is more than one believable version of events. The jury's job is to decide who wins; its method is to evaluate which version of events is more probable.” (Godfrey-Susman, 2007, in para 7). Expertise is the fact-based opinion on which juries decide. However, Godfrey argues that “decisions juries make about experts mostly concern credibility-and not accuracy or correctness.” (Godfrey-Susman, 2007, in para 10). She believes that jurors take their responsibility seriously yet “accuracy can be relevant to credibility, because it is easier to undermine the credibility of a bad engineer or economist than a good one” with 'jurors' analyses focused more on which side they trust most rather than on the accuracy of conclusions about technical issues. Credibility is thus the key." (Godfrey-Susman, 2007, in How Does Jury Decide) In the advocacy environment of the trial, the final determination rests with “presenting expert evidence should aimed at making your expert more credible than your opponent's expert.” (Godfrey-Susman, 2007, in How Does Jury Decide).
With all the constraints, rules, and expectations placed on expertise, the operational result at trial is that “credibility” trumps facts. Many want to believe that facts standalone in their appeal. There is no more salient a demonstration of the opposite than the infamous OJ Simpson murder trial in which expert DNA evidence linked the defendant to the murder. However, defense attorneys attacked witness credibility asserting racism, bungled evidence control, and a trapped prosecutors with the infamous glove that would not fit. OJ Simpson was found innocent in contradiction to all the supposed expertise.

The digression above to discuss the legal basis of “expert”, “expertise”, and “expert testimony” informs the panel selection process of Delphi. The Courts operationalized the construct and re-examined it. The adversarial nature of trial introduces a check and balance on claimed expertise through the cross examination of the credentials and biases of the expert prior to allowing the expert’s testimony.

2.2.5.6 Experts and Behavior

Should panel selection be influenced by behavior of experts? Krishna investigated the composition of a set of advisors and their behavior to determine who belongs on a panel so that a decision-maker gets the best advice. Should those with similar ideas be included? Should those with divergent ideas be included? Should more than one of the similar mindset be included? Krishna asserts that “experts dispensing advice are by no means disinterested” and “decision makers are bombarded with advice from numerous experts, with possibly different agendas [who] strategically tailor their advice to counter that offered by other, rival, experts.” (Krishna & Morgan, 1999, p. 2). The inherent interest of the expert is a consideration in panel selection. He claims “Self-interested experts influence the decision process by strategically withholding information that they possess.” (Krishna & Morgan, 1999, p. 21).

For panel selection, the selection of more than one expert with similar biases is detrimental. The weight of multiple panel members with similar dispositions tends to cause the outlying panel member to “strategically respond by reducing the precision of the information he conveys.” (Krishna & Morgan, 1999, p. 21). Fewer advisors on a panel with similar dispositions reduce this threat. Finally, Krisha concludes,
With opposing biases, even bad advice can enhance good advice. Thus, a cabinet composed of advisors from opposite sides of the spectrum is superior to a single advisor. But there are limits to how much additional information may be garnered from a cabinet. Full revelation is still not possible. Moreover, if the cabinet consists of opposing extremists, no information is conveyed. (Krishna & Morgan, 1999, p. 21)

2.2.5.7 Judgment of Experts versus Laymen

Wright re-examines earlier research comparing “the veracity of expert versus lay judgments of the magnitude of risk” of the risks posed by life-threatening events. Wright cites earlier work by Slovic et al. that concluded “that the judgment of their experts differed substantially from non-expert judgment primarily because experts employed a much greater range of values to discriminate among the various hazards.” (Wright, Bolger, & Rowe, 2002, p. 1108; Wright, Rowe, Bolger, & Gammack, 1994). He continues with more studies concluding, “There are substantial differences in the way that experts and samples of lay people judge risk.” However, literature reports by others indicate sharp contrast with the previous findings. These findings support the view of little difference in judgment. The cited study of Bolger argued that “in many real-world tasks, apparent expertise may have little relationship to any real judgmental skill at the task in question.” (Bolger & Wright, 1994a, 1994b, p. 1109). To explain the contraction, the task’s nature may play a role. Bolger classified tasks as “ecologically valid” or “learnable.” The former tasks are those for which “experts were required to make judgments inside the domain of their professional expertise and/or express judgments in familiar metrics” as opposed to the later that “meant the degree to which it is possible for good judgment to be learned in the task domain.” (Bolger & Wright, 1994a, 1994b, p. 1109). Feedback to influence performance appears as a mediator. The lack of feedback retards the ability to improve performance if experience remains unchallenged. For example, tasks that are easily learnable yet are also within the domain of the expert tend to show lesser difference between experts and nonprofessionals.

The Wright experiment concluded “experts were a little better in their risk judgments than lay persons and that expertise did make some difference [but] the differences in performance between experts and lay persons were small in magnitude the nature of biases were common to both groups”. (Wright, et al., 2002, p. 1107; Wright, et al., 1994). The suggestion existing in the literature is that expertise will manifest when the task is “ecologically valid” due to its natural and
deep relationship to the expert's expertise. In their experiment the tasks were held to be highly "ecologically valid" yet the outcomes did not evidence of the experts outperforming nonprofessionals.

2.2.5.8 Use of Information by Experts

Shanteau reviews expert decision-making against the information-use hypothesis that infers "The amount of information used, as measured by number of significant cues, should be greater for experts than non-experts." (Shanteau, 1992, p. 2). His review of the literature demonstrates that either experts or naïve decision-makers use few cues thus implying similarly informed decision-making. After further review of more recent literature, Shanteau explains that the information-use hypothesis must consider that the "amount of information used does not reflect degree of expertise; however, the type of information used does" as "evidence from these studies is consistent: judgments of experts and mid-level novices are based on similar amounts of relevant information. Although the methods vary, the results consistently reveal novices use as much or more information as acknowledged experts." (Shanteau, 1992, p. 3). Expertise does not appear directly related to the number of cues.

Shanteau infers that there is "evidence to suggest that novices may rely on too much information and that experts are better because they are more selective." (Shanteau, 1992, p. 5). He argues that it appears that experts detect relevant information. The larger implication relates to the attempt to determine "expertise generically without reference to specific decision contexts expertise." (Shanteau, 1992, p. 6). If context remains essential, then attempts at a general definition of expertise must fail. Shanteau believes that concentration on the number of significant cues results in overlooking "what makes experts special – their ability to evaluate what is relevant in specific contexts." (Shanteau, 1992, p. 6).

2.2.5.9 Optimism of Experts

Self-rating of expertise is believed to be one approach to solving the conundrum of determining expertise for a given investigation. The question remains – is self-rating of expertise sufficient to identify expertise? Tichy summarized much of the current research to indicate that "experts" tended to rate themselves highly but the actual performance may not support the hoped
for conclusion. He cites many well-known studies in support of the self-rating and performance yet others that “do not see significant differences or tend even towards the opposite, giving top experts less reliability.” (Tichy, 2004, p. 343). Tichy points to fundamental weakness in the research approach that attempt to validate self-rating and expert performance. The usual model is to use undergraduates who perform self-ratings who then answer questions from almanac data. He argues that such a design is not consistent with what experts do as “the students in the experiments answering almanac questions are not involved, while experts assessing future developments in their field of activity definitely are.” (Tichy, 2004, p. 343). The evidence is fragmented. An explanation known to risk research is “unrealistic optimism may result from the overestimation of one’s own capabilities and the underestimation of risks inherent in one’s own work.” (Tichy, 2004, p. 344). He goes on, “Perceived controllability, commitment, and emotional investment are indeed typical constellations influencing an insider’s points of view. Experts pushing ahead the advance in their specific field are fascinated by their task and they must believe in its significance and in its future. They are strongly influenced by the desirability of the outcome and they believe in their ability to influence it.” (Tichy, 2004, p. 344).

Experts tend to be fascinated with their technology. These experts “tend to reduce complexity by closing their eyes to the fact that the introduction of a new technology entails a complex of innovations not a single technical innovation. Diffusion periods are heavily underestimated, especially for major innovations and implementation is assumed to be smooth and devoid of serious obstacles.” (Tichy, 2004, p. 344)

Tichy reports findings that highest self-ratings of expertise “tend to give the most positive of assessments on all topics raised.” (Tichy, 2004, p. 353). The Delphi in which the experts participated over optimistically estimated the period of implementation. The results tended to support the belief that strong self-reporting of expertise corresponded to overly optimistic predictions. Tichy found that the work environment influenced experts in that those in business tended to be much more likely to reflect optimism than did those in academia or in administration. Tichy concludes that “over optimism among top experts, in general, and among insiders in particular, foresight exercises should base their panels on a fair mixture of experts of different
grades, with different types of knowledge and affiliation, and not only on top specialists in the respective field” (Tichy, 2004, p. 360).

Many Delphi panels ask experts to rate their own expertise given that they are “experts”, but Murphy cites Mullen’s concerns on the risk of self-rating, as little is known as to how the rating influences results. Woudenberg concludes that accuracy and self-rating of expertise produce correlations that may be negative as well as positive (Woudenberg, 1991, p. 139).

Self-confidence does not necessarily imply best knowledge. Larreche and Moinpour believe that “experts can be affected by what amounts to uncontrolled expert halo effect” (Larreche & Moinpour, 1983, p. 112) requiring an external assessment of actual expertise. They continue by arguing that Delphi processes appear to improve estimates but that self-rating of expertise does not appear to contribute to better estimates. They attempt to provide an explanation,

This finding may illustrate the fact that highly confident individuals may be less willing to update their initial judgments and consequently draw the other members of the group toward their points of view even in the absence of interpersonal interactions. Indeed, if the Delphi technique successfully eliminates biases due to face-to-face interactions, it is still influenced by the willingness of individual judges to revise their estimates in the light of new information. (Larreche & Moinpour, 1983, p. 112)

Research by Linstone, in a laboratory setting, indicates that self-identification of expertise on successive rounds of Delphi improve the outcome accuracy,

In order to improve accuracy, a number of experiments examined self-rated expertise to sub-group the panel as the rounds of Delphi continued. Though the reduction in panel size caused by sub-grouping appears as a concern, the investigators conclude “(1) self-rating is a meaningful basis for identification of expertise, and (2) selection of expert subgroups improves the accuracy to a somewhat greater degree than does feedback or iteration” (Brockhoff, 2002, p. 228).

Self-confidence in one’s expertise and accuracy of judgment need not be correlated.

2.2.6 Panel size

Panel size is yet another contentious differentiation between experimental and qualitative Delphi. In qualitative research, there is no proper sample size except that justified by the context of the investigation. In experimental research, quantitative, statistical calculations are done.

Nothing precludes Delphi from using the quantitative calculations except neither random selection
nor random assignment are necessarily used following a determination of panel size. Thus, the value of calculating a panel size statistically has little value.

Bowles (Bowles, 1999) believes that Delphi can bring a large number of people together to enhance reliability. Of course, a “large number of people” does not satisfy experimental design expectations. He reports that Linstone and Turoff thought anything between 10 and 50 would work. He notes that the largest recorded panel prior to 1988 was 1,685 individuals. More recent panels are even larger — with several thousand participants.

Akins investigated Delphi panel size reported in the literature. He found that a “small” panel was usually no less than 10. Though no descriptive statistics are reported, he summarizes that “panels consisting of 10 to 100 or more panelists” (Akins, Tolson, & Cole, 2005, p. 2) are most common. He reports the concerns about panel size and error in forecasting and comments relative to the statistical-bound calculations of appropriate, representative panel size,

Although experiments carried out in the 1950’s and 1960’s suggested that group error is reduced with increased group size, the sample size for constructing a Delphi panel is not a statistically-bound decision and good results can be obtained by a comparatively small group of homogenous experts. (Akins, et al., 2005, p. 10)

Akin’s summary of the status of the panel size discussion is,

In general, the confusion around the Delphi sample arises from the fact that there are no standards established in any methodologically acceptable way. The current literature presents only empirical choices on Delphi expert sample sizes made by individual researchers, such as convenience, purposive or criterion sampling. (Akins, et al., 2005, p. 2)

Akins research in health care use of Delphi found that “reliable outcomes could be obtained with a Delphi panel consisting of a relatively small number of experts selected via strict inclusion criteria.” Only when the experts have “similar training and general understanding in the field of interest allow” can the conclusion of stability of results using a small group be asserted (Akins, et al., 2005, p. 3).

The Delphi does not call for expert panels to be representative samples for statistical purposes. Representativeness, it seems, is assessed on the qualities of the expert panel rather than its numbers. Confusion on in press abounds in the nursing literature: Williams and Webb (1994) for example, criticize a number of Delphi studies for their lack of ‘random samples’ (p. 182) (Powell, 2003, p. 376)
In contrast, Wild and Orgersen (2000) suggest panel sizes of 300 – 500 are usually considered sufficient. Cantrill et al. report that “published studies in health applications have used panel sizes from 4 to 3000” and recommended that size “should be governed by purpose of the investigation” (as cited in Mullen, 2003, p. 41).

2.2.7 Feedback and Feed-In and Iteration
Feedback is believed to compel change by allowing Delphi panelists to re-examine thoughtfully their beliefs. The feedback and feed-in features of Delphi remain a key attribute differentiating the technique. Rowe argues that anonymous feedback “allows participants to focus upon the content of that feedback rather than being distracted by extraneous social information, so being influenced by ‘good’ information and uninfluenced by apparently ‘poor’ information” (Rowe, Wright, & McColl, 2005, p. 378).

However, Sackman believes that feedback – especially feedback in statistical form expressed as tendencies toward a mean – corrupt the independence of thought processes of the panelist and it is the divergence and thoughtful ideas that Delphi values. Sackman writes,

Sackman argued that the Delphi system militates against independent judgment because it is destroyed once the panelists know how others have responded to each item. (Goodman, 1987)

Rowe et al. investigated opinion change and judgmental accuracy as feedback and rounds continued for a Delphi. They found that accuracy “increases over Delphi rounds when statistical summaries or written rationales are provided from other members of an individual’s nominal group.” (Rowe, et al., 2005, p. 377). Delphi with only iteration but not feedback did not increase in accuracy. If a subject made a reasonably accurate prediction in early rounds then they tended not to change substantially in later rounds. However, “convergence of opinion over Delphi rounds will not necessarily imply improved forecasting accuracy in every case” (Rowe, et al., 2005, p. 397).

To isolate the effect of written feedback content and statistical consensus feedback, they found that “majority opinion exerts strong opinion pull on minority opinion even when the majority favors an incorrect answer” (Rowe, et al., 2005, p. 377).
Unfortunately, the determination of the role of feedback in opinion change and task forecast accuracy is heavily confounded by the additional feature of iteration used in Delphi. Rowe states, “After all, it is possible that Delphi panelists may ignore feedback and still improve, by using the opportunity afforded by each new round to review their own previous estimate.” (Rowe, et al., 2005, p. 378). Rowe discusses the findings of Parente and of Boie and Mrnighan that appear to support the view that iteration independently explains change. He summarizes, “These results seem to suggest that it is iteration that is the more powerful influence leading to improved accuracy, and not feedback.” (Rowe, et al., 2005, p. 378). Rowe’s review of the particular research questions the conclusion because the feedback in Parente was “simply comprised of means and medians” and Boje and Murnighan provided “estimates and justifications without average.” (Rowe, et al., 2005, p. 397).

From these studies, it may be suggested that Delphi ‘works’ partially because of the influence of iteration, allowing panelists to reflect on previous answers, and partially because of the influence of feedback, which is most likely to be successful when more rather than less informative. The output from a Delphi process, however, is an aggregate of the individual judgments of the panelists, and this research says little about who among the panelists is affected by what aspects of the technique and in what way. Understanding the mechanics of judgment change in Delphi is likely to have implications, not only for the best structure of Delphi (e.g., best way to provide feedback), but also for the selection of panelists. (Rowe, et al., 2005, p. 398)

Unfortunately, no clear guidance is available concerning feedback and iteration. Rowe suggests that “the practitioner should think carefully about the nature of feedback they provide, what information it might contain, and how their panelists might react to it, inasmuch as a variety of studies (including this one) give contrary results as to whether iteration alone, the feedback of statistical averages, or the use of panelist arguments will lead to most improved performance over rounds” (Rowe, et al., 2005, p. 397).

Schiebe attempted to isolate the role of written feedback and consensus data expressed as means of group responses. Schiebe intentionally provided incorrect feedback expecting the group reasoning process to reject the false feedback. Instead, the group moved toward the errant conclusion supported by the false feedback. Goodman presents a finding that raises many concerns,
This finding raised worrying questions about what could be taken as the manipulation effect of the feedback mechanism. Does this imply that panelists are persuaded towards conformity as opposed to true agreement? (Goodman, 1987, p. 733)

If consensus is driven by the feedback of means but not the feedback of content, then what role does feedback play?

Feedback procedures in Delphi assume that “individual judgments may be improved and debiasing may occur.” (Rowe & Wright, 1999, p. 370). Unfortunately, feedback may be false or superficial.

In one experiment the forms of feedback – statistical and reasons – affects showed that “the greatest degree of improvement in accuracy over rounds occurred is the ‘reasons’ condition. Furthermore, we found that although subjects were less inclined to change their forecasts when receiving ‘reasons’ feedback than the other types, when they did change their forecasts they tended to become more accurate – which was not the case in the ‘iteration’ and ‘statistical’ conditions.” (Rowe & Wright, 1999, p. 371).

Using physicians in an experimental design, Campbell investigated homogeneous versus heterogeneous panel membership and its relationship to feedback and decision-making. She concludes “that panel members from different disciplines make systematically different judgments and that feedback from mixed disciplines may further influence results.” (Campbell, 1999, p. 966).

To date, searches for experiments that assess the effect of questions asked by panelists or the external feed-in of almanac data do not exist.

2.2.7.1 Decision Regret

Feedback in a group process need not be the explicitly allowed form of Delphi. Of course, direct interaction and the human factors associated with such interaction are specifically controlled but not all forms of human feedback can be excluded. For participants the Delphi process corresponds to a decision-making with regret consequences. Regret theory postulates that “decision maker’s fear for future regret influences their behavior” and that “regret induces us to make more ‘rational’ choices; that are, anticipated regret causes people to think more elaborately before making a final decision.” Zeelenberg argues
People are motivated to avoid or minimize post-decisional regret. The motivation exerts impact on their decisions because the possibility of future regret is anticipated and taken into account when making decisions, and because experienced retrospective regret promotes decisions that make this regret disappear. As a result people can become risk averse or risk seeking, depending on which of the possible choice options is the regret minimizing option. In principle this can be considered rational, because it protects the decision maker from some of the aversive consequences of decision making. Now that we have establish that regret does influence decision making and that feedback is one of the prime determinants, future theorizing could focus on the normative stations of these effects. (Zeelenberg, 1997, in Conclusions)

Feedback is the chief variable contributing to regret as decision-makers compare “obtained decision outcomes with forgone decision outcomes.” If feedback were never received, then there is a significantly lessened potential for regret – ignorance is bliss. Thus, the absence of feedback eliminates the comparison of obtained and foregone conclusions and thus eliminates regret.

Regret counters the view that rational reconsideration of a position takes place during feedback rounds in a Delphi.

The most preferred alternative is not necessarily superior to another alternative. The idea is that when there is one dominant alternative the decision maker does not spend much time thinking about the possible drawbacks of this alternative, because there is less self-recrimination when the obvious superior alternative results in a suboptimal outcome. (Zeelenberg, 1997. p. 9)

The time delay between a decision and the consequences of the decision are sufficiently substantial, regret is reduced.

Delphi rounds contain feedback but not feedback containing any consequence except anonymous challenge by other members or self-evaluation of deviation from the “average.” The lack of specific “post-decisional feedback” is a central determinant of experienced and anticipated regret. When this feedback is present people anticipate possible regrets, but when it is absent regret does not play a significant role in the decision process” (Zeelenberg, 1997, p. 20, 1999).

2.2.8 Attrition

Attrition in any study is a risk to its integrity. Many features of the Delphi increase the risk from attrition. The declining pool of participants introduces uncontrolled and unexpected bias. For example, if panelists drop out but had earlier provided feedback affecting the group decision process then the advocate is no longer there but the effect is. Some Delphi report that those with
the most extreme or non-norm feedback tend to drop out as the panel tends toward consensus – a consensus with which extremes would not agree. Unlike many experiments with a one-time intervention and thus a reduced time commitment, the repeated rounds of Delphi extend the time commitment thus increasing attrition risk.

Greatorex discusses the issue relative to opinions and participant connectedness,

Dropout may occur if minority opinions are not explored and therefore the resultant consensus may not be representative of all the experts. Other reasons for panelists dropping out of a Delphi study include: low motivation; disagreeing with the design and content of the study; and lack of faith in the initial results.. (Greatorex & Dexter, 2000, p. 1020)

Bowles looked at the issue of attrition and found that in reported studies the attrition appears high. One solution proposed was “large panels to protect the study from wastage.” (Bowles, 1999, p. 32). The worst case attrition rates exceed 50% and go as high as 78% (Bowles, 1999, p. 32).

The issue of biased results due to attrition receives some attention. Whenever a survey or survey-like instrument is used there is a risk that, “those who do not participate are different from the sample [introducing a] potential bias in the findings “ Baker reports that her review determined that only “Limited research has attempted to establish whether some experts are more likely to participate, whether there is a reason for this and what the effect is on the results.” (Baker, et al., 2006, p. 66). Of the research conducted on the attrition, the researchers concluded, “that expert panels were very similar in characteristics to their colleagues but were unable to identify further research in this area.” (Baker, et al., 2006, p. 66).

Mullen notes concerns expressed concerning the introduction bias resulting from low response rate. Walker and Selfe believe that to maintain rigor, “a 70% minimum response rate should be achieved” but offer little support for this claim. However, as Reid notes, evidence from most studies is that the larger the panel the higher the drop-out rate – with panels of 20 tending to keep their members (Mullen, 2003, p. 41).
2.2.9 **Anonymity**

Group decision-making is notorious for the effects of group dynamics. Delphi assumes that the depersonalization during the feedback stage allows each panel member to participate without threat of criticism and mockery. Keeney summarizes,

> This promise of anonymity facilitates respondents to be open and truthful about their views on certain issues, which in turn provides insightful data for the researcher. Furthermore, Couper (1984) suggests that this provides each participant with an opportunity to express an opinion to others without feeling pressured psychologically by the more influential panel members. (Keeney, et al., 2001, p. 197)

Others counter that anonymity does not allow close cross-examination of submitted feedback. Bowles reports that “Milkovitch et al. (1972) claim that valuable data are lost because panelists cannot interact directly, limiting the extent to which exploratory thinking is possible. They advocate a more adversarial interpersonal process” (Bowles, 1999, p. 32).

Goodman asks if the detachment contributes to feedback that would be more carefully constructed when a person feels they may have to defend or explain their position. Goodman writes,

> There is an inherent danger that this anonymity leads to a lack of accountability for the views expressed. Considered comment or in-depth analysis of the items under discussion may not be encouraged. The result may be hasty, ill-considered judgments since panelists know there will be little or no comeback. (Goodman, 1987, p. 730)

Woudenberg wonders if the obviating of “some intrinsic positive effects of unstructured, direct group interaction (the richness of nonverbal communication)” has negative effects on decision quality. Determining the anonymity effect is unclear but some research on satisfaction of participants with anonymity “varies strongly” (Woudenberg, 1991, p. 139).

True anonymity is not assured in many Delphi panels because the panel director will see all responses. In addition, panel members are typically drawn from an expert population that is “like being in an elite ‘expert’ club where the membership is known but they do not meet face to face to discuss the issues.” At best, there is quasi-anonymity in most Delphi studies.

Ultimately, the importance of may be important to keep members participating but “The influence of anonymity upon findings has not to the authors’ knowledge been reported in the Delphi literature” (Keeney, et al., 2001, p. 197).
The relationship between anonymity and Delphi efficacy is not yet established.

2.2.10 Researcher Bias

Two major sources of researcher bias appear in Delphi – the conduct of the process and the interpretation of the results. The researcher bias is not unlike any other qualitative study. Reid argues, “Delphi studies may have more validity and reliability than many other qualitative research methods.” (Bowles, 1999, p. 33).

Thus, the best guidelines to control for researcher bias would be the same as those in any other qualitative study. Maxwell reminds, “there cannot in principle, be such a thing as a “God’s eye view”, a view that is the one true objective account. Any view is a view from some perspective and therefore is shaped by the location (social and theoretical) and ‘lens’ of the observer.” (Maxwell, 2005, p. 38). Though this philosophic argument that bias is human and intrinsic must still address “how to incorporate this experience most productively in your research design or how to assess its effect on your research.” (Maxwell, 2005, p. 28).

Delphi is likely to converge to consensus and the convergence is likely to create optimism in researchers – a bias – toward belief in the findings. Some believe this may “may predispose participants to action” (Bowles, 1999, p. 33) instead of producing findings. Delphi might itself promote change. The self-reinforcing use of the technique leads to “a high degree of confidence in its value.” Yet, few studies are repeated or revisited “by the same authors or others – to validate, confirm, or negate the findings of the original studies. There is no corroborating evidence to suggest that the original Delphi studies have contributed to useful, accurate, or timely policies and strategies, other than the reports of the original studies.” (Bowles, 1999, p. 34). Again, like any other qualitative study the expectation of revisiting or testing is not inherently part of the design.

2.2.11 Idea Generation

The prerequisite to decision-making is generation of alternatives. This initial phase compares to Delphi’s open-ended item generation. Jackson reports on Osborne’s classic work on idea generation procedures as " (a) no criticisms or judgment of ideas, (b) free-wheeling wild
ideas are welcomed, (c) quantity is wanted and (d) combination and improvement of ideas is sought" (Jackson & Poole, 2003, p. 562).

Valacich et al. investigated the issue of idea generation in the context of structured communications environments – in this case nominal groups -- and general “interacting groups.” In the literature review they state, “in no study have interacting groups using any structured or unstructured technique ever generated more ideas than equivalently sized nominal groups, and in virtually all cases, nominal groups generated significantly more ideas than interacting groups.” (Valacich, Dennis, & Nunamaker, 1992, p. 50). They further report that the “number of ideas generated by nominal groups typically increases with group size and this has not generally been found to occur in interacting groups.” (Valacich, et al., 1992). They attribute these differences to “process loss” defined to be “production blocking [can’t get your thoughts heard], free riding [groups don’t push for your participation], and evaluation apprehension.” (Valacich, et al., 1992, p. 50).

Valacich et al. conducted a number of controlled experiments using a computer supported nominal grouping procedure. They found clear evidence that the computer supported context resulted in more idea generation. Moreover, “computer mediation [enforcement of the structured communication rules of nominal groups] enabled relative large interacting groups to outperform smaller groups.” (Valacich, et al., 1992, p. 51). Anonymity “did not enhance group outcomes.” (Valacich, et al., 1992, p. 63) Smaller groups seemed more “satisfied” and “perceived themselves to be more effective than groups from other conditions.” (Valacich, et al., 1992, p. 64).

They suggest that anonymity plays a role in idea generation but anonymity is not a constant. They suggest there are various forms of anonymity at work. They state, “These findings, combine with findings from previous group research, lead us to suggest that anonymity is not simply either present or absent in computer-mediated environments; instead, there may be various levels and types of anonymity.” (Valacich, et al., 1992, p. 64). For example, content anonymity is the “ability of a group to identify the source of a specific contribution” while process anonymity refers to “the ability of group members to determine who is participating.” (Valacich, et al., 1992, p. 65).
From this discussion, the anonymity assurance given to participants should be directly addressed in the solicitation of participation.

Jackson investigated the role of group decision support system in idea generation. All idea generation procedures – computer-based or not – are "conducted within a controlled situation in which one or more central persons guide and structure the group process" (Jackson & Poole, 2003, p. 561) with technology often used in such a process. Their experiment concluded, "Medium does make a difference" when a technology based system was compared to a paper-and-pencil exercise.

2.2.12 The Questionnaire – Construction and Scoring

In most Delphi, when the items of concern are identified for inclusion for a Delphi investigation, the items are listed on an instrument that appears to be a survey instrument. The instrument is used to collect feedback from the Delphi panel. The instrument items are usually either rank ordered or responded to in a Likert-type scale.

Sackman was very clear – if this instrument is a survey instrument and statistics are going to be returned as "feedback" then he believes the researcher is required to consider all the issues of creating a valid instrument. Others agree. Some actually attack the Delphi process at this point for being nothing more than a weak implementation of a survey. Mullen counters Sackman claiming that sloppy design of instruments is not unique to Delphi,

Although some Delphis have been criticized for poor questionnaire design, similar criticism could be equally directed at poorly design questionnaires used in conventional surveys. (should be professional) But, Delphi does give the researcher or monitor an unusual degree of power. (Mullen, 2003, p. 44)

Those who analyze the survey instrument problem believe that the researcher must justify any variation from professional standards.

However, there is at least one major difference between a survey instrument and a Delphi instrument. A survey instrument is created by the examiner following all the “rules.” A Delphi instrument is created by the Delphi panel itself. There are Delphi that use a survey instrument created by the researcher and those Delphi should carefully contrast and compare the process by which the instrument was constructed with professional standards. In a Delphi, the
first round is more than likely an item identification round. The items have construct validity within the group but may fail on other measures. Turoff argues that,

The statements which comprise the elements of a Delphi exercise inevitably reflect the cultural attitudes, subjective bias, and knowledge of those who formulate them. (Brockhoff, 2002, p. 224)

Powell claims that questions originating with the panel itself lead to “richness of the data collected.” (Powell, 2003, p. 378). His review indicates that some Delphi use semi-structured initial questions. Sometimes the instrument used is borrowed from other research or survey work. However, “Delphi purists reproach such modifications” (Powell, 2003, p. 378).

If the open-ended approach is used in the first round, then the experimenter or monitor team must “structure questionnaire [from the open ended data] which is circulated to the panel in round two”.

Mullen reports that first round open-ended development of concerns remains central to successful Policy Delphi.

Judge and Podgor, in their study using Delphi in a citizen participation project conclude that allowing “respondents to make the first round choices with a seed … will assist in developing a set of choices more representative of the wants of the participants.” (as cited in Mullen, 2003, p. 44)

Mullen makes one point very clear: “open-ended first round has been used as a criterion for judging whether a study is well-conducted” [emphasis added]” (Mullen, 2003, p. 44). Though her findings also indicate that “initial questionnaire is developed by the researcher, often after an extensive literature review, or by a sub-panel of “experts” or a monitor team.” (Mullen, 2003, p. 44).

Turoff et al.did some investigation on panel response to items on the feedback instrument. Their concern is with “Statements may be too concise, leading to excessive variations in interpretation, or too lengthy, requiring the assimilation of too many elements. Consequently, we would expect a constraint on the number of words leading to the widest agreement in interpretation.” (Brockhoff, 2002, p. 226). Statement length and complexity appears to increase consensus in non-expert panels while experts "come to very high consensus with moderate statement lengths but fall to a very low level of agreement with long statements.”
Minutiae, systemic in longer questions, appear to cause experts to raise more concerns about the content and meaning of the item.

2.2.13 Consensus, Rounds, and Accuracy

As previously argued, the Delphi results in a consensus. The consensus process is a negotiation recognizing content outcome agreement. Consensus should increase as the negotiation agreement nears. Closure should result in the “many are better than one” outcome – the more accurate outcome than any one single member of the group might propose. In a Delphi, consensus, the number of rounds, accuracy and all the other factors previously discussed all interact. Separating the process into constituent components to be manipulated independently of other components challenges the best researcher.

The fundamental assumption is that a highly agreed to outcome is the most accurate representative of the “answer” sought. If the Delphi were a prediction of dates when an event might occur, then the dates proposed with the highest consensus would be the most accurate estimate. If the Delphi sought political agreement, then the proposal garnering the strongest consensus is the best alternative.

Advocates of Delphi see implicit correlation between accuracy and consensus. This assumption is directly challenged. Woudenberg refutes the consensus and accuracy interaction,

Although consensus can be important, it can never be the primary goal of Delphi. **High consensus is neither a necessary nor sufficient condition for high accuracy.** [Emphasis added] In most Delphi, a slight increase in accuracy over rounds is found. In contrast, consensus increases very strongly. Also, a direct comparison shows that consensus increases more strongly than accuracy. Together with indications that group pressure to conformity is very strong, this makes consensus in a Delphi suspect and in no way related to genuine agreement. (Woudenberg, 1991, p. 146)

To achieve consensus in a single Delphi feedback round is an unlikely event. The opportunity to revise responses is fundamental to Delphi and “Such provision obviously requires two rounds. Beyond that the number of rounds required is disputed and individual studies have been found using two, three, four, and five rounds.” (Mullen, 2003, p. 46).

Delphi is not a closed system assuring closure toward consensus but all the literature found and reviewed for this dissertation find closure. Woudenberg found Delphi to be “extremely efficient in achieving consensus” and for estimating tasks the “consensus is almost always
maximum after the second round.” (Woudenberg, 1991, p. 146) Unfortunately, the operational definition of consensus remains a challenge. Bowles reports that Williams believes “the definition is often arbitrary and post hoc.” (Bowles, 1999, p. 32).

The issue for research is – how does one measure consensus as the representative of closure?

Measuring consensus relates directly to the number of rounds in a Delphi. A researcher must know when a point of diminishing return is reached so that the process can be terminated. However, the rounds should not be stopped so early as to prevent reaching a “true” agreement. Schmidt states,

First the researcher must know when to stop polling. If the polling stops too soon, the rankings may not be meaningful; but, too many rounds would tax the researcher’s resources and was the panel member’s time. Second is how many items to carry over to subsequent rounds. Too many items for ranking can cloud consensus. Third, researchers have not made good use of available statistical techniques to support their conclusions. (Schmidt, 1997, p. 764)

Fink, writing about consensus methods in the context of health, writes, “Many Professional Standards Review Organizations have relied on consensus methods to help choose among the many areas of medicine that might be justifiable subjects for evaluation and to set standards of quality.” (Fink, Kosecoff, Chassin, & Brook, 1984, p. 979). They focused less on the predictive items within a Delphi but more upon the model as a source of aggregation and consensus building. They provide a number of examples, one of which notes Delphi use to reach a consensus to select “preventive treatment for isoniazid-resistant tuberculosis.” (Fink, et al., 1984, p. 979). Given all the limitations of Delphi as a consensus building technique, the need to create credible consensus remains paramount. To this point, they write “the supporters of consensus appear to be the more vociferous, and the use of consensus appears to be increasing.” (Fink, et al., 1984, p. 979). Hearnshaw writes, “Consensus methods are increasingly used to develop clinical guidelines.” (Hearnshaw, Harker, Cheater, Baker, & Grimshaw, 2001, p. 174). Fink et al. make it clear, “health and medicine should be based on definitive information obtained from scientifically sound studies.” (Fink, et al., 1984, p. 981). The problem for health professionals remains the translation of basic research to practice as research seldom leads to actionable results. Instead, research is combined with other research and other knowledge and
opinion. The result is “The hope for formal consensus techniques is that, when properly employed, they will create an environment in which experts are given the best available information, and will allow their solutions to problems to be more justifiable, valid, and credible than otherwise.” (Fink, et al., 1984, p. 981). Fink goes on to suggest that the use of consensus in health care demands that “decisions should be justified by reliance on available empirically-derived data as well as on judgments and experience”; “participants should qualify for participation because they are representative of their profession, have the power to implement their findings, and are not likely to be challenged in their field”; “objective and skilled leaders should administer the consensus process”; “consensus results should represent clear and specific actions”; and, all findings of the panel can be “improved through careful dissemination of the findings” with an attendant monitoring capability to assess the value of the consensus. (Fink, et al., 1984, #685).

Greatorex proposed a graphical representation of the convergence in successive rounds. He reports that by reviewing the graphs between rounds that summarize item scores that one can determine the point at which convergence toward consensus begins and when it is reached. He observed convergence but did not explain underlying causes of convergence. He states,

Whether the increase in agreement is because the feedback has worked in a constructive way to help the experts who are out of line with the consensus to refine their judgments, or whether such experts have just conformed to the majority view is difficult to tell and is a conundrum of the methodology. One approach to decide whether it is the Delphi process itself which induces the conformity would be to identify the conformist panelists who abandoned the majority of their original opinions, and remove them from the analysis. (Greatorex & Dexter, 2000, p. 1022)

Interestingly, Greatorex suggests that the observations between rounds allow the detection of participants not moving toward the consensus but that “damaging to the decision making process as experts who always conform.” (Greatorex & Dexter, 2000, p. 1022).

As much as this section concentrates on consensus, many argue that consensus brings with it “a danger that possibly important variations in views will be concealed – a risk that extreme opinions will be masked by statistical analysis.” (Mullen, 2003, p. 48). The panel could reach strong consensus but the answer could be completely wrong. Even unanimity does not
guarantee correctness. Mullen argues that “the importance of exploring disagreements as the outlier might be correct.” (Mullen, 2003, p. 48).

Doyle comments, “the role of the researcher is to maintain the integrity of the technique by accurately reflecting responses. Achieving consensus of expert opinion within Delphi is a matter of tabulating quantitative data [but] instead, a working toward consensus by the sharing of thoughtful and detailed opinions of experts, who then reconsider them in the following round by comparison to the comments made by other experts.” (Doyle, 1993, p. 142). When trends appear, the moderator should “conceptualize these into a format that makes sense of the accumulated responses” and “sense-making is the trickiest part of the process.” (Doyle, 1993, #2245).

2.2.14 **Knowing when to stop**

Schmidt states that “researchers must know when to stop gathering input from Delphi rounds with the caveat that stopping too soon may reduce the potential consensus and accuracy. Pooling beyond a reasonable point gains no valuable increase in consensus or accuracy and wastes the time of the researcher and panel members (Schmidt, 1997).

Erffmeyer conducted a number of controlled, experimental Delphi panels of MBA students to determine an optimal number of rounds. His results confirmed Martion that “four rounds are usually sufficient.” (Erffmeyer, 1986, p. 125). Other research questioning Delphi used fewer rounds. Erffmeyer contends that research using fewer rounds fails to follow Delphi fundamentals and therefore the findings that question Delphi as effective in decision-making is suspect. He writes,

> Before researchers tend to write off Delphi as less effective than other problem-solving techniques, it seems reasonable to ensure it is being implemented effectively, that is, according to the guidelines of those who devised it. (Erffmeyer, 1986, p. 126)

He counsels researchers that “a tendency to limit the number of rounds to as few as possible” is unwise (Erffmeyer, 1986, p. 126).

2.2.15 **Delphi and Policy Delphi**

Regardless of the protocols, the underlying assumption of a Delphi remains the consolidation of expert opinion on technical topics. However, Delphi appeals to a broader range
of foresighting and non-foresighting activities in which competitive views cry out for consolidation and consensus on topics not inherently technical. The classic response within organizations is to form a committee and Delphi is “not a replacement for the committee process.” (Turoff, 2002, p. 81). Unfortunately, the negative canonical group dynamics and group decision-making accrue to the committee process like domineering personalities, power positions, inability to contradict with fear of reprisal, not being flexible on a position, and the fear of losing face when an idea is farfetched. Organizations grow to the point where they become top heavy with “power so diffuse and no one feels he has the authority or jurisdiction to act as an advocate on the broader issues that arise on the policy level.” (Turoff, 2002, p. 83). The Policy Delphi as discussed next should not “in any way be a substitute for studies, analyses, or staff work” but rather it is an “organized method for correlating views and information processing to a specific policy area and for allowing the respondents representing such views and the information opportunity to react to and assess differing viewpoints.” (Turoff, 2002, p. 83).

The Policy Delphi originates from the need “to generate the strongest possible opposing views on the potential resolutions of a major policy issue” while recognizing that “participants in a Policy Delphi are advocates of competing points of view.” (Brockhoff, 2002, p. 509). Turoff, on his home page, provides a further explanation for the modification of Delphi to create a Policy Delphi,

The Policy Delphi recognized that on many policy issues people could not agree because they represented certain unique interests, values, and/or organizational commitments. The Policy Delphi was designed to allow people to express the strongest possible disagreements so that the sponsor or final decision maker(s) could see what the best information was for any of the alternative resolutions to the policy issue. It was the first Delphi designed to promote disagreement and encourage it rather than fostering a consensus. (Anonymous, 2008, para. 3).

Buck reviewed the foundations for the use of Policy Delphi finding,

Scholars of public policy have long theorized on the sources of general failure in the implementation of legislative directives. Edwards (1984), Hogwood and Peters (1985), and Jenkins (1978) attribute failure in policy implementation to lack of consensus, coordination, and vertical communication of intentions. Levin and Foreman (1986) believe that the heterogeneous nature of American society and its fragmented political system are a source of difficulty in policy implementation, and show that in the field of youth development, executives have improved implementation by building consensus.
While these studies provide use theoretical structures for the analysis of policy failure, they do not attempt to quantify the degree of conflict among players regarding specific policy objectives. (Buck, Gross, Hakim, & Weinblatt, 1993, p. 273)

The value of Delphi appears to be the “depersonalization” of communication that allows a focus on issues. The process avoids being stymied “by the dynamics of a committee or the influence of powerful or vocal special interests in the political process” with the result being a “fruitful debate and discussion.” (Buck, et al., 1993, p. 274).

Unlike the severe criticism of classic Delphi, the Policy Delphi – or its closely related technique the decision Delphi – escapes the most pointed attack, as the goal of the intervention is more likely consensus than quantitative outcomes. Woudenberg writes,

Quantitative and qualitative statements cannot be strictly divided. Ordering priorities or signal future developments implies making assumptions about which social, cultural, and political developments will occur and when they will occur. In any judgment that has accuracy as a goal or, at least, an underlying assumption, accuracy cannot be expelled as a necessary criterion for the particular judgment method used. Only the pure decision Delphi, where the primary goals is consensus, could perhaps be excluded from the demand of the criterion of accuracy. For all other applications, qualitative and quantitative alike, Ascher’s remark about forecasts holds: “accuracy is an asset because utilization of forecasts requires, at a minimum, credibility.” (Woudenberg, 1991, p. 146)

Policy Delphi is not a substitute for the final decision-making. Policy Delphi is considered an approach to establishing the parameters and exceptions needed as input to the debate. The Policy Delphi differs significantly, as, “Generating a consensus is not the prime objective, and the structure of the communication process as well as the choice of the respondent group may be such as to make consensus on a particular resolution very unlikely. In fact, in some cases the sponsor may even request a design which inhibits consensus formulation” (Brockhoff, 2002, p. 80).

Policy Delphi compares to committee meetings. The advocacy nature of participants and the “voting” resolution provide the near semblance of Delphi. However, committee processes suffer from many limitations including growing size related to organizational size growth. Other contributing problems include, according to Linstone, “dominating personalities who are outspoken and take over meetings”, “unwillingness to take a position until it is known which way the majority is headed”, “difficulty with contradicting individuals in higher positions”, “unwillingness
to change a position once it is publicly stated”, and “the fear of bringing up an uncertain idea that might turn out to be idiotic and result in a loss of face.” (Brockhoff, 2002, p. 82). Policy Delphi proponents believe that the process mitigates inherent failings of the committee. A determination of efficacy of either approach over the other is likely to fail. Linstone recommends, “Policy Delphi is not a replacement for the committee process [but] can be used to revise the effectiveness of the committee approach.” (Brockhoff, 2002, p. 82). He goes on to make very clear, The Policy Delphi, therefore, is not in any way a substitute for studies, analyses, staff work, or the committee. It is merely an organized method for correlating views and information pertaining to a specific policy area and for allowing the respondents representing such views and information the opportunity to react to and assess differing viewpoints. Because the respondents are anonymous, fears of potential repercussions and embarrassment are removed and no single individual need commit himself publicly to a particular view until after the alternatives have been put on the table. Even in those cases where the Policy Delphi uses only the committee or sponsoring body as the respondent group, it has the advantage of eliminating the principal bottleneck in the committee procedure by providing a clear delineation of specific differing views, thereby providing an opportunity for the committee members to prepare their respective cases adequately. (Brockhoff, 2002, p. 83)

The Policy Delphi mechanism differs from the Delphi. The communication process differences include “formulation of the issues”, “exposing options – what policy options are available”, “determination of initial positions”, “exploring reasons for disagreements”, and “re-evaluating positions.” (Brockhoff, 2002, p. 84). Three other factors impinge: “(1) the monitor team devoting a considerable amount of time to carefully pre-formulating the obvious issues; (2) seeding the list with an initial range of options but allowing for the respondents to add to the lists; (3) asking for positions on an item and underlying assumptions in the first round.” (Brockhoff, 2002, p. 84).

The Delphi traditionally uses a polling technique on generated item lists. Linstone warns, “The Delphi technique is not just another polling scheme, and the practices that are standard in polling should not be transferred to Delphi practice without close scrutiny of their applicability.” (Brockhoff, 2002, p. 85). Policy Delphi mimics Delphi in the construction of items and voting. Linstone reports that many different dimensions are used for item response. Four dimensions commonly apply. Unlike traditional scaling, that allows for “no opinion” or “no response”, the
scales used in Policy Delphi force choices as a motivation for respondents to make decisions. The scales suggested in the research are (1) desirability (effectiveness or benefits), (2) Feasibility or practicality, (3) Importance, priority, or relevance, and (4) Confidence in the validity of the argument or premise (Brockhoff, 2002, pp. 86-87).

Jillson provides additional guidelines summarized here as,

- a minimal size of at least 2 in the design team
- pretesting of questionnaires
- avoiding compound sentences in questionnaires
- summarization of respondents must favor the diversity from the respondent in contrast to the desire to reach group consensus
- item wording changes should be allowed for participants as participants in Policy Delphi are far more sensitive in their voting to wording than are traditional Delphi
- respondents must be able to see previous votes and positions on items for their personal reasons and to assure them that their previous work is being kept and honored (Jillson, p. 121)

The invitation to participation must establish the participant as part of a group if for any other reason than to establish significance of participation. Brockoff goes on to warn, “As can be seen, there are many things to be considered in running a Policy Delphi, or any other Delphi for that matter. The Delphi concept seems so simple that many people have thought it an easy thing to do. Consequently there have probably been more poorly done Delphi than ones that have been well done”, (Brockhoff, 2002, pp. 82-83).

Alberts reports a case study contrasting the use of subject matter experts versus stakeholders for the resolution of a “not in my backyard” environment decision related to wind power development and deployment of turbines. Democratic principles encourage consideration of divergent ideas. The failure to resolve issues may result in opposition. Delphi processes may be used to inform and build consensus. Alberts reports that in the windmill case, “All participants appreciated the opportunity to learn about the issues, but many possessed insufficient prior experience or knowledge to apply the newly acquired information, engage in meaningful
discourse on the issues, or exert any measure of control over the process” (Alberts, 2007, p. 2336). The key learning for those conducting the exercise is that “participants need extensive prior knowledge of an issue to successfully participate in policy development.” (Alberts, 2007, p. 2337).

2.2.15.1 Policy Delphi Mechanics

The panel for Policy Delphi should be representative of the varied views of policy area. Panelists in Policy Delphi seldom want to educate the Delphi management team. Instead, there is an expectation “that the monitors of the exercise understand the subject well enough to recognize the implications of their abbreviated comments” (Turoff, 2002, p. 85).

A typical Policy Delphi may require up to five rounds but the appropriate seeding of discussion issues may reduce this to three. The option must exist for input from the panel to create their own items and to comment on those proposed. Items added by panel members tend to garner fewer comments primarily because they may be introduced later in the process. Policy Delphi tends to increase the monitor team effort to pre-formulate issues and initially seeding the items. Turoff identifies six phases of a Policy Delphi,

1. Formulation of the issues. What is the issue that really should be under consideration? How should it be stated?
2. Exposing the options. Given the issue, what are the policy options available?
3. Determining initial positions on the issues. Which are the ones everyone already agrees upon and which are the unimportant ones to be discarded? Which are the ones exhibiting disagreement among the respondents?
4. Exploring and obtaining the reasons for disagreements. What underlying assumptions, views, or facts are being used by the individuals to support their respective positions?
5. Evaluating the underlying reasons. How does the group view the separate arguments used to defend various positions and how do they compare to one another on a relative basis?
6. Reevaluating the options. Reevaluation is based upon the views of the underlying “evidence” and the assessment of its relevance to each position taken. (Turoff, 2002, p. 84)

2.2.15.2 Example Uses of Policy Delphi

The Japanese Government repeatedly uses Policy Delphi for Science and Technology investment. Japan was not interested in a prognosis, rather Policy Delphi is an “instrument to
systematically look into the longer-term future with the aim of identifying the areas of strategic research likely to yield the greatest economic and social benefits." (Cuhls, 2001, p. 555). One Delphi indicated the emergence of the fax machine while those attempting to predict earthquakes failed miserably. Consensus activities do not dictate a particular truth but these activities have the consequence of having companies and others "thinking about the future and to a certain extent they are integrated into the shaping of it." (Cuhls, 2001, p. 556). Japanese officials believe that the thinking orients one toward action rather than reaction as events unfold. Consensus does not preclude competition but "creates a network by or during the foresight activities to cooperate in pre-competitive stages or to build up strategic alliances." (Cuhls, 2001, p. 556).

A Kentucky use of Policy Delphi explored the contentious issue of the economic value of tobacco growing to the detrimental health effects. For states with a long history of tobacco growing, any decision-making to curtail growing becomes extremely contentious. A Policy Delphi found that despite "Kentucky's long tradition of economic and cultural dependence on tobacco, former state legislators were more supportive of tobacco control policies than expected." The Policy Delphi appears "a promising tool for building consensus and planning legislative tobacco control initiatives." (Hahn & Rayens, 1999, p. 1).

The Swiss Policy Delphi assisted in the "in trying to confront very divergent opinions in order to identify the most acceptable solutions" to AIDS treatment (Zuber, Mann, Paccaud, Reich, & Turoff, 1996). UNESCO's Policy Delphi focused on life-long learning. The agenda for its use was a recognition that UNESCO is worldwide with difficulty for local experts to participate in the larger audience. Cookson writes, "Lacking among such exchanges have been opportunities for adult educators at the grass roots level to interact with their counterparts overseas." (Cookson, 1986, p. 3).

The Austrian experience with Policy Delphi modified the procedure to be a Decision Delphi. Tichy writes that a Policy Delphi calls for experts to "advertise their respective standpoints and use the Delphi technique to clarify their standpoints [but it] does not lead to consensus but to a clear-cut definition of the different standpoints." (Tichy, 2001, p. 758). In contrast, the Decision Delphi emphasizes the inclusion of those who "are in the actual position in
the decision making hierarchy” (Tichy, 2001, p. 759) as opposed to experts. The interest of the Austrian government was not in forecasting the future alternatives but in “mapping out those fields and niches, in which Austria could reach a leading position.” (Tichy, 2001, p. 758). The fundamental aim of a Decision Delphi is that “reality is not predicted or described; it is made” (Tichy, 2001, #2138).

Rockwell used Delphi to create a focus on research and evaluation needs for distance education. The experiment did not contribute to further understanding of Delphi methodology. The resulting recommendations of the group were of value: (1) cooperation and collaboration, (2) designing the educational experience, (3) teacher training, and (4) educational outcomes (Rockwell, Furgason, & Marx, 2000, p. 1). The policy study aims to guide research.

2.3 Delphi Software Review

Anonymity is a foundational assumption of the Delphi. If anonymity were not foundational then Delphi would be a structured, face-to-face (FTF) committee-like meeting. A corollary to anonymity is asynchronous communication – the participants work on their responses on their time without regard to the choices of times of other participants. Delphi’s ranking, input, and feedback cycles require considerable management and paper communication – usually by mail. Computer communication – and in fact a computer managed Delphi – is the obvious extension of a procedure of anonymous, asynchronous, and intense paperwork communication. Price reports,

In a 1968 article computer conferencing was proposed as a means of speeding up the rate of interaction and the data-processing operations involved in Delphi studies. (Brockhoff, 2002, p. 497)

Given this early start, Delphi was an overlay on text messaging systems. Databases were just being invented, the personal computer was yet to come, and being “online” was rare, expensive, and slow. Twenty years would pass before the worldwide web appears.

In 1974, Johansen asked, “How will a given medium of communication affect the way in which groups of people communicate? What are the most promising near future directions for research considering this question?” (Brockhoff, 2002, p. 508). Johansen assumed computer conferencing was just another medium to research like the telephone or face-to-face communication. He believed that our familiarity with face-to-face communication made it the
standard against which to compare. For him “computerized conferencing and other telecommunications media are not necessarily surrogates for face-to-face communications.” (Brockhoff, 2002, p. 508). The article in which Johansen’s question appears attempted to structure how research might proceed but in retrospect the structuring fell short.

Sheridan, in Technology for Group Dialogue and Social Choice, writes “usually the best way to discuss and resolve the choices that arise within groups of people is face-to-face and personally.” (Brockhoff, 2002, p. 525). Within a prescribed radius of a social network, people build a “circle of intimacy” that he believes “is dictated by human behavior and is not greatly affected by urban population growth, by speed of transportation and communication, by affluence, or by any other technologically induced change in the human condition.” (Brockhoff, 2002, p. 525). This circle contains about 700 people. However, the change in technology has made humankind more interdependent necessitating communication “beyond the circle of intimacy both more awkward and more urgent” thus compelling us to “communicate with and understand the whole cross-section of other citizens.” (Brockhoff, 2002, p. 525). With dramatic increases in interdependence, size, and problem complexity, Sheridan argues that citizen feedback needs to quicker and more inclusive of wider, divergent ideas.

The Delphi method lends itself to the gathering and consolidation of citizen input.

Brockhoff attempted a direct comparison of performance of forecasting groups using computer dialogue and face-to-face discussion with particular emphasis on the Delphi methodology. He writes,

The conditions of group performance have been investigated in thousands of studies. Only few studies have been devoted to the question whether the peculiar organizational structure of the Delphi group leads to higher group performance than the face-to-face discussion in a group. Beyond this the unwieldy, nonhomogeneous, and inaccurate definition of types of tasks by which the performance of groups is judged and thus the classification of concrete formulations of the question make it very difficult to derive statements as to the particular capacity of groups in forecasting.

A large proportion of the statements as to the superiority of certain forms of group organization compared with others was obtained by observing group performance in solving certain kinds of problems and by assuming that the results would apply to tasks which appeared comparable. Thus references to the ability of groups to forecast particular future events were judged on the basis of their performance in responding to almanac-type questions. (Brockhoff, 2002, p. 286)
The results of the Brockhoff study in which he cautions the reader not to overly generalize due to group sizes in the experiment and what he believes to be an invalid assessment methodology based on short-term prediction against almanac data. Nevertheless, he found that increased group size did not increase accuracy; face-to-face groups must be measured in size by active participants not just participants; though as many as five rounds of Delphi are used the results are known by the third round; self-rating of expertise was an unclear measure of accuracy in both Delphi and face-to-face; increased group communication in a Delphi tended to increase forecast accuracy; and in a group that had similar expertise content credentials, there appeared to be less of a chance to improve an estimate via sharing as compared to more inter-disciplinary groups.

2.3.1 Panel Selection

The literature review makes clear that Delphi is a qualitative technique in which some quantitative but non-experimental data are collected. The literature encourages experimenters to provide better controls without sacrificing the nature of qualitative research. Panel selection – size, composition, definition of expert, and sample representativeness – remains a weak point in most Delphi research. Sackman’s critique based itself on sampling within the experimental paradigm and found Delphi wanting. Meyer and Booker argue that scrutiny of panel selection increases when and if other part of the research investigation are questioned. In a Sackman-like critique, they find concentration on panel representativeness immediately challenged when appearances indicate themes of less diversity as in selecting a panel within a specific organization. They write,

The selection scheme is likely to receive close scrutiny if other aspects of the project, such as the results, are questioned. The most frequent criticism is that the scheme failed to select experts who were representative of the larger population and that their answers therefore were skewed. It is commonly believed that skewed results arise from selecting the majority of experts from one place such as the same organization or class of organizations or one point of view. Our studies have not found that the expert’s affiliation or education to be a significant factor in explaining similarities or differences. However, in the interest of trying to represent different views and to avoid criticism, we recommend selecting a balanced group. (M. A. Meyer, Booker, & Bradshaw, 2001, p. 88)

Ultimately, panel selection is a purposeful process in qualitative research and as Meyer and Bookers’s final comment above indicates the criticism might be reduced by the appearance
of balance – whether achieved in reality or not. But they also counter that their own research fails to establish that affiliation or education are the significant factors in explaining differences in views of experts.

Though sampling frames and random sampling within a frame are the hallmark of experimental research, there exist many investigations challenged by strict experimental sampling requirements. A number of alternatives appear in the literature. One specific sampling regime for social networks links participant's uses of members to locate other potential members of an interest. Experts and interested parties form such a social network – though often informal – around areas of concern.

The Delphi literature reports the use of social network nomination. Wedemeyer proposes a "swarming" procedure prior to the first round of Delphi during which panelist nominate others for inclusion. Wedemeyer’s proposal follows the logic of “snowball sampling.” Both will be discussed herein and eventually adapted as the panel selection method in this investigation.

2.3.1.1 Snowball Sampling and Swarming

Though random sampling of a population is the gold standard in experimental research, convenience sampling – grabbing a sample that is available – appears in the social science literature. Those who assume that the sampling frame – a fundamental prerequisite to drawing a sample – exists in all cases fail to understand that not all research interest are capable of identifying the target frame. These purists fail to understand that, for example, many populations will not self-identify – they are hidden. Hidden populations by definition have no boundary and thus a sampling frame cannot be constructed. As an example of hidden populations, consider those who would research AIDS/HIV or active drug use. Researchers are unlikely to find respondents identifying themselves as such in household surveys or other commonly used sampling frames.

Snowball sampling “is a well-known, non-probability method of survey sample selection that is commonly used to locate hidden populations.” (Johnson, 2005, in abstract). Snowball sampling relies on a “few potential respondents [who] are contacted and asked whether they
know of anybody with the characteristics that you are looking for in your research.” (Department of Sustainability, 2008, p. 2) The process is characterized as,

Snowball sampling is especially useful when you are trying to reach populations that are inaccessible or hard to find. For instance, if you are studying the homeless, you are not likely to be able to find good lists of homeless people within a specific geographical area. However, if you go to that area and identify one or two, you may find that they know very well who the other homeless people in their vicinity are and how you can find them. (W. M. K. Trochim, 2008, in snowball sampling)

Each of the recommended additional participants is then contacted to ask them to participate and to recommend others to participate.

Snowball sampling as a form of chain-referral is effective in finding hidden populations. The method does leave researchers with a fundamental problem that chain-referral methods increase the “difficulty of making statistical inferences” and these methods “produce samples that are not even close to simple random samples.” (Salganik & Heckathorn, 2004, p. 197).

To be clear, experts on a specific subject are not intentionally “hidden populations.” Experts do not hide from their expertise due to social stigma. However, they are a social networks not obvious to a single researcher. The advantage for the research stems from using experts who identify other experts due to common interest. The same social network is extensible to interested parties. Thus, the techniques of “snowball” nomination remain applicable. Additionally, as pointed out by Meyer and Booker, the use of the technique reduces the absolute involvement of the researcher in the panel construction.

2.3.1.2 Respondent Driven Sampling

Heckathorn (Heckathorn, 1997, pp. 175-177) outlines four issues with snowball sampling: the criticality of the super-seeds who are chosen first to initiate the network; the participants are more cooperative subjects; participants will mask others for privacy concerns; and, due to the size of each person’s network of relatable members the person’s with the largest network will have undue influence. He proposes respondent-driven sampling (RDS) as an alternative whose primary, claimed advantage is that “RDS produces samples that are independent of the initial subjects from which sampling begins.” (Heckathorn, 1997, p. 177). The RDS approach differs from snowball sampling in its dual reward system – participants are rewarded for participating but
are also rewarded for recruiting others. Also, RDS, unlike snowball, does not require subjects to identify their peers to the investigator but to recruit them directly into the study (Heckathorn, 1997, p. 177).

Salganik and Heckathorn claim that eventually the RDS procedure results in sample independent of the seeding, they write "seeds should be well motivated and enthusiastic, and hence willing to try to recruit their peers; they should also be sociometric stars, individuals whose high regard among their peers enables them to recruit their peers, while also instilling in them motivation to continue the peer recruitment process." (Salganik & Heckathorn, 2004, p. 109). Initial seeds should be "diverse with respect to the factors that most strongly determine the formation of social ties within the population." (Salganik & Heckathorn, 2004, p. 109) However, contamination might be more suspect due to the social network power of the seeds described, "Such individuals can more easily promote participation, and accelerate recruitment. Paradoxically, careful seed selection speeds the growth of recruitment chains and thereby accelerates the point at which seed selection becomes irrelevant and thus helps reduce bias." (Wejnert & Heckathorn, 2008, p. 109).

Salganick and Heckathorn provide mechanics (1) insure that panelist inclusion is not duplicated, (2) verification that the nominated panelist is part of the population sought, (3) the degree of a participant – the number of other participants that a given panelist "knows" or has "social contact" with in the network, (4) track who is recommended by whom for inclusion, (5) make a judicious choice of how many additional panel members for inclusion any given included member can recommend, and (6) follow the recommendations on the initial seeds. If the sampling procedures are followed then calculations are presented for an "unbiased estimate of the proportion with a specific trait based only on data collected during respondent-driven sampling." (Salganik & Heckathorn, 2004, p. 219). They conclude that chain constructed samples are convenience samples that many believe are "so hopelessly full of bias as to be useful only for exploratory purposes." (Salganik & Heckathorn, 2004, p. 229). Countering this, they argue that correctly done chain referral "can produce estimates that are asymptotically unbiased." (Salganik & Heckathorn, 2004, p. 229). They believe that the estimation is robust
even in the face of initial seed selection that is "traditionally considered to be one of the largest problems with chain-referral methods." (Salganik & Heckathorn, 2004, p. 229).

2.3.1.3 Delphi Swarm

Wedemeyer argues, “we have entered policy-making and planning environments of increasing complexities, where rapid innovation will continue to heighten policy and planning uncertainty for everyone as technological, economic, political, environmental and social trajectories become more difficult to anticipate.” (Wedemeyer, 2008, para. 2). Change and futures press decision-makers who seek reduction in uncertainty. Wedemeyer proposes the need for a “futures forecasting engine” (FFE) looking much like a web-based Delphi. The engine is characterized as "a systematic and automated process that involves multiple techniques and stages, with multiple outputs." (Wedemeyer, 2008, in FFE section). The engine assists with participant selection, development of proposals and their prioritization, collection of responses, results summary, and reporting.

Of particular interest for this investigation is the "swarm" stage proposal. The process emulates the snowball-type, network nominated membership of a target population. In the case of the Delphi, the target population is the panel of experts. Wedemeyer's Futures Forecasting Engine automates the selection from an initially seeded list by allowing seeds to nominate other experts in the social network via email. Each nominee is given the opportunity to participate, to nominate additional panel members, or refuse to participate via a web interface. The nomination process continues until,

This branching “swarm” activity continues until the desired number of distributed Social, Economic, Political, Technical and Environmental participants are obtained along with each e-mail address which are captured in an overall study database. (Wedemeyer, 2008, in FFE examples section)

Some similarities exist between the swarm described by Wedemeyer, snowball sampling, chain-respondent sampling, and Heckathorn’s respondent-driven sampling. However, there are major differences.

As described, Wedemeyer uses the general concept of respondent-driven sampling but does not provide theoretic foundations for the methodology. In part, this is understandable from
the perspective of the qualitative epistemological foundations of Delphi where panel selection – regardless of Sackman’s and others attacks – makes no claim of statistical representativeness of a population. The work of Heckathorn (Heckathorn, 1997) is more concerned with statistical implications of hidden populations, sample frames, and the ability to use response-driven sampling in classic statistical reporting. Heckathorn wants to establish a solid approximation of the characteristics required for sample selection in experimental research so that parametric statistical tools may be validly used in analysis.

There is value in discussing the differences. Improvements might be made to the “swarm” technique without attempting a metamorphosis into the realm of experimental sampling in which Heckathorn RDS operates.

<table>
<thead>
<tr>
<th>Table 2 Respondent Data Sampling versus Swarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heckathorn respondent-directed sampling</td>
</tr>
<tr>
<td>Assumed relationship is network and uses</td>
</tr>
<tr>
<td>directed graph</td>
</tr>
<tr>
<td>Prevent duplicate inclusion especially with</td>
</tr>
<tr>
<td>populations where duplicity may be normal</td>
</tr>
<tr>
<td>Verification of population inclusion</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Measure of degree of participant</td>
</tr>
<tr>
<td>Judicious choice of how many additional</td>
</tr>
<tr>
<td>members an included panelist may recommend</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Description of characteristics of initial</td>
</tr>
<tr>
<td>seeds</td>
</tr>
<tr>
<td>Reward structure for participating and</td>
</tr>
<tr>
<td>nominating a panelist who additionally</td>
</tr>
<tr>
<td>participates</td>
</tr>
</tbody>
</table>

As documented above, there is considerable difference. The fundamental difference might originate with the difference between a hidden population that is hidden for unknown reasons as opposed to experts whose social network is being used to include others. There is a design question fundamental to the FFE model in using a tree versus a directed graph that reflects the difference in perception of the process as using a social network or a disjoint set of non-overlapping experts. As previously indicated, the differences are also due to the assumed
qualitative selection of experts versus experimental sampling. A number of the differences are simple enough to match and would add to the explanatory power of the “swarm”.

2.3.2 Delphi Conference versus Delphi Exercise

The Delphi exercise is primarily a “paper and pencil” exercise during which panel participants record their thoughts and precede through the controlled feedback processes. The content of responses is unknown ahead of time. “To a degree, this form of Delphi is a combination of a polling procedure and a conference procedure which attempts to shift a significant portion of the effort needed for individuals to communicate from the larger respondent group to the smaller monitor team.” (Brockhoff, 2002, p. 5).

The Delphi conference replaces the monitor team and face-to-face communication with computer based procedures that aid in the discussion and summarization. There is a significant reduction in time between rounds making the process near real time. Physical location and size of group no longer constrain the process. “However, it does require that the characteristics of communication be well defined before Delphi is undertaken, whereas in a paper-and-pencil Delphi exercise the monitor team can adjust these characteristics as a function of the group responses.” (Brockhoff, 2002, p. 5).

The HERO Project studies global environmental change as a human by product. The founders believe that “people experience and respond to global environmental changes in localities. Consequently, according to this site, there has been a proliferation of research centers and sites dedicated to studying the local implications of the human dimensions of global environmental change.” Because the participants are geographically dispersed, the HERO Project created an infrastructure including a Web based Delphi. Their analysis of a Delphi conference had to address the procedure as well as the role of the moderating team. The experimenters provide this discussion on their web page of the role of the moderator,

Nearly all Delphi activities require, to a greater or lesser extent, a human facilitator to manage the feedback, initiate activities, and focus the panel. In traditional pen-and-paper activities, the moderator plays a central role; all

---

4 Human-environment regional observatory, http://hero.geog.psu.edu/index.jsp
responses pass through the moderator's hands before being presented to the group. The moderator may edit responses, form new questions based on those responses, summarize contributions, and decide on the order and style of each round. Electronic Delphi activities can reduce the role of the moderator, but generally do not eliminate it. Panelists' contributions can be posted directly for the group to see (as in a message board), and panelists may be given authority to call for a vote on certain topics if they feel that common themes are beginning to coalesce. Nonetheless, depending on the topic of the Delphi activity or the style of contributions, it may be desirable for a facilitator to screen responses, removing identity revealing information or discarding contributions that do not address the problem. A moderator will also be able to detect when discussion has reached an impasse, and what an appropriate response might be. While a successful Delphi panel may be able to remain focused and develop appropriate solutions without intervention, having a human moderator able to step in and refocus the group will likely be a component of each e-Delphi exercise. (Anonymous, 2002)

2.3.3 Computer Supported/Mediated Delphi

Delphi exercises are usually paper-based exercises but the ubiquitous Internet and World-Wide Web motivate an ITC-based Delphi. Given that the vast majority of research uses paper-based Delphi exercises, this investigation is compelled to review the literature. To put it simply, does the use of ICT – and specifically a web-based Delphi -- change the functioning of the Delphi? This question in and of itself could be the source of rich research streams. Considering Clark’s (R. Clark, 1994) admonition, the question might also ensnare in the debate as to whether any medium is separable from the content of the message. Regardless, some consideration for variations in results that might originate from the web-based approach appears. Some might argue that such consideration is without merit.

However, consider one simple change from paper-based to web-based. A paper-based investigation puts all participants on the equal footing of using handwritten responses assuming no special handicaps. A web-based collection procedure requires typing skill. Does typing skill therefore moderate the number of responses and the length or detail of responses simply due to the workload imposed on a participant who is a non-touch typist? If there is this possibility, then should a strictly paper-based system be used or one that mixes both media? The mixture would itself generate more questions. The questioning extends further. How does the structure of the web interface itself change the process? How will feedback be structured? Will the decision processes --both individual and group – like the belief that reasoned feedback moderates
decision outcomes be altered by the medium? Are the group decision-making processes in a computer-mediated decision-making activity equivalent to those of non-computer-mediated?

There is no simple way to answer the questions above or all the other questions that might also arise from a careful consideration of the change in medium from pencil-and-paper to web-based data collection. At best, a survey of related research to seek out any highly efficacious mediators for consideration might inform design decisions in this dissertation. From the general issues, research specifically aimed at a web-based Delphi might contribute insight.

Turoff et al. believe that a Delphi exercise with hundreds of participants is possible with computer-mediated communication. They believe most systems to date do not exhibit “content oriented discourse and visualization structures.” (Turoff, Hiltz, Bieber, Fjermestad, & Rana, 1999, p. 7). They believe that such tools will expand the problem domain to “support problem solving and learning communities of scores to thousands of participants.” (Turoff, et al., 1999, p. 7). Such systems can evolve “the discourse structures, visualizations, voting procedures, and human roles appropriate to its nature and application.” (Turoff, et al., 1999, p. 8). These changes will expand the possibility for Delphi exercises to a very wide community as opposed to the current limits imposed upon the exercise due to its paper and pencil foundations.

2.3.3.1 GDSS Literature Generally

In GDSS literature, Gallupe and DeSanctis are frequently cited. They found that computer supported decision systems are “particularly helpful in groups receiving the task of higher difficulty” and “decision confidence and satisfaction were lower in computer supported groups than in non-supported groups.” (Gallupe, Desanctis, & Dickson, 1988, p. 291). They found that group solution time is “significantly greater than the average individual solution time for all difficulty levels.” (Gallupe, et al., 1988, p. 292). The use of GDSS increases the number of alternatives considered which they believe contributes to better decision-making. Task difficulty moderated results. They found the “use of the technology improved decision quality for both higher and lower tasks difficulty conditions, the effect was particularly strong in groups expose to the more difficult task.” (Gallupe, et al., 1988, p. 293). Research is needed to determine how task complexity and group decision processes interact. Surprisingly, GDSS decision-making results in
“in more negative sentiments from group members regarding decision confidence, agreement with the final solutions, satisfaction with the decision process, and inter-member conflict.” (Gallupe, et al., 1988, p. 293).

Walther’s investigation of computer-mediated communication supports sweeping claims of efficacy as,

Results indicate that some conditions of computer-mediated communication in use by geographically dispersed partners render effects systematically superior to those obtained in other mediated conditions and greater or lesser than effects obtained through face-to-face interaction. (Walther, 1997, p. 342)

Baltes et al. completed a meta-analysis comparing decision-making in face-to-face versus computer-mediated. They found “computer-mediated communication leads to decreases in group effectiveness, increases in time required to complete tasks, and decreases in member satisfaction compared to face-to-face groups.” (Baltes, Dickson, Sherman, Bauer, & LaGanke, 2002, p. 159). They also considered group process modifiers like anonymity, time to reach decisions, group size, and task type.

Almost immediately, Baltes et al. identify problems in the research base from which they drew articles for inclusion in the meta-analysis. A wide range of dependent variables – quality of outcomes measured subjectively and objectively, satisfaction with the process, time consumed, quality of interaction, and patterns of influence. Then, they also found difficulty in the variations of computer-mediated interventions including email, chat, video conferencing, formal decision support systems, and so on. Two characteristics shown commonly in the literature – synchronicity and the ability convey “paraverbal and nonverbal aspects of communication.” (Baltes, et al., 2002, p. 173). The successful outcome of a decision-making exercise appears related to highly rich, synchronous communication. The authors point out that most studies used ad-hoc student groups and that email dominated the communication. The dependent variables for Baltes et al. were “group effectiveness, length of time to decision, and satisfaction” moderated by anonymity, discussion time, group size, and task type.

Having setup their investigation, Baltes et al. surprisingly observed, “Despite the wealth of writing and research on the topic of computer-mediated communication, there are relatively few studies that include the basic methodological rigor and statistical detail to be included in a meta-
analysis.” Even more perplexing is the absence of resultant variables related to the effectiveness of the resultant decision and instead a plethora of results “focused solely on member satisfaction” and “satisfaction measures without effectiveness measures to be only minimally useful to organizations and researchers interested in the question of whether organizations should make use of computer-mediated communication systems for group decision making.” [Emphasis added] (Baltes, et al., 2002, p. 173).

In sum, the meta-analysis concludes that little difference exists on outcome measures when “groups are anonymous and have unlimited time to reach decisions” but these assumptions are unrealistic in the common context of organizations where “time pressure is high and the individuals are selected for the decision making-group based on their expertise and thus their identities are quite salient to other members.” (Baltes, et al., 2002, p. 175). Ultimately, the resulting point is – “Is the reliance on computer-mediated communications for organizational decision-making appropriate?” The infatuation needs considerably closer examination in the decision-making context.

Baltes et al’s analysis concludes,

Computer-mediated communication may be an efficient and rapid means of disseminating information, but the research to date suggests that it is not the most effective means of making group decisions. While researchers must begin to attend to the increasingly common media of electronic mail and videoconferencing, managers must make the decision as to whether the cost savings in travel expenses and time outweigh the potential decrements in quality of decisions reached. Unfortunately, primary research on these other increasingly important communications media is presently in short supply.

All in all, our results suggest that computer-mediated decision-making groups are rarely if ever more effective than face-to-face groups, that CMC group members are rarely if ever more satisfied than members of face-to-face groups, and that CMC groups rarely if ever take less time than face-to-face groups. At best, CMC groups are not significantly worse than face-to-face groups, and even these results occur only in very unusual and uncommon organizational conditions. Managers, researchers, and theorists are thus cautioned against blindly joining in the current trend toward embracing all things Internet. (Baltes, et al., 2002, p. 175)

Bordia completed a narrative literature consolidation comparing face-to-face communication with computer-mediated communication [CMC]. Bordia notes, “There are no nonverbal cues to embellish meaning or social context cues regarding gender, age, or status. Not only can the absence of cues hamper communication efficiency, but also it seems to create a
semblance of anonymity and lack of awareness of the social context. These conditions, in turn, have been held responsible for a perceived higher incidence of rude, offensive, and uninhibited behavior (see Kiesler, Siegel, & McGuire, 1984).” (Bordia, 1997, p. 175). Bordia presents a number of propositions from the literature review. CMC groups take “longer to complete the allotted task” and “in a given time period produce fewer remarks” about ideas advanced by others yet “CMC groups perform better that FTF groups on idea generation tasks.” (Bordia, 1997, p. 100). The last point moderates by the procedures in place, as FTF groups using structured communication techniques like nominal grouping appear to be equal in idea generation. CMC groups provide “greater equality of participation.” (Bordia, 1997, p. 101). CMC groups perform worse that FTF groups when time is limited and social emotional interaction could be used to clarify. If time is not limited, CMC and FTF do about equal. CMC groups reduce normative social pressure. CMC understanding and perception of other members is weaker. Again, time is a factor as understanding is reduced in the time constraint of CMC. CMC tends to higher incidents of uninhibited behavior due to a “state of de-individualization.” (Bordia, 1997, p. 108). Finally, consistent with other research, “CMC groups, as compared to FTF groups, exhibit less choice shift of attitude change” (Bordia, 1997, p. 109).

Benbunan-Fich et al. investigated the decision-making behavior of face-to-face versus asynchronous groups noting that “asynchronous system leads to different communication behavior and to unique coordination problems and opportunities.” (Benbunan-Fich, Hiltz, & Turoff, 2003, p. 458). They conclude, “the breadth of asynchronous discussions is higher than the number of ideas mentioned face-to-face.” (Benbunan-Fich, et al., 2003, p. 465). Asynchronous groups “adopted more loosely couple interaction approaches, where they could work in parallel” (Benbunan-Fich, et al., 2003, p. 465). When attempting to complete their task, the wider discussion of issues in asynchronous communication “had a positive impact on completeness of a final report.” (Benbunan-Fich, et al., 2003, p. 465). Neither face-to-face nor asynchronous communication demonstrated a significant drop off key ideas to the final project.
2.3.3.2 Generation of Alternatives

Zigurs and Kozar (Zigurs & Kozar, 1994) contrasted manual group decision-making with computer-supported systems. They found many more communication acts initiated in computer-mediated decision-making. However, they found that considerably more initiated communication acts in the computer-supported environment focused on “process.” These acts sought “requests or suggestions for procedural direction and are concerned with how to use the structure.” (Zigurs & Kozar, 1994, p. 293). Further, computer supported groups “used significantly less goal-oriented and integrative statements than did manual groups” where “a concern with summarizing progress and integrating contributions should have a positive effect on decision quality.” (Zigurs & Kozar, 1994, p. 293). Citing Gallupe’s finding that “GDSS should improve information exchange and, therefore, group outcomes” that in turn “increased effectiveness in GDSS groups”, they found “computer support helped even out influence behavior, but it did not result in higher quality decision then were made by manual groups.” (Zigurs & Kozar, 1994, p. 293). They believe the difference in findings may be the result of differences in task complexity.

Kerr and Murthayl investigated idea generation and an “integrated bargaining task” of teams of face-to-face versus computer-mediated. They found that computer-mediated groups generated “a significantly greater number of unique ideas” but made no judgment about the quality of ideas (Kerr & Murthay, p. 416).

2.3.3.3 Structure’s Influence on Decision-Making

When considering evidence before the Courts, an essential ingredient of establishing validity is the ability of the opposing party to cross-examine and counter presented evidence. In Court, the Judge and judicial procedures control the conflict of the parties. Normal decision-making groups are not required to adopt such externally imposed structure. True, norms of behavior exist but conflict resolution is seldom explicitly agreed upon prior to discussion initiation.

Poole et al. found that structure influenced conflict level and management but that computer-supported systems as compared to non-supported systems had an “effect beyond structure.” Manual groups “dealt with conflict in a low-key fashion that did not develop obdurate oppositions between group members [thus able to] engage in hard bargaining without escalating
the conflict too much, and this led to high consensus change.” (Poole, Holmes, & Desanctis, 1991, p. 949). In GDSS, members were unable to do “hard bargaining” thus using “integrative discussions tempered by avoidance behavior.” (Poole, et al., 1991, p. 949). If GDSS groups attempted bargaining, more members opted out with avoidance responses. GDSS groups tended to “focus on written materials, somewhat more depersonalization, and greater expression of positive affect than in other conditions, which might be expected to “cool off” the conflict.” (Poole, et al., 1991, p. 950).

GDSS group’s negatives included the use of “voting to cut off or end discussions more in GDSS groups than in other conditions” (Poole, et al., 1991, p. 950) whereas manual discussion threats of votes caused more discussion. Manual groups were more willing to explore reasons that members held views different from the majority.

GDSS generated less consideration of alternatives to presented proposals and the researchers posit the belief that entering ideas “froze” them. This freezing effect might also explain why GDSS groups did not “achieve as much consensus change as manual groups.” (Poole, et al., 1991, p. 950).

Poole et al. overall believe that structure in GDSS or manual groups both reduce conflict.

Karacapilidis et al. investigated the role of decision-making criteria on group processes. They found considerable group effort exhausted on establishing the criteria used to evaluate alternatives. The metrics used to compare alternatives are essential to comparing, accepting, or rejecting alternatives proposed in decision-making.

The inclusion of particular criteria may cause one to consider alternatives that would otherwise not figure in the process, while the exclusion of certain criteria may automatically eliminate certain alternatives that would prima facie be included. They can, therefore, be a preliminary battleground for power struggles between factions involved in the process. The decision function, however, is where most argumentation is centered, since it is here that the relative value of choice criteria is established and applied to select between the alternatives. The argumentation used is often authoritative or based on voting. (Karacapilidis, Papadias, & Pappis, 1999, p. 9)

As with the Courts, GDSS or GSS often provide decision rules as part of the system.
2.3.3.4 Consensus and Closure

Consensus reached on decision-making outcomes appears to improve satisfaction with the decision-making process. Lemus et al. studied computer-mediated groups to conclude “the number of members in support of a proposal relative to those in opposition, as well as arguments in support of a proposal, are both significant predictors of decision outcomes.” (Lemus, Seibold, Flanagin, & Metzger, 2004, p. 308). More textual responses in support of a proposal only enhanced its adoption. In their experiments, status and physical cues were absent allowing preponderance to be isolated as a driving factor. Though they do not dismiss content of argument for or against a proposition in the computer-mediated group, “the proportion of members offering endorsement or objections to a proposal” (Lemus, et al., 2004, p. 314) appears to have yet more weight. They found that CMG participants tended to generate more “arguments in support of decision proposals” than “the development of arguments opposing them.” (Lemus, et al., 2004, p. 414). Again, the depth of evidence did not appear to be a major determiner of eventual outcome.

Wilson et al. found that in CMG groups over five, there are “individuals, who appeared to disassociate themselves from the CMG, producing hurried messages that were poorly proof read.” Groups detected these individuals considering them as negative. These members “could reduce the formation of team spirit and camaraderie in a potentially escalating cycle of alienation and disassociation of the individual.” (E. V. Wilson & Zigurs, 2001, p. 74).

Small found that “closure is a significant predictor of decision satisfaction” for a CMC group. Using the Cognitive-Motivational Model, Small found the “Need for closure influences adequacy assessment.” (Small & Venkatesh, 1995, p. 553). When groups were compelled to speed their decision versus given no guidance, “Information deemed extensive in need to expedite conditions may be deemed inadequate in need to delay closure conditions, suggesting that the uncertainty cutoff may be set relatively low in need to expedite conditions and relatively high in need to delay closure conditions.” (Small & Venkatesh, 1995, p. 554). Satisfaction and closure relate. Small did not address effectiveness of the achieved decision.
Alge found that teams that have a “future” are more likely to compromise “in order to preserve team harmony.” (Alge, Wiethoff, & Klein, 2003, p. 33). Teams that developed openness and trust “performed better on the high interdependence negotiation tasks; teams higher in information sharing did not.” (Alge, et al., 2003, p. 33).

2.3.3.5 Feedback

Hiltz et al. described two types of GDSS – those meant to replace FTF in a decision room and “some are constructed to structure or support decision making in a distributed communications environment.” (Starr Roxanne Hiltz, Johnson, & Turoff, 1991, p. 82). In their experiment, participants used constant statistical feedback as to current rank orders of decision items and vote changes. In a sense, the activity was a real-time Delphi exercise. The analysis of the data indicates, statistical feedback encourages “participants to change their decisions in order to conform more closely to the group average, and lessens the influence of the most knowledgeable member.” (Starr Roxanne Hiltz, et al., 1991, p. 100). Statistical feedback did not significantly improve group agreement nor did it increase satisfaction. It appears that statistical feedback “tended to be detrimental to quality of decision rather than improving it.” (Starr Roxanne Hiltz, et al., 1991, p. 100). The implication for design of computer-mediated decision-support systems is the need to “ensure the exploration of disagreements, and quantification or statistical feedback on “average” opinions should not be allowed.” (Starr Roxanne Hiltz, et al., 1991, p. 102). Participant knowledge of the computer system appears to mediate and “groups composed of participants with some previous experience using computers and groups with competitive rather than cooperative social histories.” (Starr Roxanne Hiltz, et al., 1991, p. 102) produce better results.

Information overload negates the value of feedback imposed by computer-mediated communication due to an expectation of “the imposition of unwanted and useless messages.” (S. R. Hiltz & Turoff, 1985, p. 680). The classic dormitory study examined a crowded dormitory in which “the crowded living situation is characterized by frequent and unwanted or unmediated social interactions.” (S. R. Hiltz & Turoff, 1985, p. 681). To mediate the overload, screening skills appear that “reduce the stress of numerous inputs by construction a priority-based pattern of
attention to information to cope with the high rate of social interaction in a deliberated and organized manner.” (S. R. Hiltz & Turoff, 1985, p. 681). Individuals will self-organize communication with “different individuals having different needs and preferences” thus systems must “offer a number of options for information organization and handling instead of imposing a single solution for all users.” (S. R. Hiltz & Turoff, 1985, p. 681). Hiltz and Turoff report Miller who “found that individuals tend to focus on filtering and omitting (ignoring) information as the primary effective ways of coping with high rates of information overload.” (S. R. Hiltz & Turoff, 1985, p. 681). Systems are but a partial answer to overload and system users learn to “cope with information overload after they gain sufficient experience” with the capabilities of the system that help to organize its flow (S. R. Hiltz & Turoff, 1985, p. 682). The effectiveness of the individual deteriorates “when the workload is too great or too small” but “intermediate users seem to suffer the most from information overload.” Interestingly, “individuals in decision-making situations seem to do just as well using summary data as raw data.” (S. R. Hiltz & Turoff, 1985, p. 683). The user assessment of effective systems appears to support “the flexibility it allowed them” for organizing information.

As stated, GDSS should “positively alter information exchange to improve decision outcomes” but Hilmer and Dennis question this assumption because “groups in these studies made no better decisions when using groupware than when discussing the information verbally, despite identifying and discussing the critical information omitted from verbal discussions. The conclusion from these studies is that group members often discount or do not attend to information from other members. Simply put, individual information processing often fails in group settings (Hilmer & Dennis, 2000, p. 2).

Most feedback in GDSS is “contained in text in a pool of information” making it easy to ignore because “no one is placing it in front of the participants.” (Hilmer & Dennis, 2000, p. 3) As factual data requires cognitive processing, the authors argue, “normative influence may dominate information influence” (Hilmer & Dennis, 2000, p. 7) simply due to ease of processing during exchanges. If feedback is important in a GDSS, then the authors believe that active categorization and analysis of incoming information should be compelled by the GDSS. In attempting to compel consideration, they found participants paid attention to new, incoming information but “paid no more attention to already known information.” (Hilmer & Dennis, 2000, p. 3).
7). The experiment required participants to actively construct categories and integrate incoming information. Requiring each participant to categorize each piece of information "makes the participant not only look at items grouped together in chunks, but forces one to consider how each piece of information fits into those chunks." (Hilmer & Dennis, 2000, p. 8). In their experiment, they recognize that simply increasing attention "to information alone did not improve decision quality but increased integration of information did" thus "a whole bunch of words do not lead to good decisions but evaluation of the comments can." (Hilmer & Dennis, 2000, p. 8).

2.3.3.6 Measuring Decisions Made

Poole and Holmes note that GDSS are assumed to "improve group decision-making which in turn should result in more effective decisions." (Poole & Holmes, 1995, p. 91). Normal evaluation of the improvement in effectiveness might be based upon quality, timeliness, and commitment. The belief is that GDSS's "improve decision making from their ability to enhance communications and information exchange, processing of complex information, and the coordination of group activities" (Poole & Holmes, 1995, p. 93).

Poole and Holmes make a key point prior to laying out their investigation. Their concern centers on the use of post meeting perceptions (attitudes) about effectiveness of the decision-making process. Of even more concern is the research that indicates the recall of what happened during decision process is often not the same as what actually happened. They write,

A number of previous studies have focused on members' postmeeting perceptions of meeting processes, and many of these have assessed the degree to which members believe the GDSS helps the group organize its process (e.g., Connolly Jessup, & Vdacich, 1990; Easton, Vogel, & Nunmaker, 1989; Gallupe, Bastianutti, & Cooper, 1991; Lewis, 1987; McCartt & Rohrbaugh, 1989; Steeb & Johnston, 1981; Watson, DeSanctis, & Poole, 1988; Zigurs, Poole, & DeSanctis, 1988). These are not particularly helpful in assessing the degree to which GDSSs affect the structuring of work because they ask for postmeeting recall of group processes. [Emphasis added] A study by DAndrade (1974) indicated that postmeeting recall of interaction on global scales does not accurately reflect the actual rate of occurrence of corresponding behaviors. Instead, such recall seems to reflect members' overall feelings about the interaction and its outcomes and is often shaped by an implicit theory of how the interactions should go. (Poole & Holmes, 1995, p. 94)

For Poole and Holmes, "direct observation of groups using GDSS" allows "reliable assessment of the types of activities and the order in which they were enacted." (Poole &
Holmes, 1995, p. 95). Their experiment was designed to “determine whether, in the rough and tumble of group interaction, an information technology specifically designed to structure group behavior actually does so, and if it does, whether this structuring leads to beneficial outcomes.” (Poole & Holmes, 1995, p. 97).

Their results indicate that groups use a wide array of decision paths even if they were encouraged to use a set approach. They were surprised to find that even with a GDSS the resulting decision paths “did not result in more orderly decision paths than the baseline condition.” (Poole & Holmes, 1995, p. 97). They noticed that GDSS groups repeatedly had to take time to see “where they were in the agenda” and the manual groups with the same agenda “had paths close to the order of the items on the agenda suggesting they had fewer problems” but they did recycle more and had less orderly paths (Poole & Holmes, 1995, p. 121). Decision paths were not determinative of quality outcome. Groups that followed the normative path closest were most satisfied with their decision.

2.3.3.7 Limitations of GDSS Research

Sharda’s meta-analysis notes that DSS have grown very popular but their effectiveness is yet to be demonstrated as research “claiming usefulness of the DSS has relied largely on anecdotal or case data.” (Sharda, Barr, & McDonnell, 1988, in para. 1). Sharda used MBA students participating in an executive decision-making game. They found “the groups with access to the DSS made significantly more effective decisions in the business simulation game than their non-DSS counterparts. The DSS groups took more time to make their decisions than the non-DSS groups at the beginning of the experiment. However, the decision times converged after the third week.” (Sharda, et al., 1988, in para. 2). Though their research appears experimental, the same limitations of decision-making in a game environment and using students apply.

Whetherall and Hailstones conduct seminars on risk assessment during which participants use statistical techniques. They conclude that “developments in statistical theory in the twentieth century have not resolved this limitation” and that “however rigorous the statistical theory attempts to be, personal judgment will be required in real life business situations.”
(Weatherall & Hailstones, 2002, in Sect 4.2). They cite Hacking who claims that people on one side use stochastic laws but “On the other side it is epistemological, dedicated to assessing reasonable degree of belief in propositions quite devoid of statistical background.” (Weatherall & Hailstones, 2002, in Sect 4.2). Bernstein claims that decisions often ride on “animal spirits” and not “as an outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities.” (Weatherall & Hailstones, 2002, in Sect 4.3). They conclude,

Asch [1] has shown how people’s judgments are influenced by their peers, even to the extent of people giving clearly incorrect answers to simple and definitive questions, simply because other people gave wrong answers (in these experiments the ‘other people’ had been primed to give wrong answers). Hertz [6] analyzed business executives’ attitudes to risk, and illustrated that peoples’ unwillingness to take risk in a business context might be counter-productive for the business, in that people were unwilling to take risks that would be advantageous to the business. Weatherall [12] concluded, “No decisive theory exists which will satisfactorily indicate, on an analytical basis, the optimum decision in a situation of risk.” (Weatherall & Hailstones, 2002, in Sect 4.3)

Most GDSS have embedded statistical forecasting models and intense data analytic capabilities. If Whetherall’s observations are correct, then the role of these systems is that of support for professional judgment and not an end in them.

2.3.3.8 Social Constructionism and Group Norms

Social constructs appear in computer-mediated communication in spite of the asynchronous, non-face-to-face interaction style. Postmes argues, “mediated groups can develop a meaningful and strong sense of identity through interaction despite the fact that many of the factors traditionally associated with social and interpersonal attraction and with normative influence are absent in such contexts.” (T. Postmes, Spears, & Lea, 2000, p. 344). More so, “groups can be very real to their members psychologically despite lack of physical contact” and the “social influence and attraction can actually be stronger in anonymous groups because anonymity may accentuate the interchangeability of group members.” (T. Postmes, et al., 2000, p. 344). He continues, “social influence does not require physical presence or visible social cues but stems from psychological power of the group qua common identity.” (T. Postmes, et al., 2000, p. 345). This finding challenges the belief that social norms establish only with physical presence and are usually determined prior to the computer-mediated interaction.
The identity that develops is believed to “define how group members should think, feel, or behave.” (Tom Postmes & Lea, 2000). Norms of a group are inferred from "prototypical properties of the group" and these “inform a group member what behaviors are typical and, hence, appropriate, desirable, or expected.” (Tom Postmes & Lea, 2000). Individuals assign themselves the attributes of the constructed social identity.

What role does social constructionism play in CMC and even more so in Delphi? Anonymity is a hallmark of Delphi. Like Delphi, CMC “renders users relatively anonymous” reducing the “individuating information usually abundant (through physical appearance, cues, and interactive behaviors).” (Walther, 1997, p. 346). Participants in CMC, judge “judge one another on the basis of group similarity or difference”. Consequently whatever subtle social context cues do appear in CMC take on particularly great value. CMC users build stereotypical impressions of their partners based on language, typographic, and contextual cues. They engage in an over attribution process without tempering their impressions in light of the relatively meager information base on which they are built (Lea & Spears, 1992). This tendency to project stereotypical attributes on others occurs precisely because of the lack of individuating information communicated by the medium and is promoted by the deindividuating conditions. (Walther, 1997, p. 346)

Therefore, even in anonymity there are sufficient social cues to develop a network and establish norms of behavior that will guide individual action. Guiding individual action from socially constructed norms against which anonymity was supposedly guarding presents direct concerns for Delphi. Delphi exercises are short-lived but they are exercises for which there is “an expectation of future interaction.” Walther finds,

Research on mediated interpersonal interaction suggests that the anticipation of future interaction prompts communicators to seek more information about one another, to act more friendly to cooperate in negotiations, to more highly value self-disclosure-in essence, to enact more relationally positive communication (Walther, 1997, p. 349)

Walther goes on to assert “CMC is an amplifier or magnifier of social psychological and communication phenomena, and the question becomes how to use CMC or rather how to design the social circumstances of its use” and “the interaction of long term/short term existence and group versus individual identity factors have profound implications, leading to significantly different communication, interpersonal impressions, and outcomes” (Walther, 1997, p. 360).
Lee sought understanding of depersonalization’s relationship to conformity in anonymous groups. She found “depersonalization diluted within-group differences and fostered perceived homogeneity” which is consistent with theory. Depersonalization heightened attention to messages and in particular to the norm. She argues “depersonalization amplified group influence not only by elevating in-group feelings but also by modifying cognitive representation of the group norm” (E.-J. Lee, 2006, p. 425).

Ho and McLeod believe that face-to-face discussion of controversial content reduced participant comment when compared to online discussion. They believe that communication dysfunctions of self-suppression of comments is ameliorated in web-based chat thus allowing discussion of controversial issues (Ho, 2008).

Reward behavior in computer-mediated group activities shows that “individual-based incentives and computer-supported distributed work environments” are a better fit than “these environments and group-based incentives when there is little task interdependence across group members.” (Barkhi, 2008, p. 666). Individual rewards based on individual contribution motivate more than “payoff from the grand coalition when incentive is group based.” (Barkhi, 2008, p. 668). Group-based reward structures appear to increase the propensity of the free rider participant appearance.

Alge considered the moderating effect of temporal expectations on idea generation in computer-mediated decision-making versus FTF. Ad hoc teams and teams that believe they will continue to work together “reported higher openness/trust and shared more unique information than virtual teams communicating through a synchronous computer-mediated medium” in FTF. Virtual teams “were able to communicate as effectively as FTF teams in terms of openness/trust and information sharing.” (Alge, et al., 2003, in abstract) when they shared a history.

Postmes et al. conducted a meta-analysis of the effect of anonymity in GDSS. Many GDSS assume that “to remedy the problems of decision-making in co-present groups” and “remove the social obstacles that prevent individuals from attaining their full potential.” (Tom Postmes & Lea, 2000, p. 1255) that anonymity is paramount. Structured communication techniques including Nominal Grouping and Delphi informed the decision to include anonymous
capabilities in GDSS. Anonymity fosters deindividualization – “a state of decreased self-awareness that fosters disregard for social norms and conventions.” (Tom Postmes & Lea, 2000, p. 1255).

Sometimes the anti-normative may also become anti-social. GDSS believe that “anonymity and deindividualization, in addition to their negative effects, diminish the social influences of the group over the individual that are dysfunctional for decision-making” including “pressures to conform or social loafing” and “liberate the individual from undesirable group influences” and “counter excessive pressures to arrive at consensus and thereby neglect critical decision alternative” (Tom Postmes & Lea, 2000, p. 1256).

Postmes’ meta-analytic conclusions directly contradict all the assumptions about the role of anonymity and decision-making.

Overall, it appears that there is no support for predictions derived from the GDSS literature in these findings. The meta-analysis demonstrates that anonymity has no reliable impact on performance indicators of group decision-making in GDSSs. Anonymity does not improve decision quality, increase quantity of ideas and solutions, or increase satisfaction. The absence of effects was consistent across studies. This pattern of results deviates from previous meta-analyses on examining GDSS performance in comparison with unsupported groups, which demonstrate reliable benefits of GDSS use on participant satisfaction and decision quality (McLeod 1992, Benbasat and Lim 1993). The present analysis therefore provides an important qualification to these studies by demonstrating that the effects in these previous reviews probably were not produced by anonymity. Instead the positive effects of GDSS use may have resulted from factors such as concurrency (Valacich et al. 1993, Benbasat et al. 1995) or the structure a GDSS imposes on the decision making process. (Tom Postmes & Lea, 2000, p. 1265)

These results directly challenge the assumption that anonymity within a GDSS is a basis of superiority over unmediated, face-to-face interactions. The results question the believed positive impact of anonymity. The results report “a relatively consistent absence of effects of anonymity” and anonymity’s impact on decision-making is “less solid and less straightforward than often assumed”. (Tom Postmes & Lea, 2000, p. 1265). The key assumption that “group decisions are best served if the influence of the group is reduced” is now suspect and the literature has “over-emphasized the importance of group dysfunctions and their adverse impact on group productivity.” (Tom Postmes & Lea, 2000, p. 1265). Many studies report an increase in remarks – critical or otherwise – in anonymous GDSS exercises. Deindividualization because of anonymity was believed to be the source of increased comments – theory would “predict that
anonymous settings foster more unrestrained and deindividuated behavior." (Tom Postmes & Lea, 2000, p. 1267). However, the findings of the meta-analysis contradict this belief.

However, the present results show that people were most critical precisely when they were partially identifiable. Anonymity increased the number of critical remarks only in settings in which people were visually co-present (hence identifiable to an audience), not when they were completely anonymous. Hence, the present findings suggest that deindividuation did not cause anonymity’s effect on critical remarks. (Tom Postmes & Lea, 2000, p. 1267)

Postmes questions almost all the research that became part of the meta-analysis. He claims the research assumed “the groups operate in a social vacuum” and “the studies were conducted and reported as if groups have no history, no background, and no purpose beyond the meeting.” (Tom Postmes & Lea, 2000, p. 1267). They believe that the experimental subjects are likely to “approach these experiments with clear sets of norms regarding appropriate behavior for group interaction, participation in scientific research, and use of computer systems.” (Tom Postmes & Lea, 2000, p. 1267). In addition, even if these norms are not pre-existing then for groups – anonymous or otherwise – group norms are constructed.

Rains’ meta-analytic literature centers on the GDSS relationship to democratization of group processes. Citing previous research, Rains summarizes literature on GDSS as heralding the “ability to level the organizational playing field and produce more democratic processes” that “reduce inequality between group members thus minimizing barriers to communication.” (S. Rains, 2005, p. 221). He found that “GDSS resulted in increased participation equality, influence equality, and production of unique ideas” thus “GDSS leads to great equality in member communication than that which occurs in face-to-face meetings.” (S. Rains, 2005, p. 221). Rains reports surprise that anonymity in GDSS did not appear as a significant moderating factor influencing outcomes as “the body of research in GDSSs [claim] anonymity an integral component.” (S. Rains, 2005, p. 223).

To determine the effect of anonymity on source credibility and influence in computer-mediated communication, Rains used adaptive structuration theory to “determine the impact of anonymity on receiver perceptions of sources and messages in computer-mediated communication.” (S. A. Rains, 2007, p. 118). Some theory predicts that anonymous source messages have stronger credibility and other theory predicts messages will be less credible and
influential. At a summary level, Rains was unable to find a difference. Further subdivision indicates that groups are more likely to discount anonymous messages and “perceived anonymity was negatively associated with perceptions of source competence” resulting in reduced ability of the anonymous participant to “influence his or her team members.” (S. A. Rains, 2007, p. 106). Anonymity reduces group dysfunctional effects but “its possible negative impact on message receivers may outweigh these benefits” and “the plethora of benefits anonymity allegedly affords message senders might be offset by the detrimental impact it has on receiver perceptions and behavior.” (S. A. Rains, 2007, p. 118).

2.3.3.9 Influence of Task Type

Many articles refer to task attributes as mediating efficacy of decision-making processes. Tasks frequently gain classifications of simple or complex without reference to an agreed upon, determinative scale for the classification. Task attributes demand consideration as “characteristics of a group’s task have be shown to account for more than half the variation in group interaction.” (Zigurs & Buckland, 1998, p. 313). Given the contribution of task attributes, Zigurs proposed the development of theory for task, technology, and computer-mediated group support systems. The search for theory arose from inconsistent research findings that sometimes found GDSS better on complex tasks but sometimes better on simple tasks; GDSS were not sometimes better in idea generation and sometimes no worse than face-to-face; in negotiation GDSS performs worse. At best, “these findings are equivocal and reveal a considerable opportunity for enhancing our knowledge of task-related circumstances.” (Zigurs & Buckland, 1998, p. 314).

2.3.3.10 Nature of the Task

Dennis and Wixom determined that satisfaction with a GDSS used for decisions was “significantly less satisfying than using one to generate ideas, and significantly less satisfying than working without the GSS.” (Dennis & Wixom, 2001, p. 246). They believe that most systems make idea generation easy but decision-making requires the “group to either use electronic communication, which does not easily fit the demands of the task.” (Dennis & Wixom, 2001, p. 247).
2.3.3.11 Organizing Data

If a GDSS contained advanced features” to help groups organize and rank information by voting and modeling improved the group’s understanding of the issues under discussion” resulting in better decisions (Dennis & Wixom, 2001, p. 247). Though virtual teams and FTF teams generated approximately equal ideas, the “virtual teams are poor at making decisions compared to teams that work face-to-face” because they “do not have a rich verbal context.” (Dennis & Wixom, 2001, p. 249). Dixon and Wixom conclude,

For effective performance, organizations need to ensure that distributed teams working on decision-making tasks have some access to face-to-face meetings, a video conference, an audio conference, or a series of telephone calls to augment GSS. (Dennis & Wixom, 2001, p. 249)

2.3.3.12 Group Size

GDSS group size moderators indicate that larger groups took less time and were more satisfied with the process as opposed to smaller groups. Larger groups gain more with computer-mediation than do smaller groups. Smaller groups were “smaller GSS groups were significantly less satisfied with the process than groups without GSS.” (Dennis & Wixom, 2001, p. 249).

2.3.3.13 Facilitation

GSS facilitation is not uncommon. GSS groups “with a facilitator made better decisions and were more satisfied with the process than did GSS groups without a facilitator.” (Dennis & Wixom, 2001, p. 249). Research indicates that facilitators “help groups successfully appropriate the GSS into their work processes.” (Dennis & Wixom, 2001, p. 249).

2.3.3.14 Learnings from GDSS

The salient points that might inform the Delphi exercise and the design of the web-based protocols need highlighting. For this researcher, this section raised very problematic issues that for some reason do not appear in the Delphi literature itself.

One might expect that Delphi’s requirement of anonymity strengthens in computer-mediated, non-synchronous activity and those social norms and structures reduce significantly. Apparently, group norms and roles emerge in this setting. Norms influence the decision process by bounding comment. The bounding of comment strikes at the heart of Delphi and the literature
of structured communication. What is unclear from the literature is the degree to which these social constructs then limit diversity of response and encourage tendency to consensus and agreement. It is somewhat clear that expectations of being more than a one-time group strengthen the social construction. What is not clear is how long the interaction must be before the social construction builds to its perceived dilatory effects.

The research indicates that idea generation should be no worse than paper and pencil exercises. Worrisome from the research is the implication that negotiations on content of ideas generated are difficult in any form of meeting not FTF. The complexity of the statements themselves contributes to the difficulty. Even more worrisome are the findings that unless actively required through some mechanism to digest, integrate, and compare feedback from other members on discussion items that the feedback is ignored. Delphi always argues that the reasoned review of the feedback of others is the essential driving force for clarification of one’s own thoughts and that this clarification is the basis of good decisions and consensus. The research indicates that unless required – participants just ignore the commentary.

Consensus construction in Delphi remains paramount but computer-mediated decision-making experiments tend to indicate that once a proposition appears that the group “fixates” on the proposal with little change. The supposition is that the written statement “locks” the discussion. Further, it appears that quantity of commentary on an item further locks the decision. It is unclear if positive or negative make any difference – and just sheer numbers of comments appear to indicate to participants the commitment – and thus consensus – to a specific proposal. Closure seems to be a significant factor in driving satisfaction and closure may result from processes other than deep consideration.

The literature reports that ICT enhanced decision conclusions are probably no worse than non-ICT enhanced decisions. However, almost all the research focused on “game playing” where a “right decision” existed – something very different from futures foresighting. None of the research used real situation, real experts, or real problems.
This researcher was unable to find general research on the interface, interaction, or construction of decision-making software. All systems seemed to be proprietary unless they were small, academic implementations.

Member commitment appears no worse than non-computer-mediated group processes but the data are very spotty and the use of students confounds determining commitment – students need to participate for external reasons.

The lack of clarity around “task” performed further muddies the water. Clearly, whatever a “complex” or “difficult” task is – that category of task is no worse in result. Unfortunately, the literature is unclear and not standardized on tasks or the activity to accomplish the task. Thus, it will be unclear as to whether a Delphi exercise is a “difficult” or “simple” task. Too much of the literature uses satisfaction with the GDSS process as a success measure for decision-making outcomes.

2.3.4 Other Web-Based Delphi

Given the considerable number of dissertations and theses produced yearly using Delphi and its classic paper intensive process, the availability of the Internet and web based form input would seem a natural process improvement.

Turoff et al. (Turoff, Hiltz, & Yao, 2006) published a case study using an eDelphi adapted from a social decision support system. A main screen orients each user to activities. Each item developed for evaluation in the system is “voted on” using a scale from critical to unimportant. The results appear in a summary table. Descriptive summaries of vote count by category are presented. In addition, a “vote change” column indicates the number of people who change their importance ranking since the last cycle. Vote changes diminish with time and indicate convergence and consensus.
Each item in the list can be expanded to provide textual comment.
In a 1981 Delphi exercise, Turoff (Turoff & Hiltz, 1970) et al. used the exercise to determine the ideal structure of a CMC system. As early as this exercise was, many of the identified concerns would influence a web-based system. The system must be comprehensible given that “the level of effort to completely understand a rich system might be far more than beginning users are willing to expend.” (Turoff & Hiltz, 1970, p. 94). A CMC should allow users to adapt their styles of communication. The system must be predictable, protected and allow for evolution – brought on by user feedback for system improvement. The system should provide a flexible voting capability but Turoff writes “there is a great deal of disagreement about the relative importance of voting.” (Turoff & Hiltz, 1970, p. 97). This article predated the web by 15 years, but the concerns for the functionality of a system provide insight today.

Chou (Chou, 2002) designed a simple web system. The items were displayed and as rounds progressed, the mean score (not rank) showed along with a graph of distributions of responses. A text box with provided for input of comments. It appears only one round after the initial ranking was allowed. Cho used a very simple web based Delphi to contrast student activities structured by Delphi versus unstructured CMC. Cho concluded that “Delphi structure is more effective than the unstructured approach in generating unique ideas.” (Cho, Turoff, & Hiltz, 2003, in Discussion) No system details appeared.

Yao and Liu characterize their implementation of a computer-mediated Delphi as a dynamic survey. However, they preserve anonymity, they used “pen names” for participants to introduce the ability of others to identify communication, to view participant data like view on the subject, expertise, and so on. They believe that in a, “Delphi survey, there are seemingly intuitive advantages: the expert panel can obtain more understanding of why specific individuals are agreeing or disagreeing with certain concepts; experts can be prompted to consider their positions more comprehensively, which may increase the opportunity of forming consensus.” (Yao & Liu, , in Sect. 3). Yao and Liu apply Parente’s work that describes members of a decision-making group as swingers and holdouts with the former having a higher likelihood of changing due to input as opposed to holdouts that waver little. They believe that “By establishing direct member interaction channels for socializing among the panel, the dynamic Delphi provides extra
opportunity for the swingers to calibrate their individual confidence. Therefore, we can speculate
that the dynamic Delphi may be much more in favor of improving the appropriateness of
individual judgment change than the conventional Delphi.” (Yao & Liu, , in Sect. 3). Yao and Li’s
expected quicker convergence of the group with the added direct communication but they found
“a dynamic Delphi survey may result in sharp changes of individual opinions and worse
convergence of the collective group view when panelists are impacted by different local views.”
(Yao & Liu, , in Sect. 2).

Lindqvist and Nordangar created a web-based Delphi to capture a teacher’s practical
knowledge – something that is “deeply contextual.” They recognized that the typical
questionnaire appearing in a Delphi exercise lacked the versatility to capture the practical
knowledge. They turned to more qualitative responses thus turning Delphi into a “process that
could be likened to a problem solving conversation.” (Lindqvist, 2007, in De-Contextualization).
They had assumed that the experts would be motivated “to speak eloquently about their practice”
but the “method [Delphi] does not seem, despite the absence of peer pressure and our
encouragement to write freely, to encourage production of rich descriptions of professional
knowledge.” (Lindqvist, 2007, in De-Contextualization) They argue that there is a need for
queries shared by all to motivate discussion and for the moderators to edit the responses with the
caveat “even though researchers feel they have edited the relevant data with great care, the
informants do not recognize this data” resulting in an “ineffectual basis for discussion.” (Lindqvist,
2007, in De-Contextualization).

Colton and Hatcher also introduced pen names into a web-based Delphi but additionally
included “a calendar to keep the panel on task and as an archive to hold draft versions.” (Colton

Mayfield et al. created a web-based Delphi “minimized wasted time that clientele might
not find useful.” In their work, they cite Ladner et al. to conclude “web-based and traditional
paper-based survey methodologies were equally reliable for social science research” but “a
significant difference occurred in response rates” with “no difference found between the groups’
opinion on computer usage” (Mayfield, 2005, in Conceptual Framework). The Delphi experience
using a web-based approach took less time. From a practical view, they write “be careful not to change original meanings of the responses” when condensing them into items; report rank orders; and collect all statements from participants. The Ladner study cited by Mayfield concludes,

In all other hypotheses tested, no statistical differences were found. These findings indicate that Web-based surveying methodology has the same reliability and criterion related validity as traditional paper based survey methods. The findings support the idea that Web-based surveying is a valid and reliable method of conducting social science research. (Ladner, 2002, p. 49)

Ladner indicates that academicians will continue to investigate that efficacy of web-based versus paper-and-pencil surveys but an assessment based on cost and convenience for the researcher strongly favors the web-based methodology. They suggest further study of mixed-mode surveys, invitation letters, and other factors.

In a web-based Delphi used to review contract proposals, Harman and Baldwin observed, “difference of opinions was usually due to interpretation of the document” (Hartman & Baldwin, 1995, p. 248) and no important differences appeared in text comments based on area of expertise and background. Participants drew satisfaction from “the confidentiality of the origin of comments” but also the “entertainment value of the [new] process.” (Hartman & Baldwin, 1995, pp. 248-249).

Delphi’s strength is the gathering of expert opinion for complex problems. The U.S. Defense Advanced Research Projects Agency (DARPA) is funding a round-less Delphi for “improving the speed and efficiency of collecting judgments in tactical situations where rapid decisions are called for.” (Gordon & Pease, 2006, p. 322). The driving idea is that a decision-maker requires alternatives and judgments generated from multiple sources. Classic round-based Delphi and even web-based Delphi would be far too slow for real-time decisions. The experiment underway uses artificial intelligence and natural language processing to consolidate the diverse information flows from the comment data of a Delphi round. The system remains in a proof-of-concept and classified state.
2.3.5 Turoff/Hiltz Summary of Web Based Delphi

Turoff and Hiltz investigated computer-based Delphi processes. They claim Delphi’s inherent asynchronous structure is “the most important and least understood property of the Delphi” (Turoff & Hiltz, 1996, in Asynchronous Interaction) because it allows participation when desired and participation on topic of importance to the Delphi panel member. Synchronous, face-to-face problem solving usually exhibits a set of steps in which all participate at specific times with “the group forced to take a sequential path.” (Turoff & Hiltz, 1996, in Asynchronous Interaction).

Asynchronous communication “comes at an obvious price” as there is still a need to coordinate and organize. There is a need “to define a total communications structure and put it into a framework that produces a group view.” (Turoff & Hiltz, 1996, in Asynchronous Interaction).

Turoff and Hiltz state anonymity as a rule is “not one that should be considered hard and fast for all aspects of Delphi” and “sometimes the user of anonymity is carried to far.” (Turoff & Hiltz, 1996, in Asynchronous Interaction). They believe that the “expert panel” nature of a Delphi exercise requires that the exercise provide the sense of a “group that will be able to contribute valuable insight about the problem being examined” that acts as a “primary factor motivating participation.” (Turoff & Hiltz, 1996, in Asynchronous Interaction). They believe experts participate actively if they receive value as the result of the process with “the value received at least equal, in their minds, to the effort expended to contribute information.” (Turoff & Hiltz, 1996, in Anonymity). The sense of group and importance is “one reason why blanket invitations to participate in a Delphi that do not specify who will be involved and what the feedback will be to the group members often result in low participation.” (Turoff & Hiltz, 1996, in Anonymity).

Turoff and Hiltz make specific recommendations for Policy Delphi implementation:

1. Provide each member with new items that they have not yet seen.
2. Tally the votes and make the vote distribution viewable when sufficient votes are accumulated.
3. Organize a pro list and a con list of arguments about any resolution.
4. Allow the individual to view lists of arguments according to the results of the different voting scales (e.g. most valid to least valid arguments)
5. Allow the individual to compare opposing arguments.
6. Provide status information on how many respondents have dealt with a given resolution or list of arguments.
7. The role of the Delphi Coordinator or human facilitator is very minimum in such a well-defined structure. The software powers or special privileges that such an individual needs are:

8. Being able to freeze a given list when it is felt there are sufficient entries to halt contributions, so as to focus energies on evaluation of the items entered to that point in time.

9. Being able to edit entries to eliminate or minimize duplications of resolutions or arguments.

10. Being able to call for final voting on a given item or set of items.

11. Being able to modify linkages between items when appropriate.

12. Reviewing data on participation so as to encourage participation via private messages (Turoff & Hiltz, 1996, in Structure section).

2.3.6 Paper and Pencil Versus Web-Based Delphi

Bates and Cox investigated paper-and-pencil versus web-based survey “perceptions of anonymity and confidentiality” (Bates, 2008, in abstract) finding that though there were differences in perception there did not appear a difference in content response. Their experiment found, “the computer based method of administration [was perceived] to be more anonymous and confidential” and the respondents “preferred this method over the office paper-and-pencil condition” (Bates, 2008, p. 904).

Hayslett and Wildemuth joined the web versus paper-and-pencil response rate and content comparison research model. They compared the methods on “(1) response rate, (2) response time or quickness, (3) the potential for sample bias, and (4) the content of the responses provided.” (Hayslett, 2004, in abstract). They found that paper-and-pencil survey response rates remain higher than web-based response but regardless of method the usual response rates are usually less than 50% thus leaving “more than half of the population” unaccounted for in the results. They found that web-based responses are initially considerably faster but that after one week the difference disappears. As positive result, “web surveys do not automatically result in sampling bias” and “very few mode effects were found [so that] if sampling bias is avoided, responses to a Web survey would be comparable to those from a parallel paper survey.” (Hayslett, 2004, p. 89). They extend this finding to conclude that “paper administration in some longitudinal survey research, or use both paper and Web administration and confidently combine the results for analysis.” (Hayslett, 2004, p. 89) Brecko and Castens conclude similarly,
“although the completion rate was higher for paper based questionnaires, other analyses of the quantitative data show there is very little mode effect on the quality of the data.” (Breko, 2006, p. 11).

Archer investigated non-response rates for web-based surveys for in-house use finding that those that ballots achieve 62% response rates where generic needs assessments – or surveys with less decision-making power – achieve 50% or less (Archer, 2008).

Asserting that others have addressed issues of comparison of web-based and paper-and-pencil survey instruments, Jones et al. focus on the design of a web-based instrument to conclude the need for a printable copy; the longer the questionnaire the stronger the need to be able to print it; validation is accepted as an unavoidable consequence by respondents; and “respondents expressed the need to be able to save a draft” of responses (Jones, 2005, p. 11). They suggest that in the design of the instrument, the respondent’s view should not be ignored.

Fraze et al. compared response rates using web, email, and pencil-and-paper surveys. They conclude that “traditional paper surveys yielded the high response rate of 63% with a significant drop off to 43% for web-based and 27% for email.” (Fraze & Hardin, 2002, in abstract). Technical issues for email developed when the instrument used other than plain text. The research agrees with others on speed of response finding “significant difference in the days between the mail survey and the other two modes” (Fraze & Hardin, 2002, in conclusion) with mail surveys being returned steadily but email and web-based surveys came in “clumps immediately following contact days” (Fraze & Hardin, 2002, in conclusions) with the majority of email and web-based surveys being returned within two days of contact. Respondents had been pre-coded by demographic attributes like age, gender, and education – none of which appears to change response rates.

Archer used a web-based post event survey instrument and inspected response rates. In his research, he comments,

Some factors that have been found to increase response rates include personalized email invitations, follow-up reminders, pre-notification of the intent to survey, and simpler formats (Cook, 2000; Solomon, 2001). Other factors that increase response rates include: incentives, authentic sponsorship, and multi-modal approaches (Johnson, 2005). There are some factors that may influence response rate for which no formal studies have been identified. These include
the age of potential respondents, population to which the survey is administered, and the purpose of the survey. (Archer, 2008, in introduction)

He found, “The lowest response rates found in this study were from needs assessment types of questionnaires.” (Archer, 2008, in discussion). He speculates that the belief of the respondent that they can make meaningful contribution relates to the response rate.

2.4 Purpose of the Study

As has already been argued, an institution may passively adapt as its futures unfold. The contrasting approach allows for the possibility that some foresighting of an institutions’ futures may result in proactive favoring of selected outcomes to increase the likelihood of those outcomes while simultaneously decreasing the potential of other outcomes deemed less desirable.

Distance education, in any of its various manifestations, is a force in most higher education institution’s futures that drives organizational change for which proactive outcome selection is possible and necessary. Organizational change brought about by distance education is unlikely to abate. Waight, Willgig, and Wentling argue that the higher education stakeholders must participate in meaningful discussion to frame higher education’s future,

Therefore, as business organizations discuss the growing markets of e-Learning and develop business models to attract investors, the government targets policy makers, educators, employers and the public at large, while professionals organizations mainly focus on the workforce and workplace. it is imperative that all three entities become aware of their convergences and divergences. [Emphasis added] As Banas and Emory, (1998) stated, “while there is a growing recognition of and attention to distance learning in higher education, its growing inclusion in academia significantly changes the educational environment and experience. These changes need to be acknowledged and discussed by all stakeholders.” [emphasis added] (Waight, et al., 2002, in conclusions)

Consistent with Waight, Willgig, and Wentling ‘sentreaty, this study seeks a policy, long-range forecast that describes how distance education and the university will interact. What is long range? Armstrong writes “I have struggled with various definitions [and] the definition is not as clear as I would like it to be.” (Armstrong, 1985). For this study, 15 years is long range planning. History helps in picking the value. In the 1994 period the Netscape Navigator changed the face of the Internet and the revolution in the 15 year time is a remarkable story that few would have foreseen then.
Therefore, in the 2009 period, a span of 15 years forward is about as far as anyone should look.

A policy-setting forecast is part of planning. For many, Armstrong asserts, “Forecasting is often confused with planning.” (Armstrong, 1985). Forecasting and planning have different goals though each activity relates to the other. Armstrong writes,

A forecasting model can be used in an attempt to find out what the world will look like if you leave it alone … or if you make different assumptions about the future … or if you make changes. (Armstrong, 1985)

Armstrong goes on to indicate the relationship between forecasts and decision-making,

A forecast is often required whenever a decision is [to be] made. This is especially true for decisions that have long term consequences. Organizations must make decisions involving the location of facilities, the hiring of key personnel, the types of products to provide, contracts with suppliers and unions, and financing. Because such decisions must be made, it is desirable to forecast their consequences. (Armstrong, 1985)

This study proposes to identify individuals within the University of Hawai‘i system with expertise in distance education or who are stakeholders who influence decisions about distance education or thought leaders who may have no particular expertise in distance education but who influence organizational directions. These individuals will participate in a Policy Delphi exercise to identify those organizational changes that the Policy Delphi participants foresee and, more importantly, those changes that they prefer that might result from potential changes originating from further implementation of distance education.

Many Delphi exercises identify future events and assign to them probabilities of occurrence and relative time-frames. This exercise is interested in a future period but not in occurrence prediction within a relative time-frame. The goal remains the identification of events, themes, and trends and the assignment to these trends assuredness, desirability, a sense of priority, and some relative importance. Every attempt will be made to encourage divergent thoughts and to capture these thoughts because – as warned by any number of futurists and by Armstrong in his evidence-based forecasting work – futures research must encourage multiple outcomes rather than to forecast a singular outcome.
2.5 Proposed Research

Sackman – introduced in this dissertation’s literature review – challenged the Delphi methodology believing that the Delphi methodology claimed itself derived from the experimental research paradigm. Sackman’s measuring stick – *Standards for Educational and Psychological Tests and Measurements* – demonstrated that Delphi, as an experimental methodology, failed on many levels and on many attributes common to the experimental paradigm. Sackman concluded Delphi is a “specious consensus” gathering technique and that “consumers of information on the future need far better advice and protection. The future is far too important for the human species to be left to fortunetellers using new versions of old crystal balls. It is time for the oracles [referring to Delphi] to move out and for science to move in.” (Sackman, 1974) When Sackman wrote, Delphi appeared to wear the mantle of experimental research given its use of survey-like instruments, structured participant selection, and various statistical treatments common to survey research that frequently appeared in articles that used the methodology. Sackman’s comments were fair in context and period because the articles did not make clear a distinction and may, in fact, have sought the mantel of the experimental paradigm. As an alternative formulation of explanation at the time for Sackman’s critique, qualitative research was, at very best, an emergent second cousin in the social science research literature.

The literature review did not leave Sackman unchallenged. A number of other researchers questioned Sackman’s analysis. Additionally, the continued refinement in epistemology increases the credibility of qualitative research as an alternative to the rules of experimental research as applied by Sackman. Regardless, as all the authors noted, Sackman’s critique forced a much more careful look at Delphi’s methodology, which, in turn, compelled refinements.

With time, qualitative research gained credibility and Delphi researchers realized Delphi exercises to be methodologically and epistemologically qualitative rather than quantitative-experimental. Delphi is unquestionably qualitative but also pragmatically quantitative and not experimental. The quantitative attributes originate from the need to make ranking choices about items that is used to show convergence on the three measures used on each item – importance,
feasibility, and desirability. Given that the items are not construed to be a validated psychological instrument, the only appropriate numeric representations are minimum counts and relative weights. Any claim by a researcher using a Delphi exercise to the assumptions necessary for null hypothesis statistical testing is inappropriate. Delphi’s confused foundation stems from literature and researchers who may have intentionally blurred lines between Delphi’s qualitative nature and the experimental methodology to gain credence for the Delphi method by appearing to be what it was not. For far too many researchers, the use of quantitative results is tantamount to experimental research – a fundamental failure on their part to understand the design differences and the role of numbers in the communication of results. The use of numerical results is quantitative but quantitative and statistical results may be necessary but clearly not sufficient for research to deem itself experimental.

For example, many articles appear as Delphi research concentrating on statistical techniques for rank ordering the items in the Delphi instrument. Others concentrate on statistical or graphical representation of emergent consensus assuming that consensus equated to forecast accuracy. Yet others investigate change in rank order of items during rounds of the Delphi. The appearance of numbers does not magically transform Delphi into an experimental methodology nor are numbers barred from qualitative research when appropriate assumptions for their use are met.

Even within the qualitative paradigm, Delphi is handicapped. Many Delphi exercises utilize its structured communications strength for decision-making where no foresighting is involved. For example, the Delphi appears many times in medical literature as a technique to integrate practice and widely varying basic research into a treatment regimen. A treatment regimen is thus the consolidation of practice and theory as constructed by a group of “experts”.

The handicap attaches when the Delphi exercise is a foresighting exercise, which, interestingly, was the purpose for creating the methodology. Reading articles about the process of research one discovers an emphasis on humankind’s attempt to understand why and how events occurred. Research constructs explanations of evidence gathered for events that have happened. The claim to evidence’s future existence is felt by many to be nothing but speculation
completely lacking for the standards of evidence. Past tense thinking systemically entwines in the proposition of “understanding” and “explaining.” Phrases like “let me understand the future” or “let me explain the future” would seem unusual to a listener.

Take for example these paragraphs drawn directly from one of the most heavily cited and used textbooks on qualitative research. In addition, even though one author is used here, the same theme exists in other research methodology texts.

The first paragraph outlines some practical differences in the kinds of questions experimental and qualitative researchers ask,

In a qualitative study, it can be risky for you to frame your research questions in a way that focuses on differences and their explanation. This may lead you to begin thinking in variance terms, to try to identify the variables that will account for observed or hypothesized differences, and to overlook the real strength of a qualitative approach, which is understanding the processes by which things take place. [emphasis added] (Maxwell, 2005)

The following paragraph contrasts the understanding of causation in the qualitative and experimental paradigms.

Part of the reason for disagreement has been a failure to recognize that quantitative and qualitative researchers tend to ask different kinds of causal questions. Quantitative researchers tend to be interested in whether and to what extent variance in x caused variance in y. Qualitative researchers, on the other hand, tend to ask how x played a role in causing y, what the process is that connected x to y. The emphasis on understanding processes and mechanisms, rather than demonstrating regularities in the relationships between variables, is fundamental to realist views of causation, which are prominent in the current philosophy of science. [emphasis added] (Maxwell, 2005)

The following paragraph provides a clearer view of the questions characteristically asked by qualitative researchers.

Qualitative researchers tend to focus on three kinds of questions that are much better suited to process theory than variance theory: (a) questions about meaning of events and activities to the people involved in these, (b) questions about influence of the physical and social context of these events and activities, (c) question about process by which these events and their outcomes occurred. [emphasis added] (Maxwell, 2005)

The last paragraph deals with the construction of survey instruments used in qualitative research,

One issue that is not entirely a matter of realism versus instrumentalism is whether research questions in interview studies should be framed in terms of the respondents’ perceptions or belief about what happened, rather than what actually happened. [emphasis added] (Maxwell, 2005)
The important issue in the paragraphs above is not the attempt to contrast experimental versus qualitative or to provide further foundations of qualitative research versus experimental research – as these topics are extensive and books and myriad journal articles are available for any reader wishing deeper understanding. Rather, the added bolding and underlining in the cited paragraphs above indicates – research epistemology and methodology centers on the past and about events that have occurred. In a sense, Popper’s advocacy of falsification of theory by evidence (Popper, 1963), even as extended to qualitative research, demands that “evidence” be presented. Evidence only exists about events in the past.

Problems proliferate. The classical construction of experimental research calls for research hypotheses construction while qualitative research calls for research questions. Many qualitative researchers who believe that questions must arise from the intense, substantive emersion into the social milieu of the investigation challenge even the call for research questions. The former characterizes “answers to what you think is going on [as the implication of] your theory or experience” while “research questions are what you want to learn” (Maxwell, 2005) without presupposing what is going on. Whether qualitative or quantitative in design, research is constantly comparing observations to constructs – researchers continually generate theory against which observations contrast thus resulting in refinement and continual change.

The most generous and flexible theory construction is “grounded theory” as described below,

Identifies unanticipated phenomena and influences, and generating new “grounded” theories about the latter. Qualitative research has an inherent openness and flexibility that allows you to modify your design and focus during research to understand new discoveries and relationships. This flexibility derives from its particularistic, rather than comparative and generalizing focus, and from its freedom from the rules of statistical hypothesis testing which require the research plant not be significantly altered after data collection has begun. (Maxwell, 2005)

Multiple searches on methodologies commonly used in qualitative research include such titles as case study, ethnography, focus groups, grounded theory, action research, discourse analysis, critical theory, phenomenography, content analysis, surveys, interviews, and audio or videotaping but no reference to Delphi. Closer inspection of each of these techniques continues
the theme of experiencing the “now in a social context” to explain the why of complex social processes with an emphasis on “evidence” and thus a non-future orientation.

2.6 Research Questions

In experimental research, the literature review and researcher’s experience lead to hypotheses that express the plausible outcomes and relationships between independent variables and the variance observed in dependent variables. Hypotheses are falsifiable statements for which experiments to falsify them must be possible. More than being falsifiable, hypotheses emanate from the ontological structure in which they appear. Failing this, a statement might be falsifiable but not central to the argument’s structure thus rendering it the nil hypothesis and not the null hypothesis.

Qualitative research replaces a strict definition of hypotheses with the concept of research questions – questions about what the researcher wants to learn through experience of intimate involvement in the social milieu of the participants. Maxwell states that others believe “The use of explicit research hypotheses is often seen as incompatible with qualitative research” but he believes “there is no inherent problem with formulating qualitative hypotheses” [emphasis added] (Maxwell, 2005). As was argued in the misapplication of experimental yardsticks by Sackman to critique Delphi, so too hypotheses and research questions are caught as “a matter of terminology and partly as a matter of the inappropriate application of quantitative standards to qualitative research hypotheses.” (Maxwell, 2005). Questions, hypotheses, propositions, and explanations each represent nuances of the same construct. Considerable literature focus is devoted to the sequence of events in a research investigation wherein “hypotheses” appear. Experimental research methodology calls for these statements to be created prior to commencing an investigation. Making the statements early is a pre-requisite of statistical testing. These hypotheses remain unchanged during the investigation whereas qualitative research methodology allows a continuing introduction or appearance of questions including those prior to the investigation, those created during the investigation, and those created after the investigation.

The discussion above thus guides this researcher to state some of the questions for which explanation will be sought while holding open the increased probability that many
interesting “questions” [also read as results or observations] develop during the execution of the exercise. Delphi exercises result in statements about the future that are a function of the communication acts of the participants. These acts are difficult to predict in a particularistic form. Hypothesizing explanations of what the participants will construct seems to be of little value and almost presumptive. Worse, the over specificity of outcomes may blind the researcher to the valuable in-execution percolation of ideas due to pre-prejudiced expectations. There must be a balance of the researcher’s preconceptions and biases brought to the research as opposed to those that appear during its execution.

2.6.1 Question 1: What should distance education become?

This research began for this researcher with a general interest in education and the use of information and communication technology to deliver instruction. Upon reading the literature difficulties that would appear when attempting to determine effectiveness of distance education, the direction of the research began to shift. During this time the researcher began to teach and take more and more web-based courses and seminars. Some of these seminars were synchronous and some asynchronous. The question thus evolved away from comparative and reflective questioning of distance education to one that assumes the tool will exist and grow more prevalent. The growth would mean change. What kind of change would spring forward?

Since higher education is ultimately a human organization that can influence its future, the interesting question for this researcher became – what does higher education want the future interaction of ICT and its instructional mission to be? This question is a futures policy question. There will not be one answer and possibly some of the future alternatives will not even be revealed for years to come. Just five years ago no one would have guessed that Facebook would become the second largest application on the web. Thus it is quite possible that we have yet to conceive of the kind of changes that might be forthcoming.

Simply restated the most fundamental question is – what does this university want to be the relationship between its instructional mission and information and communications technology. Alternatively, whither distance education in a higher education institution?
2.6.2 Question 2: How rich is item generation by participants?

Delphi exercises, as with any structured communication intervention, generate a large number of ideas. One of the claimed benefits of structured communication is an increased generation of ideas. Tacitly implied in the increased generation is the assumption that the sheer number also includes wide breadth, significant depth, and large diversity with many outliers and alternatives that the pressure of group “norms” might otherwise discourage.

In any structured communication exercise, each participant generates ideas independent of other participants. The structure prohibits consolidation during idea generation. The process results in a significant number of overlapping items. The items must be reduced to a smaller number of non-overlapping items without losing key ideas in a participants input. This process likens to the construction of a survey but without the trappings of psychological test instrument construction. The item list constructed must reflect the interpretation of the sense of a group derived from participant input. Psychological tests do not ask the participants to aid in the construction of the instrument to be used to examine themselves.

Even though each participant’s inclusion in this exercise derived from their individual expertise and experience with distance education, participants may not have thought broadly about the organizational issues – in particular those issues characterized as “future” issues or that stem from sources other than those of their current use.

As stated, Delphi generates a significant number of items. How do these items reflect the wider issues of organizational change that distance education is likely to encourage? How well does the list of items – once cleaned and consolidated for round one of the Policy Delphi – reflect a list generated from the literature review of this dissertation that may be wider in scope given the breadth and depth of the review?

2.6.3 Question 3: How successful is a web-based Policy Delphi Tool?

The Policy Delphi for this research will be a web-based tool. After a review of the literature and searches on the web, no tool was found that met the needs of this study. A tool was constructed as a major part of this research. The tool reflects the functional and non-function requirements. The non-functional requirements are uninteresting due to their generic, cross
application foundation coupled with the fact that software development in-and-of-itself is not the object of this research but the vehicle by which it takes place. The functional requirements are more interesting.

The functional requirements stem from the literature. The literature is suggestive but not conclusive on so many of these ideas. Even if there are Delphi systems on the web, none of them seemed to address many of these functional requirements. Some of these features include:

1. Establishing credibility or expertise via a background survey
2. Use of the Wedemeyer swarm to enhance penetration of an unknown population of stakeholders
3. Introduction of items as seeded and the process by which items continue to become included as the project progresses
4. The use of feedback and comments from other panelists
5. The use of almanac research
6. How many times to panelist actually change their ranking and in what direction? Ideally, we would want to know why though that functionality has not been included.

Each of the above are interesting questions. The literature assumptions of the need for such functionality can be commented upon as each of these is monitored as part of the web-based system developed for this research.

2.6.4 Question 4: Reaction to Introduced Items

In question two, the emphasis was the comparison of the reduced items generated by the participants to a list generated from the literature. There should be a difference in the lists. Because all items are generated by the participants and reduced for further rounds are automatically included, the balance of the items not generated by participants but from literature must be identified as such and included in successive rounds.

How will the group react to these introduced items? Will the group assign different weights or reflect different interest in the items feed-in to expand the discussion?
Chapter 3. Research Methodology

Maxwell warns that the subject of believability in qualitative dissertation research is often boilerplate “borrowed from methods books and successful proposals without demonstrating that the author has thought through how these strategies will actually be applied in the proposed study.” (Maxwell, 2005, p. 107) He claims such boiler plate is tantamount to “magical charms that are intended to drive away evil [with] a lack of how these strategies work in practice” seemingly based “largely on faith in their supernatural powers.” (Maxwell, 2005, p. 107) This investigation attempts to exceed Maxwell’s expectations of validity qualifications of a piece of qualitative research. Maxwell goes on,

Validity is a goal rather than a product; it is never something that can be proven or take for granted. Validity is relative. (Maxwell, 2005, p. 105).

Validity must be “assessed in relationship to the purposes and circumstances of the research, rather than being a context-independent property of methods or conclusions (Maxwell, 2005, p. 106).

The next sections attempt to establish validity for this investigation’s Delphi application.

3.1 Internal and External Validity, Credibility, and Transferability

Research validity draws extensive debate. The existing rift in validity’s philosophic foundations remains deep, unsettled, and not argued here. The rift originates with the differences in epistemological assumptions of the experimental and qualitative research paradigms – in how each school of thought attempts to explain the world. Experimental epistemology remains the “gold standard” against which comparisons appear due to its close alignment with the physical sciences and considered as ‘true’ science.

Research validity is central as Wampold argues, “if the validity of a study is high, then there are few rival interpretations and they are relatively implausible” in contrast to a study with low validity where “there are many alternative and veridical explanations for an obtained result.” (Wampold, 2007, p. 98). Without sufficient validity, the attempt at explaining events would be useless, as just about any explanation would hold.
For experimental research genre, internal validity demands that manipulated variables be demonstrable as the proximate cause of observed changes. The manipulation of independent variables must account for the variance or change in dependent variables to the exclusion of other possible uncontrolled events’ contribution to the explanation. Where internal validity focuses within a study, external validity demands that conclusions hold for other similarly situated participants at other times and places.

The parallel constructs in qualitative research are credibility for internal validity and transferability for external validity.

3.1.1 Credibility and Transferability

Marshall and Rossman interpret validity as “soundness” and she cites Lincoln and Guba who argue validity is the “truth value of the study, its applicability, consistency, and neutrality.” (Marshall & Rossman, 2006, p. 201). For qualitative research, credibility originates with “an in-depth description showing the complexities of processes and interactions will be so embedded with data derived from the setting that it is convincing to the readers.” (Marshall & Rossman, 2006, p. 201)

3.1.2 Credibility

Credibility for qualitative research differs significantly from internal validity. Researcher’s methodologies and instrumentation impose internal validity. Whereas, credibility of a qualitative study originates from the participants themselves. The researcher can simply ask the participant to verify the observation and recorded data. The experimentalist relies on procedures and instrumentation removed from the participants who have no direct review. Qualitative studies “describe or understand the phenomena of interest from the participant's eyes, the participants are the only ones who can legitimately judge the credibility of the results.” (W. S. Trochim, 2006, in sect. credibility). Tools, process, and instrumentation so fundamentally part of the validity certitude of experimental research do not exist for the qualitative researcher in the same sense as they do for the experimental researcher.
3.1.3 Transferability

External validity establishes by repetition of an experiment at other times and places. For qualitative research, the parallel construct is transferability. Justification for transferability is the consuming researcher’s responsibility – a view very different from experimental external validity in which the originating research must qualify reuse at other times and places. Assurances for the consuming researcher in experimental research are implicit in the instrumented methodology. Given that the originating qualitative researcher is expected to document assumptions about the context of the investigation, the “person who wishes to ‘transfer’ the results to a different context is then responsible for making the judgment of how sensible the transfer is” (W. S. Trochim, 2006, in sect. transferability).

Marshall and Rossman state, “Qualitative research does not claim to be replicable. The researcher purposefully avoids controlling the research conditions and concentrates on recording the complexity of situational contexts and interrelations as they occur naturally.” (Marshall & Rossman, 2006, p. 204). They claim that the flexibility inherent in qualitative research denies to future researchers the ability to replicate the investigation and they “should not attempt to do so.” (Marshall & Rossman, 2006, p. 204). The denial stems from the essence of qualitative research’s participating in “real world changes” not in laboratory manipulation. In place of replication “thorough notes and a journal or a log that records each design decision and the rationale behind it allows others to inspect procedures, protocols, and decisions.” (Marshall & Rossman, 2006, p. 207). They contend that all collected data be kept in “well-organized, retrievable form [so that other researchers] can easily challenge or re-analyze the data.” (Marshall & Rossman, 2006, p. 207). External reviewers must be able to inspect procedures, protocols, and decisions made during the course of a qualitative investigation. Logs and journals must not only record the observed, participating subjects but must also reflect decisions and intrusions of the researcher. Exposed and recorded evidence lends to validity. The process of recording must also be reviewable. Finally, they conclude,

The researcher explicates the design methods so the reader can judge whether they are adequate and make sense. a rationale for qualitative research generally and the specific include genre in which the study is situated. (Marshall & Rossman, 2006, p. 205)
3.1.4 **Evidence versus Methods**

Maxwell summarizes the experimentalist's view as one in which methods assure validity. Marshall and Rossman argue “validity threats are made implausible by evidence not methods; [emphasis added] methods are only a way of getting evidence that can help you rule out these threats.” (Maxwell, 2005, p. 105).

Experimental research claims to remove the observer from the observation but qualitative research recognizes the intimate involvement of the researcher. In place of observer-independence, “all we require is the possibility of testing these accounts [our observations and propositions and hypotheses] against the world, giving the phenomena that we are trying to understand the chance to prove us wrong.” (Maxwell, 2005, p. 108). Thus, validity is increased – not proven or established – by outlining the threats or ways that understandings of the phenomena might be wrong. Qualitative research does not have “randomization [that] purports to control an infinite number of ‘rival hypotheses’ without specifying what any of them are” (Maxwell, 2005, p. 107).

3.1.5 **Maxwell Validity Checklist**

Leading researchers consider general approaches to the mitigation of threats to validity in qualitative research. There are a number of checklists and articles outlining recommendations but many of them are specific to the methodology employed by the researcher.

Of fundamental importance in qualitative research – and clearly a threat to validity in a Delphi exercise due to the central role of the researcher conducting the Delphi exercise – is researcher bias. Given the involvement of the researcher, removing “the researcher’s theories, beliefs, and perceptual lens” is impossible. Instead of attempting elimination, qualitative research attempts an “understanding of how a particular researcher’s values and expectations influence the conduct and conclusions of the study and avoiding the negative consequences” of the involvement (Maxwell, 2005, p. 108). The influence of the researcher’s presence may also be an active contributor to reactions of participants and a threat that is impossible to prevent. Instead of attempting elimination of the threat, “a goal in a qualitative study is to understand it [the threat] and to use it productively.” (Maxwell, 2005, p. 108). Some argue that though there is an
influence, the natural setting exerts significantly more influence on the observations thus substantially mitigating the concern. Ultimately, the researcher bias is “important to understand how you are influencing what the informant says, and how this affects the validity of the inferences you can draw” (Maxwell, 2005, p. 109).

Maxwell (Maxwell, 2005, pp. 110-113) provides a “checklist” summarized and generalized from his review of multiple other authors. His checklist is below,

1. Intensive, long-term involvement. Thinking of Margaret Mead or other researchers spending considerable time with subjects – the opportunity exists to test alternative theory and expand observations. Interactions may also be intense and involved.
2. Rich data. These data are “detailed and varied enough to provide a full and revealing picture of what is going on” including verbatim transcripts.
3. Respondent Validation: To rule out “the possibility of misinterpreting the meaning of what participants say and do and the perspective they have on what is going on” there is a need for “systematically soliciting feedback” about the data and conclusions.
4. Discrepant and Negative Cases: “The basic principle here is that you need to rigorously examine both the supporting and the discrepant data to assess whether it is more plausible to retain or modify the conclusion, being aware of all the pressures to ignore data that do not fit your conclusions”.
5. Triangulation: Some believe that collecting data from a range of individuals using different methods reduces “the risk of chance associations and systematic biases due to a specific method and allows a better assessment of the generality of the explanations one develops”.
6. Quasi-statistics: Though not breaching experimental design, qualitative research often demands quantitative presentation to support claims like those that a “particular phenomenon is typical, rare or prevalent in the population being studied”.

3.1.6 Mitigation of Maxwell’s Checklist

The following numbered item list responds to the general challenges to qualitative validity enumerated by Maxwell. The responses become part of the application of the methodology in this research.

3.1.6.1 Intensive, long-term involvement

There is no direct way to eliminate the integral presence of the researcher as the coordinator of the Delphi exercise. The concern of long-term involvement grows from the intrusion factor of the researcher that may introduce unintended changes into the natural setting. For example, a researcher who lives with a remote tribe might change the tribe’s customs by the
researcher's very presence or worse the researcher could introduce a disease for which the observed participants have no inherent immune system defense. There is one significant difference in the Delphi – the presence of the researcher is in reality the presence of the web-based Delphi program controlled in the background by the investigator. The involvement of the researcher becomes a remote secondary one.

By its very definition, the Delphi exercise is an intense but limited time-frame intrusion into a group constructed for the purpose of the exercise. The exercise ends and the influences end because the group ends. Of course, participants may change their view on the focus of the particular Delphi exercise due to participation and re-evaluation of their position during the exercise. This change is an individual, long-term change not a change of the group per se. The influence on the group is gone because the group is gone but trace remnants of individual shifts may have staying power. The threat is minimal at worst.

3.1.6.2 Rich Data

The communication acts central to the Delphi exercise are all written. Nothing is oral. Nothing is visual. Nothing uses a process outside of the required recording in the system. Nothing is prohibited. Nothing is omitted. Even when reducing the larger communication to the consolidated items, all original communication remains. All “votes” are retained. The Delphi technique does not have insight into the thought processes of individual participants and it does not monitor them in their decision-making process. The software controls the only observable artifact. All decisions of the Delphi coordinator will be recorded as well as any external panels. Almost all data are verbatim thus reducing the concern that the observed is heavily screening or interpreting the events while they are happening.

3.1.6.3 Respondent Validation:

Every Delphi round allows validation by the participants. If data are consolidated inconsistent with the views of the participants then the feedback mechanism within the Delphi provides the ability to the participants to challenge and correct the consolidation. Unlike ethnographic studies, interviews, or researcher observation in qualitative research in which there exists the requirement for the researcher to record the participant’s acts, the Delphi collects only
data flowing from the participants. Finally, the written conclusion of the coordinator of the Delphi is circulated to the participants for an “out of band” review. These reviews will be collected as participant email and appended to the report but preserving the anonymity.

3.1.6.4 Discrepant and Negative Cases:

No input will be removed. No negative reaction will be omitted. Again, the nature of the process assures that all input is capable of being preserved. All data will be retained. During consolidation to item lists from the free text input, the greatest risk appears. Items usually result from a sense of the input – a consensus so that voting and ranking can identify the most dominant thematic items. If item construction relies too much on contextual agreement from the input then critical divergent items may be eliminated. The item list construction rules will favor inclusion of more divergent views allowing the voting, consensus, and feedback process to percolate items of prominence to the group. There is no value to the researcher to reduce items or force an unnecessarily concentrated consensus. Consensus is not critical to the conclusion of the exercise. Like prejudice in publication of research that concludes with no significant difference, a Delphi that does not in the end achieve solid consensus or clear ranking of items may make the results appear weak. In some Delphi, researchers have been accused of trying to compel consensus to improve the “value” of their research. There is no value for this research to forcing such consensus.

3.1.6.5 Triangulation

This process allows the researcher to look at the same events using different methods to ensure a more unbiased view. More techniques of looking at the same events are believed to create a more complete view as no one technique reliably creates a 100% understanding. Certainly, in interviews or other similar observational investigations this capability might exist. In the Delphi exercise, there is a single method and one observed collection of communication acts by a group that disbands at the conclusion of the exercise. Unlike many other qualitative methodologies, the Delphi does not directly observe the participant, thus further reducing triangulation. Multiple sources of observation are not possible. At least, the literature review of the Delphi completed for this dissertation did not reveal any modifications of Delphi that went
beyond the standard collection. Some may argue that other approaches like a committee meeting, a nominal group exercise, or even a survey might provide the triangulation necessary. There is no reason to deny categorically and finally the possibility but, as has been argued, the alternative methodologies will not observe the same set of communication acts constructed within the group and thus hobbling the provision of a different view of the same events. Another observation would provide a different view but of different events. To believe that the same group of experts would participate in a second, repetitive process given the difficulty usually associated with getting them in the first place is not reasonable. Even further, if one argues that the future as conceived should be similar in result to the measurement by different scales of the weight of a loaf of bread then one is also arguing that future is something that is static – a completely misconceived notion. Triangulation is desirable but not possible for these reasons.

3.1.6.6 Quasi-statistics

No claims are being made about this Delphi exercise as being unique or representative of a larger population. There are no control or experimental groups within the study to contrast and compare. No instruments compare scores. The instrument makes no claims to survey research or construct measurement. The group constructs item lists in the Delphi and validity stems from this process.

3.1.7 Mitigation of Validity Issues Specific to Delphi

The previous section dealt with validity from the larger perspective of Delphi as a technique within the qualitative genre. The next sections proceed to address similar questions as they have been previously documented that are specific to the Delphi methodology.

3.1.7.1 Sackman Concerns

Sackman’s concerns appeared in detail in the literature review along with the counters to his positions. The important result of the head-to-head review results in agreement that though strict application of experimental research requirements are inappropriate that many improvements in Delphi might result from reasoned inclusion to strengthen its methodology. One should note that the Sackman exchange focused upon “classic” Delphi methodology and its
application whereas this dissertation is a Policy Delphi as a derivative that further loosens the tie back to an experimental paradigm.

The detailed claims of Sackman appeared previously. Below is a recompilation and restatement of Sackman’s concerns. His original concerns appear below if they continue to apply. However, some of his original list contained overlapping ideas requiring combination. His items not supported by research and challenged by others disappear. Some of his comments as judged by this researcher did not merit response.

3.1.7.2 Epistemological Claims to Experimental Research

Delphi is a non-experimental, qualitative methodology.

Where other Delphi research has strayed in its application of the methodology to attempt to infer the mantel of experimental research to gain credibility, this dissertation recognizes Delphi exercises for what they are and creates a strong wall of separation from claims to the epistemology of experimental research. The justifications of validity for this research anchor themselves entirely in qualitative research. The use of numerical values and some descriptive statistics will not cross the line to imply inferential statistics. The numerical values are simply tools of communication and representation – not tools of variance explanation or group contrast.

Delphi interventions are in the best sense pragmatic research. The intervention uses the most appropriate tools to attempt to achieve an explanation.

3.1.7.3 Panel Selection

The essence of Sackman’s issue in panel selection is the challenge to the ability to generalize results of a Delphi exercise due to the specialized process used in its convenience “sampling.” The process of selecting domain experts for participation is a central tenet of Delphi that draws considerable concern. Classic Delphi allows the researcher to use a convenience sample of experts identified by the researcher. The Delphi methodologies is silent on the process by which experts are located, selected, and participate.

Most original Delphi exercises are clearly convenience samples characterized by Berk as “risky business” requiring deeper understanding of the generation of data as statistical issues and
substantive issue overlap and “No amount of statistical maneuvering will get very far without some understanding of how the data were produced.” (Berk, 2003, p. 17). Berk goes on to state,

If the random-sampling assumptions do not apply, or the parameters are not clearly defined, or the inferences are to a population that is only vaguely defined, the calibration of uncertainty offered by contemporary statistical technique is in turn rather questionable. (Berk, 2003, p. 1)

If a Delphi application is going to use inference statistics to generalize – as many have done – then as Sackman and Berk indicate the panel selection using appropriate methodology is required. Of course, if strict selection methodologies apply then one is conducting a survey not a Delphi exercise.

This research makes no such claims to applicability of inferential statistics and the attendant assumptions.

3.1.7.4 Instrument Creation

From the response to Sackman, Goldschmidt made clear that the instrument constructed in a Delphi exercise is not required to comply with strict rules of psychological testing.

In this research, extreme care avoids the inference that the item list reduced from participant input is a psychological test and measurement instrument. Support for construct validity of the item list derives from the Delphi exercise itself. The panel constructs its reality through their communication acts and translates them to the item list. The direct correspondence between the collection of communication acts and item list is as good an approximation of construct validity as one for intelligence or anger or some other more general construct.

Delphi exercise item lists make no claim of instrument reliability where reliability demands repeatable, consistent results. The communication acts are an ever-changing target unique to the exercise participants who are considering a construct – the future – which itself is not static. Requirements of test-retest reliability, multiple forms, inter-rater reliability, and internal consistency are not reasonable in the Delphi methodology context. The possibility that item lists as finally constructed could, through normal instrument validation processes, participate in an unrelated, future processes is not excluded. The result could be valuable to others who might take a more classical instrumented approach.
Again, the constructed item list appears in form to be a psychological instrument but the Delphi methodology makes no such claims. There are published Delphi applications that do treat the item list as if it were a psychological test and measurement instrument. This mistake is not made in this research.

In a Delphi exercise, the item list originates with participant input but other processes refine the list. Guidelines for good item construction will be constructed and applied to abstracting from participant input. The use of a set of guidelines for item construction that originates from good practice in psychological test instruments is not inconsistent with Delphi methodology. The purpose remains creating the clearest communication without proving the clarity using strict, psychological testing procedures.

3.1.7.5 Statistical Treatment

The previous paragraphs have already addressed key foundational issues. This research will use appropriate descriptive statistics and numerical results where applicable.

3.1.7.6 Efficacy of Group Processes and Alternatives

At its inception, Delphi proponents made many claims reliant on emerging tools of structured group communication. Those claims in retrospect show themselves somewhat overstated. The early Delphi methodology advocates, like any other emergent technique’s advocates, overstated its efficacy. Some efficacy claims demeaned existing approaches like committee meetings or other classical group procedures.

The additional thread of concern within Sackman’s critique challenged the efficacy claims. Sackman asked for proof that the product of Delphi exercises matched future outcome. He argued that Delphi advocates made compelling rhetorical arguments for efficacious tool use but demonstrable comparison of predictions and results remained in short supply.

The issue remains open and contentious. The fundamental issue for decision makers is more practical. There is a class of problems that demand thoughtful consideration of alternatives for which no better method appears to exist except the reliance upon the solicitation of expert
thinking. Within this context, the Delphi methodology remains the preeminent methodology regardless of the well founded concerns.

3.1.7.7 Feedback, Forcing Consensus, and Accuracy

Returning to the original definition of the Delphi, Linstone and Turoff stated, “Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem.” (Linstone & Turoff, 2002, p. 3). The structured group communication process claims that anonymous experts reviewing a list of items, who are able to give and receive feedback, will converge upon a solution thought to be better than that of other procedures. Sackman aimed many of his critique at the assumptions and causal relationships buried in the previous sentence.

The literature review in this dissertation explored all these relationships and assumptions and found, like any statement, there are cases of support, cases requiring clarity, and cases in opposition. Many assumed feedback from other experts became a singularly powerful force of change and reconsideration. Maybe it is. However, many researchers found feedback ignored and participants simply refined their positions because they had the opportunity to do so. The paramount causal position of feedback as the motivation and driving force of change, reconsideration, and refinement of positions is not

Group processes often result in accurate outcomes but there are no guarantees. There is also no clear relationship between group consensus on an item and the accuracy of the outcome. There are so many subtleties in the research that make generalization impossible. However, the counter conclusion of uselessness implied by Sackman remains unsupported. In many practical applications, the only tool available is expertise and expert opinion. There are examples cited in this dissertation of the use of Delphi as a tool in building a treatment regime from basic research and current practice shows promise and utilization in the medical field.

This research will not fall into the trap of causal relationship yielded by the methodology to produce outcomes. The literature review established that the argument is complex but the reality is a need for some way for organizations and decision-makers to chart a future and fund
potential outcomes. This Delphi application and its final analysis of the results use careful construction from the evidence collected.

Statistical feedback herein is descriptive. No graphs demonstrate consensus building around a central tendency. No claims that consensus drives accuracy appear. When ordering of items is requested by the panel member, the data will be displayed in the order but the numerical values will not be displayed.

If the final application allows participant questions, the answer will almanac in style and content. If an item exhibits some controversy, all sides appear in the response.

3.1.7.8 Expert Prediction versus Prediction of Experts

Sackman did not summarize well his epistemological concerns with Delphi though he referred to scientific predictions many times. The central question remains – is the Delphi application a scientific prediction?

Armstrong expresses in a much clearer way this concern in a number of recent articles by contrasting scientific prediction with the prediction of scientists. The former characterizes from sources like experimental research or econometric tools like moving averages, regression, and the like. Scientific prediction is evidence based. Armstrong and others developed an audit list to determine how well a particular prediction measures against principles he believes are necessary for credible, “scientific”, evidence-based forecasting. The latter is a collection of opinions of scientists potentially engaged in doing scientific prediction. Armstrong and others constantly seek, and have found, results that indicate that evidence-based forecasting provides explanation. In contrast, the prediction of scientists is seldom accurate. The Delphi literature focused on accuracy concurs with the conclusion that group prediction has an uncertain accuracy.

The accuracy concerns mediate the results and add caution to transferability.

3.1.7.9 Containment of Researcher Influence

Without overstating the obvious, the involvement of the researcher in so many phases of qualitative research -- in this Delphi exercise in particular -- demands considerable consideration to avoid the introduction of significant bias. Luckily, the Delphi methodology allows for a number of checks and balances bias at the points most suspect.
3.1.7.10 Bias

Bias is a central attack on qualitative research. Bias is believed to be “the intentional skimming [away from its counterpart] that which is considered unskewed or not biased.” Implicit in this general definition of bias is the view that what is not biased in reality or truth.” (M. A. Meyer, et al., 2001, p. 2). There is always a difference between the report and the thing reported and “our memory for the event, our perceptions of the present, and our anticipations of the future can never be objective because they arise from an inevitable subjective and personal model of events.” (M. A. Meyer, et al., 2001, p. 2). Bias opens direct attack on the credibility of an investigation. Meyer writes, “Regardless of whether or not bias presents problems, a study is open to criticism unless it addresses bias through experimental design. Bias must be monitored, or controlled, or analyzed for its impact.” (M. A. Meyer, et al., 2001, p. 41).

As stated, the role of the researcher is paramount but other sources of bias drawn from a yet larger list are most likely to affect a Delphi application. Meyer (M. A. Meyer, et al., 2001, starts on p. 42) presents a long list but the most applicable are summarized and restated below,

1. Motivational bias – Subjects consciously or unconsciously accommodate responses to please the interviewer.

Mitigation: This application will reduce the presence of the researcher. The interviewer is, in reality, the software system with which the participant interacts and pleasing software should not have the same affect.

2. Cognitive bias – develops when the researcher’s world view is not supported by the results of the investigation. Cognitive bias is a primarily a problem of the researcher but some research in Delphi indicates that cognitive bias is a motivator for participant drop out.

Mitigation: Until the research report is written to summarize the panel’s input, there are no large or critical results for the researcher to respond to. The panel receives a written report of the findings and this panel makes further comment.

3. Social pressure – the influence of the researcher on the participant during data elicitation. This is primarily a problem in face-to-face or direct contact.

Mitigation: The Delphi’s structured communication approach and a reduction of the presence of the researcher contain social pressure.

4. Group think – stems from the “basic need of people to be loved, respected, and recognized” and “individuals in groups conform to a greater degree if they strongly desire to remain a member, if they feel satisfied with the group, if the group is cohesive, and if they are not a natural leader in a group.”

Mitigation: The operational assumption of structured communication is that the procedures – in so many phases of the process – mitigate groupthink from the listing
of alternatives to the consolidation of results. Groups in Delphi are also short-lived and anonymous. Groupthink has only limited opportunities to manifest.

5. Wishful thinking – likely to occur in any elicitation method. “if experts have a special interest in an answer, wishful thinking will likely occur. In general, wishful thinking will be most pronounced when the experts do not have to explain their reasoning. The experts’ highly optimistic responses are checked by having them disaggregate the problem and explain their problem solving.”

Mitigation: Individual participants are not challenged by the researcher. Recent literature in Delphi encourages anonymity but pen names so that participants may challenge individual lines of argument by a participant. No specific check or balance can exist.

6. Anchoring -- humans – researchers and participants – tend to select an original response for a particular situation and continue to hold the response and modify successive responses from this anchor. Anchoring is “one of the major sources of inaccuracy in forecasting of future possibilities [because participants] extrapolated from old patterns that no longer represented the emerging or future patterns.” The Delphi is particularly susceptible to this source of bias. “Anchoring is most prevalent in situations where the expert is not likely to experience the opposite bias of being influenced by the knowledge engineer [or group feedback directed at the participant] such as in the Delphi method.” Seldom do people “throw the baby out with the bath water” but instead they cling to the baby as long as possible.

Mitigation: Following the literature, pen names are introduced to allow a kind of cross-examination but participants still have the option to be completely anonymous. To challenge an anchoring response, either participant or researcher would have to first identify the response, which assumes that other see more than the participant does and then the researcher or group would have to formulate a challenging response. Further, the original participant would have to agree to engage in the continue dialogue that challenges their anchored assumption. Anchoring will remain a source of bias. The planned introduction of items into the Delphi from the literature review by the researcher aims at expanding the worldview and potentially provides the opportunity to un-anchor some views.

7. Underestimation of uncertainty – humans are generally very uncomfortable with high amounts of uncertainty and they systematically attempt to contain it. For foresighting exercises, this bias is particularly problematic as the future is by nature highly uncertain.

Mitigation: Like anchoring, this bias is baked into the participant and will not be undone. The participants in a Policy Delphi provide categories of “certainty”.

3.1.7.11 Delphi Bias Sources

For Delphi, there are a number of specific areas of the methodology that are most likely to incur bias,

1. Expert Panel Selection
2. The methodology of the particular Delphi methodology and application
3. The consolidation of feedback where such warrants change in items
4. The introduction of “feed in” almanac data from questions or background data
5. The maintenance of the item list
6. Final Report

3.1.7.12 Oversight and Review

Most researchers understand and accept bias’ existence. To mitigate the influence of individual bias where item lists are created, where feedback is consolidated and reviewed for item change, where feed in of data or the answer of questions is required, or where the item list is to be created or maintained, a small committee of researchers will either do the work or review the work of the main researcher.

The original plan for oversight was to use Delphi itself. After all, the activities of the oversight committee are themselves decision-making and should structured communication techniques be the “better” way then applying them would be an obvious extension. Unfortunately, the cursory review of modifications to the base Policy Delphi methodology and to the constructed Delphi software does not sufficiently match. A completely different software application would have to be developed. Even further, the Delphi process as a meta-process to the basic application is likely to slow down the functions needed and thus change the Policy Delphi itself.

Where item creation is required, the committee uses a guide developed on writing good questions. The guide is a consolidation of research and advocacy articles on the process of writing questions for psychological research instruments and survey instruments. The Delphi item list is not a series of questions and codified responses are not the central demand so the literature on writing questions is, at best, the major literature source for writing good statements that behave like questions.

The creation and maintenance of items from textual feedback is another source of resource bias. Tools exist to aid in the analysis of text. However, most of these tools are for much larger sets of textual material. The Delphi feedback is normally short and to the point. Essays are highly unusual. The responses are generally focused. The committee creates rules. The rules become part of the record.
3.1.7.13 Remaining Sources of Bias

The committee mitigates some of the most obvious sources of bias. Other concerns follow.

3.1.7.14 Panel Selection

The selection of participants in a Policy Delphi is particularly knotty if one makes claims of representativeness of the participants to some larger group. Such claims align with the experimental paradigm where g and assignment provide the basis for the extension. As discussed, this research is not interested in such claims and, even if it were, the substitution of transferability for external validity would be more consistent. The burden of reuse of the findings falls to the consuming researcher.

Many Delphi exercises use panels constructed by the researcher conducting the exercise. Researchers claim the best position to know which candidates are most expert or most involved in the exercise’s focus. The researcher must mitigate personal bias and make clear arguments for panel inclusion of a participant. Goldschmidt summarized this concern that “the ‘experts’ selected may not be representative of all experts and may even be purposely biased by the Delphi monitor to increase the probability of achieving a prediction that is favorable to his interests” but “such behavior is obviously unprofessional and should be deplored but it is more related to ethics than to science.” (Goldschmidt, 1975, p. 201). A researcher might rely on the documentation requirements previously discussed but experimental research on "hidden populations" provides an alternative approach to panel selection that might adapt to a Delphi exercise.

The literature review already outlined foundations of the approach in the section on snowball nomination of members of a “hidden” sample frame.

This Delphi application combines the functionality of Wedemeyer’s Future Forecasting Engine and Heckathorn’s Respondent Driven Sampling. The former establishes the solicitation procedure while the later mitigates the bias from researcher selection of the panel. As a reminder, Heckathorn’s approach for the identification of hidden populations intends establishing the basis for the use of inference statistics by arguing that it effectively identifies the sampling
frame and its representative membership. Those claims are unnecessary for Delphi but the approach meshes with Delphi providing needed improvement and reduction of bias introduced by the researcher. In addition to reducing researcher bias, the RDS process may increase the diversity of the identified population. Diversity of the panel aids mitigation of anchoring as explained,

Diverse experts play an important role in cases in which the experts forecast future events or situations. In forecasting, the tendency is to anchor to the status-quo situation and not extrapolate sufficiently in considering the future. Having multiple experts with different viewpoints helps the group overcome the human tendency to anchor to one, usually conservative, reference point. (M. A. Meyer, et al., 2001, p. 87)

As stated, RDS exists within the experimental genre of social science research to generate an acceptable approximation of a sample. Heckathorn writing on bias states,

An ideal sampling procedure yields not only a sample independent of its starting point, but also an unbiased sample of the underlying population, with a known degree of consistency from which confidence intervals can be computed [however] because of the absence of probability samples in studies of hidden populations, a more modest goal has been to devise means for drawing samples that produce ‘a good cross-section of the target population’ or the ‘coverage of heterogeneity in the target population’. (Heckathorn, 1997, p. 190)

The detailed examination of RDS appears in Heckathorn’s research. This research is not attempting to claim experimental credibility but the essence of RDS may assist in panel diversity. At least, as a procedure, RDS reduces the role of the researcher as a major source of bias.

The list below summarizes the key elements of RDS applicable to a Delphi exercise. The literature review explains each list element.

1. Selection of initial seeds. RDS suggests the identification of well-connected initial seeds. Well-connected seeds should be able to recommend more additional participants and more participants who in turn have a wider network. For this research, initial seed selection will be solicited from leaders at each campus identified by administrators.
2. Incentive structure for initial seeds. See incentive discussion.
3. Incentive structure for all other participants. See incentive discussion.
4. Objective Traits. A sampling frame in experimental research is rigidly specified. The traits for inclusion in this study will be part of the solicitation process. Though most panel members self identify, RDS requires qualification for entry. Qualifications are required.
5. Be either a user of current distance education or in the position to influence the decision.
6. Have thought about the issues of a future university and an interest in shaping its future

7. Be a member of the University of Hawai’i system as either faculty, staff, student, or other directly related person

8. Eliminate Duplicates. The software will eliminate duplicates based on email address and prevent seeds from adding duplicates.

9. Ends when “saturated or stable composition with respect to goals.” Concluding the selection process relates directly to the panel size selection in a Delphi. The recommendation of the literature is to achieve a large number to mitigate attrition. The literature indicates that Delphi attrition may exceed the rate of attrition expected in comparable research. The initial goal is 150 participants.

3.1.7.15 Motivating Participation

Motivating panel members to recommend others and motivating panel members to stay engaged is a challenge in Delphi exercises. Motivation in clinical trials and other social science research typically uses remuneration. Many researchers believe that the motivation of experts and stakeholders differs from the general population due to professional and personal involvement. Meyer writes,

Generally, we believe that paying the expert should be reserved for those situations in which there are no aspects of the project that can be used to motivate participation or in which participation requires major investments of the expert’s time and thought. Focusing on how the intrinsic aspects of the project can be developed to encourage participation can produce more effective motivators and improve the design of the project. Paying experts for their time should be the last resort for several reasons: it is expensive; it may attract one type of participant and slant the results; or it may have unexpected effects on the participants’ view of the project. (M. A. Meyer, et al., 2001, p. 89)

Meyer’s comments that paying experts in this context may not yield the desired results to motivate participation remains reasonable but idealistic. The practical dictates that some reward be included within the ethics constraints of the University.

The proposed method, consistent with the RDS approach, is,

1. A pool of $2000 will be available for distribution to the campus’ scholarship fund. No payment will be made to an individual participant.

2. This decision may introduce significant challenges in gaining participation. The problem with attempting to do individual participation lies with the Human Subjects requirement that the payment not be conditioned on actual participation or completion. If the solicitation went to hundreds of potential participants but only a few responded the payout to actual participants would be so minimal. If the amount were fixed, then cost of this project could be far beyond the actual resources available.
The above are subject to approval of the Human Subjects committee of the University and the availability of some of the prizes. The commitment of $2000 comes from the researcher.

3.1.7.16 Participant Identification and Solicitation

Reviews of Delphi frequently focus on how panel members were identified and solicited. Practical research suggests that the solicitation message itself is important. In this Delphi exercise, the solicitation uses email.

Each email can be personalized to come from another community member during the swarm phase. The content of the personal message is up to the sender but the personal message increases the likelihood of participation.

In addition to the personal message content, the balance of the solicitation should include the attributes below:

1. Reason for the contact. Phrase in a way designed to motivate. Recognize expertise. Be inviting and inclusive in the language. Do not be technical. Emphasize the importance of participation.
2. Recognition – from the selection, from working with other experts
3. Something new, different: most people enjoy a break. Active curiosity.
4. Need for meaning: interested in how their work fits in large picture
5. Who is sponsoring the project
6. How much time and effort, over what period of time, and when will it start
7. How they were selected and who referred
8. What tasks in detail will they perform. Avoid technical jargon. They may seek help on understanding the tasks – provide “training” or help. Want to know if beyond their scope. “In general, the experts can be told that they will not be forced to give judgments when they are not expert because this practice would detrimentally affect the quality of the data. Emphasize that the project’s goal is to collect judgments that are based on careful consideration and experience”
9. Will judgments be anonymous and how will this be maintained.
10. The anticipated product of the project and the experts access to it
11. Whether participation will be required or voluntary

3.1.7.17 Panel Size

The literature is inconclusive as to implications of the size of the panel on the results of the exercise. Unlike experimental research where the sample size is calculated, no similar process exists. A major concern for panel size is the likelihood of significant attrition and thus the
literature recommends a sufficiently large panel to compensate for attrition, though no specific guidance appears. Secondly, the literature recommends a sufficiently large panel to insure diversity of ideas. The suggestion contained in the literature is the intentional structuring of participants whose ideas compete and diverge from one another as simply adding sycophants has little decision-making value or information content.

In this specific exercise, the panel size originates with the Wedemeyer modified swarm or snowball generation of members. How many participants the swarm will identify and how many invited experts actually agree to participate cannot be known ahead of time. If 150 participants are identified and agree to participate then there is a reasonable chance of reaching some valuable results. This is a significant threat that cannot be controlled. Capabilities in the software allow re-contacting participants who have agreed to participate but who might drop out.

3.1.7.18 Initial Instrument Creation

Classic instrument construction from test and measurement theory does not directly apply to the construction of items in the Delphi instrument. Fundamentally, the instrument is the product of the group’s conceptualization of its purpose. As argued by Goldschmidt in his analysis of instrumentation according to the American Psychological Association, the standards do not apply. Therefore, a quasi-instrument remains that participants validate in succeeding rounds. Validity originates with the panel and how the panel evaluates items and contributes clarifying or extending items.

Delphi methodology assumes participants will review the items. If items are not representative of participant understanding, feedback mechanisms must allow clarification. To make item modification, the researcher must detect that the feedback demands item re-formulation and the researcher must accommodate the change. There is not a perfectly unbiased researcher observing the communication acts of others thus moving the construction of items closer to the participants reduces the filter.

One approach to eliminate researcher consolidation is to allow the participants to vote on each item presented. As ideal as this may be, the practical implications might be the bogging down of the entire Delphi into an item construction exercise to the detriment of thinking about the
implications of the items for the institution’s future. The process could degrade into an endless debate on semantics.

Ultimately, someone has to make a decision. There is no way to construct a meaningful item instrument except via professional judgment in a somewhat autocratic way.

3.1.7.19 Item Construction

The literature on psychological testing and the construction of items for a given instruments tends to indicate that the construction of a good item is an art form and one that comes with experience. There are statistical tests like factor analysis to determine how respondents group items. Grouped items may be evaluated against the construct sought. There is no statistical test for the quality of an individual item. Since the Delphi is in appearances only a psychological test instrument, the strictest rules applicable for item construction and review go beyond the needs of the Delphi.

Further, in discussions with Murray Turoff during his review of the items he suggested that some items might be intentionally provocative. Instead of items that sound removed or abstract, a provcational position should be taken. For example, instead of “Students should be encouraged to have their own computers” the more strident item would be “Students shall own their computer and the University will not provide public access computers.” The audience needs to be encouraged to react, not to passively “vote.” He suggested that a large number of items be seeded. Review by a number of people whose position made them equivalent to potential participants drew the opposite response arguing that smaller numbers of items would be a better direction. They found a large number of precise items stifling. They suggested far fewer items that read more reasonably as policy items with a larger scope and timeframe.

A has been noted elsewhere, a small panel of reviewers will act as a quality control check and balance for items. Some guidelines for the reviewers is necessary and follows.

3.1.7.20 Item Construction Guidelines

This list will be available for use by the panel that is reviewing items for inclusion in the study.
1. Each item suggested for inclusion should be edited to the minimum degree possible to ensure fidelity to the author’s language. The trust to interpret and rank the items will be left to the participants.

2. Items should be reasonable in length.

3. Items must be phrased as “wants” or “desired” future states of distance education.

4. Items that deal with specific implementations will be generalized. For example, “The University will use Sun Unix computers.” Besides the fact that Sun no longer exists, this item is about implementation and technology. If the item can be saved or edited to generalize then that will be done. If not, the objections will be noted and returned to the writer to request clarification.

5. Items should never be so broad as to cover too many concepts. Avoid compound conclusions or compound arguments toward a conclusion.

6. Reduce the number of items. The larger the number of items the more daunting for the participant and the more likely they will not respond.

7. Organize and structure the questions categorically.

8. Include in the question all information necessary to understand the question including background, definitions, and assumptions or provide access to these.

9. Avoid compound questions. Avoid compound, complex sentences. Write more and shorter sentences to express ideas that are more complex.

10. A good sentence contains a single thought.

11. Simple questions are less likely than complex questions to require subdivision.

12. Experts require less information to formulate answers from simple questions. Typically, only those data, definitions, or assumptions that the experts require are provided.

13. Regardless of a question’s complexity, careful phrasing is always critical to the expert’s understanding of the questions.

14. Remember that questions carry information. Remember that we assimilate “new information when it fits within the context of the previous knowledge or mental models.”

15. Unclear results in multiple interpretations -- “experts are likely to interpret the question differently and give answers to essentially different questions.”

16. Do not imply an answer.

17. Keep the question on one dimension.

18. Do not presuppose a certain state of affairs.

19. Do not use emotionally loaded or vaguely defined or emotion words.

20. Do not make the response to one statement conditioned on a previous statement.

21. Avoid double negatives.

22. Avoid recall questions.


25. Avoid emotional language.

26. Avoid prestige bias.
27. Avoid double barrel items like “I support distance education and the increase of salary to faculty for the workload it causes”

28. Avoid leading items.

29. Avoid false premises. In items meant to stir discussion, the premises should not be false or misleading.

There is no easy procedure for combating biased questions – the best strategy is to be sensitive to this issue and to scrutinize carefully the wording of the question.

3.1.7.21 Initial Seeded Items

One of the research questions proposes examination of the richness of items generated by the panel in comparison with a list of researcher created items derived from the literature review. Each item constructed will be subject to the same review as items created from the initial round.

In a “classic Delphi”, the researcher constructs the entire instrument. Modifications in this research allow the participants to create the initial issues list. The researcher identifies differences between the lists of items created from the participants and from the literature. The difference of items will be included as if from the participant to broaden out the consideration.

3.1.7.22 Questions and Initiated Feed-in

Some research indicates that participants generate questions that require a response from the Delphi coordinator. For example, a participant might ask for a comparison of higher education budgets for FY2008 as compared to FY2009 in order to complete their feedback or ranking decision on an item. Alternatively, they might ask for enrollments in distance education courses by campus. The software will provide the capability to post a question. Even though it is a form of feedback, posting a question is a process separate from the feedback.

Only questions that have almanac or fact based responses will be replied to. Questions that call for speculation, projection, or conclusions outside of known evidence will all receive the same message stating that the question is out of scope. The response message will always be the same.
Most responses will come from the literature review. Some questions might appear that are not in the literature review. If almanac answers are available and quickly identifiable from sources other than the literature review, the answer is given. In all cases, citation to the source appears. In general, contested answers will not be provided unless positions are easily identified and documentable.

3.1.7.23 Rounds, Consensus, and Calendar

Delphi literature investigates the relationship of the number of rounds and consensus building and the relationship of both to accuracy. No dominant theme emerges. More rounds generally result in more consensus but many researchers argue that this is just regression to the mean and worn down panelists. Many argue that rounds should only be stopped when consensus is achieved. Unfortunately, what consensus means in a Delphi context is also problematic. Is consensus item based or cumulative over all items? Should means be used or some other statistic? This discussion then leads one back to the instrument itself that was never intended for the statistics applied to it to measure consensus. Of more concern is the finding that consensus and accuracy of forecast do not necessarily correspond. Put simply, a panel could reach consensus on the wrong answer. However, the additional challenge for consensus, accuracy, and rounds comes from the nature of the research conducted to determine the relationship. Most of the research was “clinical”, did not use “experts”, and used problems for which answers are known which is in direct violation of the purpose of Delphi.

As for Policy Delphi, there are suggestions in the literature for the number of rounds being more than a “classic” Delphi. However, there are also practical arguments that Policy Delphi are even more difficult to keep interest of the participants.

For this reason, the arbitrary decision is made to have one round for clarifying question generation and two additional rounds of feedback, review, and ranking. A pseudo-fourth round will be the review by the participants of the report developed as the result of their work.

In addition to limits on the rounds, a number of articles indicate that structure relates to a panelists decision to participate. The literature suggests that a timeframe for rounds, feedback, and digestion of ideas be part of the solicitation. A calendar is required and the calendar relates
to rounds and consensus but the practical matter for participants is the need to explain the time commitment they are making.

3.1.7.24 Computer Mediated Communication

Will the use of a web-based Delphi change the results of a process that is typically a paper-and-pencil exercise? The research is inconclusive. Resistance to typing shrank as communication using web-based applications increased. The public is being forced to use a keyboard.

In a computer mediated social group, anonymity can be preserved in the absolute sense but research indicates that social identity does appear after interactions. The research indicates that pen names in place of actual names in threaded conversation do not dissuade participation.

There exist many guides for writing web-based applications but that research gives minimal specific direction to a Delphi application. Human computer interaction at the software construction level reduced its complexity with the powerful tools available today.

There are no compelling or overpowering findings that indicate a web-based Delphi would prejudice the results. No behavior or panel results are expected to be different in a paper-and-pencil versus a web-based Delphi. The social identification in anonymous interactions may be a confounding variable. These issues are less likely in paper-and-pencil exercise but the short-term, focused nature of the project will reduce the bias introduced.

3.1.7.25 Attrition

The literature indicates that attrition rates for Delphi exercises can be “significant.” The sheer numbers failing to participate in additional rounds is often in the 50% range. There is also a problem with attrition where those representing the most diverse opinions from the “norm” of the group tend to drop out.

Some research indicates that participants attempt to evaluate the other participants. If they decide others are of equal or better standing as to their expertise then they tend to continue. If they decide others demonstrate less knowledge, they tend to drop out. Participants use the sophistication of feedback by others to build their judgments.
All these factors combine to threaten the results. To counter this threats the following measures were planned,

1. Repeated emails
2. Enticement to continue
3. Appeal from their sponsor
4. Incentive prizes for quick response – into a pool
5. Overall grand prize – 2 night neighbor island
6. Limit the rounds

After a preliminary review by Human Subjects Committee, all of the enticements had to be dropped as rewards and incentives violate current policy. The only remaining options became the limiting of rounds and repeated emails. Neither of the remaining options provided strength of enticement. Lack of enticement increased the difficulty of securing participants.

Meyers believes that the involvement of the participants will mediate the attrition. She states,

“generally, if the individuals have control over a process, they feel better about it and will lend support to methods used or conclusions reached. This optimistic attitude stems from the belief that if I did it, it has to be good. Also, if the experts judge that the project will be done in a manner that will bring credit to them or that their reputations will benefit from being included in the company of experts, they will be more willing to participate. (M. A. Meyer, et al., 2001, p. 89)

3.1.7.26 Expertise Survey

The ability to encourage participation and to gain good feedback is often conditioned on the perception of a participant that other members of the group are as expert as the participant is. The ability to judge the quality of the other participants by accessing certain demographic data that establishes expertise without exposing the person’s identity increases group identity.

In addition, certain demographic data will be valuable to justify the diversity of the panel. The same demographic data might be interesting for some thematic comparisons of responses.

To this end, the registration process of a participant will include a demographic survey. Each participant can elect not to participate or to participate but not have their results shown to others. Results may be shown but are never specifically identified with the individual.
3.1.7.27 Results and Write Up

The final element of bias is the production of a final report. Goldschmidt makes that comment that “investigators, Delphi or otherwise, are above ignoring all empirical data and simply adjusting the outcome of their experiments.” (Goldschmidt, 1975, p. 207). The obvious control is the careful reporting of all data collected. Finally, the report itself will be circulated to those interested for any additional comment.

Ultimately, the faculty committee and the researcher all expect honesty and control in the presentation of results.

3.2 Informing the Design

The Policy Delphi exercise in this dissertation will be conducted via a web-based application developed specifically for this exercise. Software development requires design statements that drive the functionality displayed by the system. The literature and most textbooks infer that “user requirements” dominate the structuring of software and that “user acceptance” comes only with meeting “user requirements”.

User requirements derive from the literature review. Professional requirements originate from multiple sources that are not the focus of this research and would clutter the work. These include security, stability, modern object oriented design, infrastructure, database, and so on. Readers will have to accept that the researcher is sufficiently skilled to include necessary elements without documenting them here, as this is not a software project.

3.2.1 Fundamentals

The Delphi software application written for this dissertation relaxes many of professional requirements. The application is intended to be one shot. That does not mean that good practice goes out the window but it does mean that many of the more detailed considerations will not come into play.

The most important issue for this application is that it implements those capabilities needed for a Policy Delphi and it does so in a manner that is easy for the user and reliable.

The points that follow are Policy Delphi application requirements that originate from the literature.
3.2.2 Solicitation – Panel Selection

The process of identifying initial seeds is a consideration external to the software. The software will expect a list of names and unique email addresses. The software will mass email an invitation. The software will record each invitation. Email addresses must be unique and the system will enforce the uniqueness. Seeds who then invite others to participate will use the same capability.

3.2.3 Accepting the Solicitation

Email addressed to invited participants will contain the link to the web page that allows them to accept or reject participation.

3.2.4 Human Subjects Consideration

If a user clicks the accept button, then the human subjects participation form is displayed. The user reads the form. If they accept the conditions of the study and agree to participate, they check the “yes” box and they type in their email address into a text box.

The human subjects form will discuss all the details of the study. The form explains the handling of contributed participant data.

3.2.5 Validation of Expertise

Once the user accepts, a short survey form will appear. The panelist may opt out of the form. If the panelist agrees, then they answer the set of questions developed to establish their expertise or stake holder position to be included in the study.

3.2.6 Anonymity

Original Delphi assumed complete anonymity to avoid the deleterious effect of power relationships in-group face-to-face decision-making. Research indicates that social order forms in anonymous groups as long as they have a more than one-time interaction. Further, research indicates that pseudonyms or pen names encourage more interaction and exchange of during textual feedback. Participants can trace the thought stream of feedback contributed by specific participants.
During registration, participants will be asked to provide a pen name or to decide not to provide a pen name. Regardless of choice, further personal identifying data will not be provided like name, email, and phone number.

3.2.7 Data Retention
All data are recorded. All responses in each round are recorded. All questions and requests are recorded. All data are backed up.

3.2.8 Calendar
The main page will display a calendar to indicate the current progress through the Delphi.

3.2.9 Voting and Ranking – Display Options
Each item has three attributes – importance, reliability, and feasibility. The panelist will be able to rate the item on each attributes. In classic Delphi, items tended to focus on a limited construct related to panelist expertise. Distance education is a much more diverse construct that touches many different categories of expertise. Ranking will occur within each category of item and by each of the three attributes separately. The panel member can request the rank order. Items sort on a weighted distribution that is the composite of the three basic attributes. The composite weight score is not displayed. Composite weight differences is not displayed. This is consistent with research that indicates that feedback of means is distortive in the judgment and ranking process.

3.2.10 Feedback Options
Feedback and feed-in might be sent or received by a panel participant. The information flow could be handled by a generic feedback mechanism but separating the functionality provides more clarity and allows better review of behavior.

3.2.10.1 Notice of Directed Feedback
The literature indicates that completely anonymous feedback is problematic because participants try to establish the consistency in an argument. To establish consistency, an identification is required and thus the penname. Panelists also want to direct their feedback to specific feed-in. For example, if a panelist states “x will never happen” then another panelist might want respond by identifying the source of the quote. Participants are looking for a way to
cross-examine or clarify. The system will provide this capability if the participant agrees to allow identification and selects a penname.

3.2.10.2 Option to read Feedback and its Display

The web page will not automatically display the threaded discussion of an item. Panel members will see an indication that a discussion thread exists. Participants display the indicated thread. Participant options allow the showing of either all or just new feedback.

3.2.10.3 Option to check Background

Background data supports a statement. For example, enrollment figures or budget figures may provide context to a statement. If background data exists, then the panelist notification supports the display the data. There is no interaction except to read the background data.

3.2.10.4 Option to ask a question

The panelist clicks a button to ask a question.

3.2.10.5 Option to challenge Item wording

Though panelists might challenge the wording of an item via general feedback, the challenge button provides a more specific feedback to the researcher. This feedback is available to others and is round specific.

3.2.10.6 Submitting Feedback

General feedback is always available. The feedback is unlimited in size. Feedback revision is allowed. Feedback is round specific.

3.3 Software Development

The web-based Delphi software developed for this investigation implements the functional requirements contained in the introduction, the literature review, and the research questions of this document.

The software design and development for this research evolved in many major phases. Phase 1 was the working prototype demonstrated during the proposal defense. Phase 2 and beyond reflects the alterations and refinements based upon review by software professionals and
potential participants. Each major rewrite or redesign blurred together during the development. Attempting to number specifically each phase is impossible. Phase 2 is a shorthand reference to all post Phase 1 development. For example, one forced phase revision originated when Microsoft released a new version of the Visual Studio development environment, the C# language, the support libraries, the Windows server, and the SQL database server. Many might argue that making a version move in the middle of application deployment makes little sense. However, in the world of the web so much of the infrastructure changes rapidly. Failing to remain up-to-date with major releases introduces other risks. The safest path appears to be the path of change.

Each phase caused significant reworking and fine-tuning. Most of the changes resulted from questions arising from user review. Some change originated because the reviewers in any given phase were not necessarily the same reviewers as those of previous phases. When you are dealing with software professionals, content professionals, and even novices, the ability of reviewers to concentrate during the delays introduced in the software revision processes often exceeds the reviewer’s willingness to cooperate and the time reviewers have to assist. The reviewers were often software and design experts known to this researcher from professional contacts over many years. All reviewers requested anonymity and no compensation.

A stable and somewhat complete Phase 1 Beta appeared at the Proposal Defense. Refinements continued throughout Spring 2009 with the objective to collect data in early Fall 2009. A substantial part of the non-aggressive timetable to implementation based itself on the recommendation of some members of the dissertation committee who felt that social and political environment of the University in that semester appeared soured by budget cuts, staff reduction threats, and the massive economic downturn. Soured environmental factors might sour results. The period following this delay was unfortunately further complicated due to serious health concerns for this investigator. On the brighter side, the delays pushed the data collection beyond the tumultuous U.S. economic crisis period with the hoped for results being a less negative worldview of the potential participants.

Two driving forces characterize Phase 2: the need to finish this work and fundamental changes to a continuous Policy Delphi as suggested by Murray Turoff. Phase 2 received more
professional review and comment from software experts. The Turoff changes included the
elimination of one evaluation attribute. Turoff’s original work indicated four evaluation factors. He
suggested that a Policy Delphi eliminate the entire concept of “rounds.” Rounds exist in all classic
Delphi exercises.

Turoff suggested software modifications to capture participant behavior data. The
captured data might prove valuable to shed light on some issues plaguing the Delphi technique. A
number of event capturing snippets introduced into a code revision capture panelist behaviors.
For example, the system captures each change of ranking for each attribute of each item. The
system also captures the use of feedback/comments and the use of research requests. Even
more importantly, the use or non-use of certain of the literature driven, functional features may
prove to be less than useful when deployed in this application. The data might lead to the
elimination or to a refinement of operation of these extended features. For example, the literature –
not necessarily Delphi literature – describes the desire to have available the credentials of a
panelist. The literature suggests that the participant credentials are part of the weighting factors
used to assign a value to a panelist’s comments. This research will not be able to get directly at
this question but tangentially will collect usage data of the function that does expose the
expertise.

As the web-based tool is central, some of the issues that have emerged are not
necessarily technical tool issues but are structural issues related to the Policy Delphi. So far,
many of these issues have not appeared in the literature either in Delphi or in a more abstract
fashion in other fields and thus remain open to resolution. These appear in the limitations and
further research section.

The conclusion of the Phase 2 updates changed the underlying structure in a sufficiently
significant format to compel re-notification of the University’s Committee on Human Subjects. The
Committee in a response email noted the revision submission. The Committee required no
changes and the research remained in the exempt category. The Committee has access to the
system for deeper review.
3.3.1 The Web Tool

As with all software, requirements drove the functions included or intended for inclusion in the implementation. Software literature identifies two major contributory streams for requirements – the functional and non-functional. The software literature definitions and distinctions between these two major sources remain contentious. Non-functional can be viewed as those requirements that are extractable across nearly every software solution including such categories as architecture, servers, security, development style, language platform, and other categories that stem from the professional information technology knowledge. Unlike non-functional requirements, functional requirements speak to the system’s functioning requirements. For example, a general ledger system must follow the rule that within a transaction the sum of all credits must equal all debits.

The non-functional requirements of this software development have remained constant. Given the researcher’s familiarity with the Microsoft Windows platform, all tools, servers, software, system software, development software, and supporting applications remain Microsoft based. The infrastructure decisions include the Microsoft dotNet framework including Windows Server 2008 R2, SQL Server 2008 R2, Visual Studio 2010, Internet Information Services, and the C# language. The software runs on the Server 2008 behind a firewall. The server is dedicated to this single purpose now. The server is a relatively new Hewlett-Packard machine with large memory, fast processors, and more than ample disk space. Stress testing, using available tools, has indicated that no bottlenecks are anticipated for the numbers of potential simultaneous users. Data base indices and access have been reasonably fine-tuned for rapid response. Most tuning is automatic and suggested by the systems software. During the execution of data collection, additional tuning steps may take place.

3.4 Summary of Phase 2 Reviews

The following paragraphs note some of the issues identified by the reviewers. Approximately 25 different professionals participated in software review over a period of more than a year. The well-established reviewers all work in the software industry. This researcher, through professional association, knows each reviewer. The reviewers did not have expertise in
Delphi or Policy Delphi but many were able to comment on the system in abstract as detached from the specific implementation at hand as they have implemented many systems over their lifetime in the industry.

3.4.1 Instruction on System Use for the Participants

The web-based Delphi is a non-trivial software application whose options and features derive from the literature requirements. The target audience in a Policy Delphi is the stakeholders for the policy. Significant differences exist between the expert use of the expert Policy Delphi system and the subject matter expertise of the stakeholders. The mental model of a tool to gather input about the future state of ICT and higher education is complex. Reducing the dissonance between the subject matter stakeholder’s understanding and the system’s requirements is always a large challenge. In the case where the stakeholder has claim to knowledge about the subject matter but little attachment to the process, the challenge increases. As is well documented in the literature, the amount of code dedicated to user interaction facilitation is often more than half of the total effort. The cursory review of the current implementation supports this notion.

The most intuitively appealing but highly simplistic argument calls for a system design so simple, so clear, and so obvious that the stakeholder would understand the system’s functionality without need of instruction or training. Simply, the system functionality is intuitively obvious. The mental model of the participant and the model of the software coincide. Unfortunately, this is not the case. The dissonance increases as the Policy Delphi’s voluntary participants are stakeholders in Distance Education not experts in the web-based tool use or in Delphi itself. If the target audience were the traditional Delphi with a limited number of experts, the software might better match their common experience. One would hope that if the target audience closely aligns with the content that the audience would forgive some system complexity or annoyances.

Two approaches to explain and teach the system were tried with the reviewers. The results appeared mixed and this dissertation could spend its entire effort in simply dealing with the issues raised. The first approach to explaining the system presented a textual tutorial. Each tutorial appeared as a series of web pages that explained the entire system prior to actually interacting with the system. This approach also included more on-page instruction specific to
available page actions. The second approach to explanation replaced the textual material with video explanation.

Neither approach drew conclusive support. Each had strengths and weaknesses. The fundamental point made by every reviewer was that each of the explanations was too long, too complex, and had too many options. These questions and comments stir considerable interest but this researcher had to remind himself continually that this project was not a human computer interaction project. Such interests could totally swamp the intent of this research. The review of interaction remains a highly seductive area of investigation carefully ignored for now.

Conversely, reviewer requests to identify instructions or tutorials to remove resulted in few if any responses. Fundamentally, a complex system is hard to explain and a simple system is simple to explain. A Policy Delphi can be very simple when it is nothing more than a survey. However, when feedback between users, multiple ranking phases, input from research, and other literature driven requirements appear then the system complexity increases. Possibly with substantial time, interfaces and patterns of interaction could be found that might simplify the interaction but these were unclear during this development. One future alternative might be a system that grows in “smartness” and “capabilities” as the user increases familiarity through use. That is, the system would expose functionality as the user learns instead of presenting all the functions at one time.

The reviewers suggested – make it simple. Use text but limit the material to one page of introduction. Keep the wording simple and avoid too many technical comments. Avoid precision in favor of generality and wider understanding. Do not spend too much time on the extended features. Note the existence of the feedback features and attempt to expose those features in the appropriate context. How do you achieve balance? The reviewers continually press for less and less explanation while at the same time they demand explanation.

Of particular concern to a number of reviewers was the threat imposed by the detailed explanation. The reviewers were concerned that mortality would increase because the system appeared too challenging. Several commented that the nature of the participants as volunteers would only increase the likelihood of drop out. As noted earlier in the Delphi literature review,
mortality is a serious issue plaguing Delphi. More than 50% dropout occurs. One of Sackman’s criticisms of Delphi identified mortality because he felt that mortality invalidated the claimed rankings and consensus reported.

The reviewers agreed that overall and specific explanation remains important. They seemed to favor brief text in place of video. They commented that video did explain in detail all the options but the explanation also significantly lengthened the time commitment one is making to understanding and participating.

The final decision for this implementation uses a shortened text format. Reviewers suggested that the primary explanation and tutorial use text but that a detailed video be available.

Video posed another significant problem. Too many novice reviewers did not have the correct CODECs, drivers, add-in components, or viewer software. Novices frequently lacked audio capability that many developers assume. Lack of audio immediately renders the video approach worthless. Too many issues about system stability introduced by the attempt to use video decreased the value of the introduction. Again, this research did not have time to investigate the findings behind the use of video. Video also demands more scripting and a more stable software implementation. Attempting to change or improve the software system and then matching the system back to the video became a daunting task. Video options are not included in the final implementation.

3.4.2 Sequence
Most systems begin with grand designs and expectations. This investigator’s original conceptualization was expansive and generalized.

The system was originally entitled Delphi. There are other web-based systems with a similar name. This system was renamed Ho‘ailoa for the Hawaiian equivalent of ‘oracle’. The design called for a system that would support multiple simultaneous Delphi panels with all the maintenance necessary to support those panels. A comprehensive security and authorization

5 Definition provided by Wayne Chang as a native speaker, language teacher, and noted kumu hula.
system would separate roles allowing distributed maintenance of each Delphi instance even though the instances executed on one server using one database. Capabilities would include the ability to create a Delphi panel; declaring a Delphi panel type like public or invitation only; maintaining participants; maintaining items; and all other normal functions expected in system maintenance.

The participant login combined with a selection of a specific panel would result in a menu. The menu would control the activities relative to the context of the panel declared.

Figure 4 Original Conceptual Design reflects the original design. A first attempt implemented this high-level diagram. As natural as this interface is, early testing indicated that general participants had difficulty with a menu from which they would select options. Further investigation on their difficulty indicated that they simply were overwhelmed and did not know which menu items to select. Reviewers demand more simplicity.
A successful login would lead to a web page allowing the participant to select the panel made necessary because the system supported multiple, simultaneous panels. This page is Figure 5 Panel Selection. This is an early draft of the working multiple panel selection. Panels appear related to the nomination to participate. A list of public panels displayed. Again, the participant had to make a selection.

The efforts at generalization disappeared approximately two months after the proposal defense. The complexities of generalization caused delays in the development of later phases of a working prototype. The code necessary to revisit generalization and the database structure itself still contain essential remnants that would be necessary to create the more general model.

The more realistic functional system requirements demanded a set of capabilities including login, password declaration, human subjects, item contribution, participant nomination, demographic data, certain tutorials and explanations, and finally the ranking page. All pages except the ranking page were relatively less complicated to construct.
Figure 6 Original First Pass Sequence illustrates the participant sequence through the system. Each of the major steps is illustrated. Details on the content and functionality of the ranking page appear later in this document. To explain the ranking page here distracts the discussion’s focus from the emphasis on sequence. Important in this diagram is the general flow from start to finish.

Phase 1’s Original First Pass user sequence strongly favored a deep tutorial on the entire project. A major motivation for the tutorial was the provision of sufficient explanation to the potential participant to judge the Human Subjects concerns that the participant might have and potentially to stir interest in the project. The explanation erred on the side of detail. Each potential participant saw a complete picture of the process and logic of this Delphi implementation. Because this first attempt used text and screen shots, the initial tutorials required three pages.
A sample of one of the early text only versions of the attempt to introduce the system appears in Figure 7 Original Detailed Introduction to the Project.
One draft of the introduction included a flow chart. The chart graphically presented the system flow while associated text explained each major box. See Figure 8 System Flowchart.

![Figure 8 System Flowchart](image-url)
A sample of the detailed tutorial is in Figure 9 Original Detailed Tutorial – Page 1 of 3.

The purpose of the tutorial at this early stage was the clarification of the overview and the hoped for familiarization with the entire system.

The next page in the sequence was the human subject’s notification. Some interesting issues arose with the human subject’s sequence. In order to get to this page, the participant login authentication must exist. There was some question as to whether or not this was in the truest of spirit of the human subject’s process. The system security design dictated the login because, as

---

6 Authentication is the login sequence that creates the access rights to the directories and resources. The choice to exclude pages from these constraints exist but many recommend that such exclusion be used only sparingly. This is a technical and security issue.
explained, the first page after the login is the description of the project. The application software secures the project description page to authorized users only. Thus, after the login the user can view the description page followed by the informed consent notification. Users can opt in or opt out. Whether the potential panelist opts in or opts out, a database record is needed. Database access requires an authenticated context. An informal request for clarification to the Human Subjects Committee resulted in a comment that they felt the identification process was not an issue. In particular, the research was exempt so it caused them less concern. Figure 10 Human Subjects Informed Consent illustrates the form approved by the human subjects committee.

The next form in the sequence allowed the participant to set a password. One of the committee members requested that a “no password” option also be part of the password process. The logic is clear – in today’s world, there are too many systems demanding too many
passwords. A password recovery process is available in the system. Figure 11 Password Screen depicts the functionality of the password screen.

![Figure 11 Password Screen](image)
Following the password screen, Figure 12 Demographic Survey demonstrates the appearance of this page. There are 25 demographic questions and thus this illustration is just a screen capture to give a sense of the look and feel of the screen. The items appear on one page. Original design showed one demographic item per page but the user reaction was clear – one item per page was simply too time consuming. When queried, the evaluators said they would rather put up with a long page that scrolled quickly versus an item-by-item presentation.

Figure 12 Demographic Survey
The two pages that followed the demographic survey completed the remaining tasks of requesting that participants nominate others and that they contribute items of concern to them.

Figure 13 Nomination (Swarm) for other participants illustrates the implementation of the nomination process while Figure 14 Item Contribution illustrates the item contribution.

Phase 1 First Pass Sequence changes significantly in the Second Pass. A pass is the completion of the login authentication. Thus, the first pass is the first ever login while the second
or later passes are all logins after the very first login. The first pass attempts to gather all initial data. The second pass skips all but the login. Participants proceed directly to the ranking page. The ranking page contains links back to all previous pages thus becoming the central page of the system.

![Figure 15 Original Second Pass Sequence](image)

Phase 1 was field tested on a number of volunteers who were a mixture of software design people, generic college graduates, and some students. The predominant feedback focused on system complexity with particular emphasis on the text used in the tutorials. One complaint focused on the fact that the tutorials came early but their use came later thus the reviewers felt they had to remember and recall too many details. System design did provide links back to tutorials from later pages but this did not satisfy the users. Consistent with complex systems like Office 2010, the text material was minimized and videos were introduced to explain the system. The videos contained the same material as the text. The videos also contained live action shots of actually clicking thru the various pages and responding to the demands.

Figure 16 Video Replacing Text in Explanation shows the revised flow once the videos appeared.

The fundamental change was the swapping out of detailed textual tutorials and explanations for the equivalent video. The general sequence remains approximately the same.
Many of the volunteers who participated in the first review participated in the second review. Some additional non-technical participants were solicited to gain feedback less on the technical but more on the logic, explanation, sequencing, and time demands. Two very clear messages came back – the video was an excellent addition but unlike the textual material that allowed quick move to the next steps, the video required complete review. The reactions claimed the videos length exceeded their willingness to listen. The videos required too much memorization and the participants reported that they had to re-watch the videos too many times.

During the Spring 2010 semester, Dr. Murray Turoff reviewed the software. He requested an addition to the sequence. He requested a status page that would inform the participant of current participation rates of other panel members.

The flow sequence redesign includes the status page. The status page requires viewing in the sequence during first pass and during all subsequent passes. See Figure 18 Sequence Revision including Status Page.
Dr. Turoff, in several emails, argued that the status page was important to encourage participant interaction and continued support. He argued that participants’ motivation increases when a participant observes that others participate actively. A revision made to the second pass sequence displays the status page immediately following the login and before the ranking page.

During the summer 2010, requests were made of colleagues who graduated from the University’s CIS program and a number of other professionals. Some of the professionals were educators while some of them were web designers. The sequence presented to them eliminated the videos in favor of text and kept the Turoff recommended status page.

Stability of the system was no longer the focus of the review. Instead, these reviewers looked at details of language, color, sequencing, and flow. The group strongly suggested the elimination of considerable explanatory or contextual material. The group proposed a number of
changes to structure and content of the pages and some revision to sequencing. They insisted on less verbose, detailed on page instructions in favor of shorter, more precise language.

As an example, the multiple page original explanation of the system and the tutorial were reduced to the following shortened text that is presented immediately after the login on a first pass. Though this text appears long, it is about 1/10\textsuperscript{th} of the original explanatory text. In terms of reading time versus video time, reviewers found written text requiring far less time. Readers could quickly read the text below – and possibly skip what did not motivate them – whereas the video required a stronger beginning to end viewing.

Relative to the sequence, the reviewers raised pointed questions about the item contribution. They felt that to request contribution of items – even though one could display all the items on the item contribution page – just made little sense to them. The reviewers argued that participants would want to get to the central activity of the system and that item contribution and review at this point was too early. Because item contribution remains an essential focus of the
system, the request for contribution moved to the closing sequence after the participant has completed rankings. One reviewer also commented that the system ended without a thank you for participating page. Reviewers also felt that the status page was confusing and did not motivate them to participate. They did not understand its value and would have to spend too much time attempting to understand what it was trying to communicate.

The final sequence resulted from the Summer 2010 reviews.

![Figure 22 Summer 2010 Full Final Sequence](image)

The second pass sequence returned to its original flow from login directly to ranking.

![Figure 23 Summer 2010 Final Second Pass Sequence](image)
3.4.3 **Item Ranking Page**

The previous discussion focused on the overall sequencing of the system of interacting web pages that comprise Ho‘ailona. Regardless of how the flow might be structured, the Item Ranking web page is central to the Delphi experience. See the conceptual model Figure 24 Main Page Conceptual Diagram.

In the simplest model, the item-ranking page would do just as it is titled – rank the items. The purpose of the ranking, as explained in the literature review, is building consensus and prioritization. Many, if not most, Delphi studies use the prioritization process. The literature indicates that an alternative approach is rating each item on three attributes of desirability, importance, and feasibility. The design complication grew due to the research driven requirements to provide a commenting, feedback system and the ability to request research. Even further, the sequence or flow through the system evolved as described in the previous section. This evolution drove the need for the ranking page to support the ability to revisit functionality including revisions to the demographic survey, adding to the nomination or swarm list, and item contribution. In addition, links to tutorials were required. Figure 24 Main Page Conceptual Diagram illustrates this functionality.
The actual main page used in the summer testing appears in Figure 25 Screen Shot of Ranking Page.

![Screen Shot of Ranking Page](image)

**Figure 25 Screen Shot of Ranking Page**

Figure 26 Explanation of Ranking Page Major Area introduces each number area on the ranking screen.

<table>
<thead>
<tr>
<th>Area Number</th>
<th>Brief Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Clear labeling of this page as the “Main Page” so that explanations can talk about the “main page”.</td>
</tr>
<tr>
<td>2</td>
<td>Every other page used in the preliminary process plus the “Contact Us” and “Withdraw” have been added. The panelist can branch out to any previous function and then return to the main page</td>
</tr>
<tr>
<td>3</td>
<td>Instructions for the page are on page and simple.</td>
</tr>
<tr>
<td>4</td>
<td>The detailed task is hidden to reduce page clutter</td>
</tr>
<tr>
<td>5</td>
<td>Categorical selection of items is provided to reduce the total length of the page of items.</td>
</tr>
<tr>
<td>6</td>
<td>Items are listed with flags to indicate new and “click here” to guide the user.</td>
</tr>
</tbody>
</table>

**Figure 26 Explanation of Ranking Page Major Area**

Reviewers provided considerable input on the ranking screen. The comments shifted focus to the ranking page as the overall sequence issues reduced in importance. During the
development, page sequencing dominated reviewer response. As the sequence simplified by removing preliminary pages, reviewers increased comments on the main ranking page. Reviewers suggested that the ranking page be more central and be raised in importance. The sequence change required more options on the main ranking page to allow access to functions removed from the initial sequence.

3.4.4 **Item Ranking**

Central to the entire Delphi is the ranking of the items. Presenting a list of items upon which some complex actions occur is a challenge. There are two basic approaches. First, present each item one at a time on individual web pages. Compel the participant to view a series of web pages each focused on the particular item. Additional navigation demands emerge to allow forward and backward movement through the item list. Many systems also provide random return to a particular page. Very early reviewer reaction was swift and clear. The reviewers did not like the single item per page approach. They found this approach cumbersome and very time consuming. More than any other concern, the participants felt that they could not see the “whole” picture on one page. They expressed the need to weigh and think about one item against another. The one-at-a-time approach did not easily facilitate this need. The reviewers were concerned that they spent too much time flipping between pages to contrast their answers.

Reviewers expressed a need for “flags” to indicate which items changed since their last visit. Participants said they did not want to revisit an item unless something had changed. The participants also expressed concern about their ability to interpret the results of voting. The original prototype simply reported vote counts. The participants asked for simpler, graphic result presentation. Beyond voting, the reviewers also noted the need to know if comments changed since a prior visit. They requested a feature not to show all comments but just those that they had not previously read on the particular subject. As to research, they requested indication of when research was available.

The Ranking Page redesign addressed the issues developed above. All item presentation changed to a list format. A simple list consumes little page length. However, the list in this research requires both the ability to rank the item and to view the ranks of others. If all the
material displayed, the page length grew to an unreasonable length. To compensate, a client side Java script control - the accordion control - displays as a list while additionally can activate to show the ranking and tabular results. Users control the display of detail.

Figure 27 Items Displayed As List demonstrates the unexpanded accordion list view. This option allows the participants to see the entire list. The "Click Here" icon entices the user to expand the item. Technically, the user could click anywhere on the item text body itself but some experimentation showed that such instructions were confusing. "Click Here" made the clicking location obvious and dramatically reduced the explanatory text.

![Figure 27 Items Displayed As List](image)

Figure 28 Item Expanded in Accordion Control demonstrates the result of clicking on a list item. The accordion control expands to show the ranking and other activities.

![Figure 28 Item Expanded in Accordion Control](image)

Each item ranking uses three attributes: desirability, feasibility, and importance. The original implementation had a submit button exposed as part of the three ranking options.
Feedback was very negative about having to press a submit button for each displayed item and its rankings. Many reviewers commented that they forgot to press the button. Failure to click meant their choices never recorded. Redesign compensated by eliminating the need of a save or commit button. Each radio button in the ranking list used Java script to passively post user selected values back to the server without the panelist taking action. The additional, inefficient network traffic places increased demand on the server but enhanced the evaluators’ approval. Tests on the server indicated that the system burden due to frequent updates showed only minor delay. During this period, a new server acquired from Hewlett-Packard significantly increased processor speed and system RAM. This server replaced the server used in early implementations. Network traffic increased but the small packet size and the elimination of page post backs allowed for a responsive server environment.

Figure 29 Screen Shot of Ranking Items illustrates the three categories of ranking exposed by the accordion control. To remind the reader, these three attributes and their values appear in the literature review.

![Figure 29 Screen Shot of Ranking Items](image)

Originally, the ranking displayed using numerical values including the mean, standard deviation, and frequency. Reviewers did not like the numeric only presentation. Charts replaced numeric rankings with charts displaying the current ranking. Chart increments are in a percentage bin large enough to allow reasonable display in the area provided without significant sacrifice of data integrity. The bars display bin frequencies not true percent. Due to the complexity of the code, bar charts update only on a page refresh. The use of the accordion control eliminated the
need to refresh each post back. Java script development might accomplish the task of instant update. When asked, reviewers thought the proposed process reasonable. They felt that a page refresh of the charts sufficient as instantaneous update of the charts added no value to their decision-making.

As was pointed out in the literature, many Delphi use “averages” to summarize participant ranking. Instead, this study uses an index.

Figure 30 Screen Shot of Ranking Results shows a mock up with fictitious results from as presented to the evaluators.

![Figure 30 Screen Shot of Ranking Results](image)

3.4.5 **Within Control Menu Items**

Relative to the current item, there are two context sensitive actions (1) adding or reading comments or (2) making research requests and reading the results of research. The links to these actions display at the item level and not at the level of the entire list. Figure 31 Item Context Action Menu illustrates the location of the context actions. The add comment, add research, and explain actions links are always active. The actions to read comments or read the results of research are active when there is data present.
3.4.5.1 Adding Comments

Figure 32 Item Context Action Menu- Detailed View makes clear the clickable content menu options. This figure also demonstrates how links change appearance when active.
A click on “Add Comment” shows a web page to allow comments entry. The current item under consideration presents so that the participant does not have to remember the wording of the item when responding. Multiple comment entry applies by saving each entry individually. With the completion of all commenting, the “Finished” button is used.

![Figure 33 Entering A Comment on an Item](image)

Figure 33 Entering A Comment on an Item
The display of comments shows in Figure 34 Reviewing Comments. The display is clear. Each comment displays in sequence of original entry. If the participant selected the “Show Unreviewed Comments” then only those not previously viewed would appear. Under the comment, an abbreviated version of the demographic data of the panelist making the comment displays.

Research indicated that panelists want to know the “expertise”, “bias”, or “viewpoint” of the commentator to use this in judging credibility and value of the comment. The participant also has the choice to click on displaying the full demographic response.

Figure 34 Reviewing Comments

Figure 35 Individual Detailed Demographic Responses illustrates the effect of the user who is reading comments and wants to see a fully explained version of the demographic responses.

Figure 35 Individual Detailed Demographic Responses
3.4.5.2 Adding Research

Figure 36 Adding a research request is the screen format to add a research request. The format follows the same format at making a comment.

![Figure 36 Adding a research request](image)

3.4.6 Access to All Previous Pages from the Ranking Page

Links are available at the top of the ranking page to allow participants to return to pages previously visited. Figure 37 Ranking Page Access to Previous Pages illustrates the appearance of the top of the ranking page.

![Figure 37 Ranking Page Access to Previous Pages](image)

The actions provided at the top of the page complete the concept of the ranking page as central to the Ho'ailona system. All other pages and actions become reachable from the ranking page. The human subject's update page was not included because one cannot reach this page without having agreed to participate. All tutorials previously presented link from this page. Detailed discussion of these actions is not required here as the links simply return participants back to pages whose functionality appeared previously in this document. No new or different actions appear during the revisit of any selected functions.
3.4.7 Adding Panelists

Reviewers and Dr. Turoff all commented about the capability to add participants during the course of the Policy Delphi. The work of Wedemeyer on “swarming” as a modified snowball sampling could be extended to a continuous process of inviting new participants as the discussion evolved. Again, classic Delphi would fix the panel prior to the commencement of round 1. The fixed panel assumed that one had either identified the statistically valid population or built a sample frame to be used in a pulling a random sample. The revised, organic approach encouraged inclusion as the process continues.

The mechanics of dynamic addition appear in the software as part of the original plan to have stakeholder identification prior to the launch of the first round. The modification to the software required that a link be added to the Ranking Page to allow participant to revisit the add participant function. Though a mechanically simple change, a number of procedural questions arose that appear later in future research.

Once the Ranking Page allowed access to the “add participants” functionality, reviewers raised the issue of why not allow access to all functions. Some reviewers suggested that even the initial login sequence begin at the Ranking Page with indications of what was incomplete. Some argued that the diagram used to explain the system should in fact be the system Ranking Page or central theme page.

The software revision now supports access to all functions from the Ranking Page.

Ranking items is central to Delphi. Classic Delphi depended upon a fixed list of items constructed by experts prior to the ranking rounds. As has been stated, many confused these items with psychologically validated instruments.

3.4.8 Participant Attributes and Expertise Identification

The reviewers commented on the participant demographic attribute identification web page. A number of them felt it was too long. A number felt it was reasonable. Quite a number of the reviewers asked why it was included. The reviewers asked what use the data would be.

The reviewers suggested a complete review of the list. Because they are not involved with higher education or with Delphi, they commented that there appeared to be too much jargon
in the demographic attributes. The suggested a careful re-reading and revision. Attributes were eliminated and choices clarified.

The reviewers also questioned the need for the attributes because the only attribute that appears in the Ranking Page is the classification of participants as Faculty, Student, Administrator, and such. If this is the only classification to be used to present results, then what value do the other attributes have?

3.5 Round-Based Versus Continuous Delphi Processes

Classic Delphi implements “rounds” as explained in the literature review. The preliminary round is usually nothing more than gathering items for consideration followed by the consolidation and cleaning of the submitted items. In fact, many Delphi do not identify the construction of the item list as a function of the targeted expert panel. Some Delphi use initial items generated solely by the facilitator thus eliminating the preliminary round. Rounds 1 to N involve the circulation or re-circulation of the fixed item list and the ranking or rating of the items, the reading of feedback, and the inclusion of additional feedback. Rounds cease when the responses converge – when the variance between rankings on item values shows little change between rounds. Most Delphi reported a stop somewhere between round three and five. Delphi rounds may also cease when the facilitator stops the process.

Determining when to terminate is more problematic for the Policy Delphi. There are more items than a classic Delphi. Convergence is unlikely because the goal is not a point estimate of a value at some future state.

In discussions with Dr Turoff, he suggested that the proper execution of a Policy Delphi would be a complete scraping of the round-based Delphi in favor of a “continuous” format. Rounds do not exist. The Delphi would begin and continue until no new participants joined or no one modified their responses. Several attempts at literature search on the subject of a continuous Delphi did not return supporting literature. There are Policy Delphi web sites but the mechanics found were more likely clones of the classic Delphi than of the continuous Delphi. Given Dr Turoff’s position in the Delphi research community and the appealing nature of the continuous Policy Delphi, a software rewrite implemented the continuous process.
The change to a continuous Delphi solved the vexing problem in Delphi research of convergence, correctness of the responses, and termination of the rounds. The literature review in this document highlights the issues that so many believed a hallmark of classic Delphi – that the convergence of the group on some value was tantamount to the correctness of the response. Elimination of rounds also eliminated the need for convergence thinking and convergence’s implication of correctness. The Policy Delphi evolves into a continuous political discussion of a topic of fundamental concern to the participants.

Termination no longer depends upon either rounds or convergence. The project continues as long as the community expresses interest in the topic. For this research, the arbitrary termination or snapshot support completing this document. As will be noted later, atrophy and disinterest supplants the problem of convergence use for termination.

3.5.1 Initial Participant Seeding

Participant selection remains a fundamental issue for Delphi. When selection is focused by the facilitator on a small group of experts having the necessary expertise, the concerns of selection bias and generalization threats are less assailable. Process credibility extends from the personal credibility of the facilitator. As the literature review indicated, selection bias seldom appears.

There are many challenges in the literature to this assertion concerning selection. Anyone trained in research is keenly aware of the issues and limitations of generalization to a population absent the use of random sampling.

As the panel moves away from “experts” toward “stakeholders” then the panel selection becomes a larger issue. The initial target population becomes more amorphous and generally much larger. The problem space is more particularistic in the expert context while more general and politically contentious with stakeholders in a policy context. The challenges to generalization stem from the convenience nature of the panelist selection mechanism.

This research struggled with the issue of panel stakeholder selection and will continue to struggle with it. At least one dissertation committee member raised the issue of generalization of the results due to selection bias and underlying issues of non-random samples. Delphi is not a
survey technique nor does it claim a random sample from a population. As such, the Delphi is neither survey nor experimental research. Delphi’s appearance as a survey compounds the problem. Further, so many researchers report Delphi as if it is survey research or psychometric research. Compounding this is the constant argument of richness versus reach when caught in the debates of qualitative versus experimental research.

Delphi has a specific problem. Empirical research must present evidence but Delphi concentrates on the future. The future has not yet occurred so evidence does not exist.

3.6 Final Software Deployment

In the weeks just prior to data collection, more rounds of participant review occurred. The participants in the review were not software experts but rather people who might actually participate in the system’s use. The objective was to fine-tune the remaining language, screen content, color, and sequencing. In each of these categories, there were numerous additional changes. At this point, the changes were largely cosmetic requiring little computer code alteration. For example, background color changed to white replacing the light yellow used throughout all the development. The final decision threw out all video presentations and video explanations of the system. All tutorials and explanations appear in text or static screenshots with text. The textual material was shortened and made specific. Parsimonious text won favor.

The group suggested final clarifying changes. Each changed seemed to gel around the theme of – make it simple, make it clear, and eliminate anything not essential. These final changes appear below:

1. The login screen omits passwords when prompting for a user name. This is a risk but now invitations to participate do not require that a participant be preregistered by either other participants or the investigator. Code was included to capture IP addresses to remove spam entries. This change also eliminated the need for password recovery and the sequence of screens required to get the password from the user.

2. Following human subjects approval, the system asks for the demographic data. The users were somewhat annoyed with having to answer every question even though they agreed that the researcher would probably want complete responses. Due to the way later processes evolved – in particular the summary charts of responses in the item ranking – all the mandatory response demands disappeared.

Only the first question in the demographic instrument remains as a required response. The first question in the demographic survey identifies the participant in
categories like faculty, student, external, administration, or support personnel. This question’s requirement stems from the dependency in the ranking display.

3. The original sequence then proceeded to a web page that asked the participant to contribute items to the discussion. After multiple re-writes of the directions and further feedback, this requirement was simply out of place. Testers were unable to visualize adding items before they actually saw the items and had done the ranking. This page moved to the logout sequence. When a participant indicates they are finished with the ranking, the system then asks them for any items they think should have been included.

4. Some options were removed on the main page. Some wording changed. The detailed instructions deletion allowed replacement by two lines. Feedback indicated that the activity to “click here” invited obvious participant action. A hypertext link is appears to a deeper explanation in case more explanation is needed by a given participant.

3.7 Final Panelist Identification and Prompting

A Delphi panel has traditionally been one of experts identified by the panel coordinator. The literature on identification and selection is sparse. The credibility of the panel appears to rest on the credibility of the Delphi coordinator. The change from expert participant to stakeholder in the Policy Delphi changes dramatically the target population. Instead of a limited number of experts the target population could be thousands, tens of thousands, or larger.

Delphi panel selection is not intended to substitute for the regime of drawing a random sample from a population. Rather, Delphi – even as a Policy Delphi – remains largely a convenience sample. To that end, two attacks appear for eliciting potential panelists. First, the participant list seeds with as many organizations, mailing lists, and groups as possible. Second, the proposal of Wedmeyer as an example of Respondent or Snowball sampling applies to “swarm” a larger number of people. Like hidden populations of drug addicts, the assumption is that panelists will recommend others to join the panel.

Here is the basic plan of attack to seed the list.

1. Find email of all faculty who have taught a distance education course in last 4 years.
2. Obtain email list from ITS of all people who attended seminars
3. Obtain email of all email of faculty with active/deployed Laulima sites
4. Obtain email of all IT staff related to distance education
5. Obtain email of all education tech personnel – every campus has a department or division that supports educational technology use
6. Obtain email of all college of education faculty related to distance education
7. Contact student governance bodies and their corresponding “education” committee members
8. Randomly select some faculty or departments who might have a “negative” opinion
9. Contact the Department of Education and invite their participation
10. Scan various university directories
11. Use the master class list to identify courses and thus faculty involved in distance education
12. Thought leaders in the faculty will probably be involved in senates and other organizations. Get their list of members.

Because panelists can join over time, it is not necessary to have all invitations and participants at launch. Instead, the solicitation can grow with time. Panelist addition occurs as each decides to join. This change will allow “continuous recruitment” instead of “big-bang recruitment.” The basic idea is to do continuous recruitment as opportunistic groups appear.
Chapter 4. Research Findings

The research findings are reported in the following two sections. In the first section, the data collected is consistent with the original subject of investigation for this dissertation as outlined in the research objectives. The underlying motivation of this dissertation is a future-state codification of policy recommendations for distance education within the University of Hawai‘i system. The first section is subdivided into three subsections (1) expertise (aka, demographic) survey data findings and analysis, (2) item ranking data findings and analysis, and (3) policy recommendations. Rather than including policy recommendations within this findings section, the more appropriate location for the policy recommendations is the Conclusions of this dissertation.

The second section of these findings result from observations gained by a review of the software tool used to collect data. The tool and its implications for the conduct of Delphi Policy research emerged as increasingly intertwined with the eventual results. Although a question was raised concerning the construction of a working web-based Policy Delphi, the importance of the issues surrounding the tool, the tool’s dependency on functional design driven from the literature, and the tool’s actual use increased to a point demanding consideration.

4.1 Expertise Survey Results

A twenty-five item expertise assessment precedes the participant’s entry into the usual Delphi ranking process. The expertise survey is presented to participants to ascertain their skills, knowledge, and attitude towards distance education. As explained in the methodology section, a mind map was constructed to provide an ontological framework from articles that reported faculty, student, and administrator concerns and speculations on the subject of distance education. Some survey items sought attitudes toward distance education including support and advocacy. Some survey items sought expertise including how many times the participant taught or took distance education courses. Some survey items were simple demographics including academic rank, college, and highest degree earned. Some survey items sought beliefs concerning effectiveness and quality of distance education in general or in comparison to tradition, face-to-face classroom instruction.
The original intent of including the expertise survey satisfied the literature review derived requirement to establish participant subject matter expertise. Absent a clear method for participants to evaluate the credibility, stature, and importance of other participant comments, each participant forms their own belief structure about other respondents. In extended interaction where participants directly identify by name, indirectly identified by pseudonym, or just by frequency of apparently similar commentary style and focus, participants construct an image of the unknown originator of comments. Rather than allow this construction to be uncontrolled, this dissertation argued that there is an increase in relative value to participants by establishing the commenting participant’s expertise explicitly. The literature review requirement was in turn derived from the relevant literature previously cited.

4.1.1 Expertise Data Detail

Table 3 reports the distribution by employment or association with the University. Due to small returns from APT (support staff) and administrators and given the fact that these categories are University employees, these groups have been consolidated into the Faculty category. Thus, Faculty is now defined to be University employee in faculty, APT, or administrative role.

<table>
<thead>
<tr>
<th>Question #1</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is your primary affiliation with the University of Hawaii?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty, Researcher, or Education Specialist (Bargaining unit 7)</td>
<td>42</td>
<td>28.0%</td>
</tr>
<tr>
<td>APT Staff (Bargaining Unit 8)</td>
<td>4</td>
<td>2.7%</td>
</tr>
<tr>
<td>Student (any student - community college, undergraduate, or graduate studies and alumni)</td>
<td>99</td>
<td>66.0%</td>
</tr>
<tr>
<td>Administrator (non-teaching and above department level holding E position type and not in bargaining unit 7 or 8.)</td>
<td>5</td>
<td>3.3%</td>
</tr>
<tr>
<td>Non-University</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 4 summaries the distribution of the participants by highest achieved academic rank for University employees and for students their academic objective. The proportion of students seeking doctoral level degrees is elevated in this survey due to the inclusion of the
Communication and Information Sciences seminar. The bachelor’s degree distribution is also disproportionately elevated due to availability of student email lists for contact.

Table 4 Demographic Question 2

<table>
<thead>
<tr>
<th>Question #2</th>
<th>Faculty</th>
<th></th>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the highest level of education you have completed or the level in</td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>which you are currently enrolled with the expectation that you will</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>complete the specified program?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PhD, JD, or any degree that is the highest in your field available in the</td>
<td>37</td>
<td>72.5%</td>
<td>16</td>
<td>16.2%</td>
</tr>
<tr>
<td>UH system</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masters</td>
<td>11</td>
<td>21.6%</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>Bachelors (4 year degree program)</td>
<td>1</td>
<td>2.0%</td>
<td>90</td>
<td>60.6%</td>
</tr>
<tr>
<td>Associate (2 year degree program)</td>
<td>0</td>
<td>0.0%</td>
<td>12</td>
<td>12.1%</td>
</tr>
<tr>
<td>None of the above.</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>2</td>
<td>3.9%</td>
<td>2</td>
<td>2.0%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100.0%</td>
<td>99</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 5 summarizes the institutional affiliation of participants. The participants originated largely with the two O‘ahu four year campuses and the community colleges throughout the State though largely concentrated on O‘ahu. Student participants were from the Manoa campus or West O‘ahu as limited by access to email and participation by faculty.

Table 5 Demographic Question 3

<table>
<thead>
<tr>
<th>Question #3</th>
<th>Faculty</th>
<th></th>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Within the University of Hawaii, what is your campus?</td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>UH Manoa</td>
<td>27</td>
<td>52.9%</td>
<td>27</td>
<td>27.3%</td>
</tr>
<tr>
<td>UH Hilo or UH West Oahu (4-year university)</td>
<td>10</td>
<td>19.6%</td>
<td>69</td>
<td>69.7%</td>
</tr>
<tr>
<td>Community College (Hawaii, Maui, Windward, Leeward, Honolulu, Kapiolani,</td>
<td>12</td>
<td>23.5%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Kauai)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System (not associated directly with any campus)</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Non-University</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>2</td>
<td>3.9%</td>
<td>3</td>
<td>3.0%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100.0%</td>
<td>99</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 6 illustrates the distribution of participants by major. The alignment of faculty and student participation is consistent with the distribution lists used to request participation in this study.
In Table 7, although the solicitation for participation was primarily drawn from faculty listed as teaching an on-line section in the schedule of courses for Spring and Fall 2010 and from students enrolled in courses that were designated on-line during a similar time period, the results show that approximately 25% of the respondents had neither taught nor taken a distance education course. Also included in the selection were members of distance education committees. Remembering that the faculty participants were drawn from those actually teaching via distance education, this distribution’s reflection of the population is skewed to more experience than might be expected should this sample been more representative of the general population of faculty as the actual offering of section of distance education is a small fraction of the total University effort. This skewing is also likely in the student data because of the participation of West O'ahu and community colleges. These institutions are far more likely to offer and use distance education.

Table 6 Demographic Question 4

<table>
<thead>
<tr>
<th>Question #4</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>Social Sciences (like Political Science, Psychology, Sociology, Anthropology, Economics, Communications)</td>
<td>17</td>
<td>33.3%</td>
</tr>
<tr>
<td>Natural Sciences (like Biological sciences, Chemistry, Math, Physics)</td>
<td>3</td>
<td>5.9%</td>
</tr>
<tr>
<td>Arts &amp; Humanities (like Art, Music, Dance, Literature, Language, Culture, Philosophies)</td>
<td>8</td>
<td>15.7%</td>
</tr>
<tr>
<td>Professional (like Law, Business, Architecture, Engineering, Medicine, Education, Nursing, Tropical Agriculture)</td>
<td>20</td>
<td>39.2%</td>
</tr>
<tr>
<td>Applied Sciences (like Food Service, Hospitality, Auto Mechanics)</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>None of the above</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>2</td>
<td>3.9%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 7 Demographic Question 5

<table>
<thead>
<tr>
<th>Question #5</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>None</td>
<td>13</td>
<td>25.5%</td>
</tr>
<tr>
<td>1 to 5</td>
<td>22</td>
<td>43.1%</td>
</tr>
<tr>
<td>6 to 10</td>
<td>8</td>
<td>15.7%</td>
</tr>
<tr>
<td>10 to 15</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>16-20</td>
<td>2</td>
<td>3.9%</td>
</tr>
<tr>
<td>More than 20</td>
<td>2</td>
<td>3.9%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>2</td>
<td>3.9%</td>
</tr>
<tr>
<td>No Response</td>
<td>2</td>
<td>3.9%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 8 indicates that both students and faculty feel comfortable and experienced with distance education. Again, the faculty solicitation pool would favor faculty with a positive view. Fully one-third of the respondents considered their skill to be novice or less.

<table>
<thead>
<tr>
<th>Question #6</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Well Versed</td>
<td>23</td>
<td>39</td>
</tr>
<tr>
<td>Novice</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>No knowledge</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>No Response</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>88</td>
</tr>
</tbody>
</table>

As indicated in Table 9, both students and faculty see growth in the deployment of distance education throughout the university. There is a strong divergence of opinion between faculty and student respondents. The faculty reports a belief that distance education will become a significant percentage of total delivery in the near futures while students see a more reserved growth rate. One-third of students believe that distance education penetration has already reached a plateau.

<table>
<thead>
<tr>
<th>Question #7</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td>Increase significantly becoming a substantial percentage of total delivery</td>
<td>34</td>
<td>66.7%</td>
</tr>
<tr>
<td>Grow but the growth rate will slow from its current percentage of total delivery</td>
<td>10</td>
<td>19.6%</td>
</tr>
<tr>
<td>Plateau at a relatively low level of the total instruction delivery</td>
<td>3</td>
<td>5.9%</td>
</tr>
<tr>
<td>Result in little total change except in specific subject areas and special needs</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>3</td>
<td>5.9%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 10 provides insight into one of the central issues of the literature on distance education. The beliefs about effectiveness of distance education by students and faculty remain a question. Even among this selective sample of those who would tend to be supportive, there is marked ambivalence. As reported in the literature review, the no significant difference finding on multiple measures between distance education and face-to-face education is the paramount finding. The faculty is far more likely to argue for face-to-face instruction as more effective or to see no difference. Students appear to argue that distance education is more effective though a sizeable group appear to support the face-to-face instruction as more effective.

<table>
<thead>
<tr>
<th>Question #8</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete the following statement. From what I know of the research, readings, blogs, or posts on the effectiveness of distance education versus traditional, face-to-face, instructor-lead, classroom, on-campus delivered instruction, I believe that as a means of teaching and learning that</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance education is more effective than traditional classroom delivery.</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>Traditional classroom delivery is more effective than distance education delivery.</td>
<td>21</td>
<td>38</td>
</tr>
<tr>
<td>There is little or no difference in student outcomes.</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>No Response</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 10 Demographic Question 8
Again, due to the nature of the participants, Table 11 shows a reasonably strong advocacy for increased use of distance education. Faculty is either cautious or enthusiastic about expansion while students are cautious.

![Table 11 Demographic Question 9](image)

Table 12 summarizes another interesting divergence. Students do not seem convinced that distance education originates due their needs but instead originate largely from faculty and administration advocacy. Faculties are rather evenly split as to the source – administration, faculty, or student request.

![Table 12 Demographic Question 10](image)

As would be expected, Table 13 highlights that the faculty participants in this study are likely to be advocates or supporters of distance education because they were drawn largely from
faculty involved with or in distance education. Again, students are far more skeptical about the effectiveness and efficiency of distance education.

Table 13 Demographic Question 11

<table>
<thead>
<tr>
<th>Question #11</th>
<th>Faculty</th>
<th>Percent</th>
<th>Student</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td></td>
<td>Frequency</td>
<td></td>
</tr>
<tr>
<td>I am an advocate, a &quot;true believer&quot;, a proponent, a thought leader</td>
<td>23</td>
<td>45.1%</td>
<td>23</td>
<td>23.5%</td>
</tr>
<tr>
<td>I am a &quot;fellow traveler&quot; in that I support the idea, might use it or take instruction using it, but I am not strong advocate.</td>
<td>15</td>
<td>29.4%</td>
<td>10</td>
<td>10.2%</td>
</tr>
<tr>
<td>I am cautious and neutral preferring others to take the lead.</td>
<td>6</td>
<td>11.8%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>I am skeptical as to the effectiveness and efficiency thus slightly opposed</td>
<td>4</td>
<td>7.8%</td>
<td>58</td>
<td>59.2%</td>
</tr>
<tr>
<td>I am opposed to further expansion and use of distance education</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>3</td>
<td>5.9%</td>
<td>7</td>
<td>7.1%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
<td>98</td>
<td>100%</td>
</tr>
</tbody>
</table>

Not surprisingly, Table 14 reports that information technology is either extremely frequently or very frequently involved in their lives.

Table 14 Demographic Question 12

<table>
<thead>
<tr>
<th>Question #12</th>
<th>Faculty</th>
<th>Percent</th>
<th>Student</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td></td>
<td>Frequency</td>
<td></td>
</tr>
<tr>
<td>Extremely frequently</td>
<td>38</td>
<td>74.5%</td>
<td>30</td>
<td>30.3%</td>
</tr>
<tr>
<td>Very frequently</td>
<td>9</td>
<td>17.6%</td>
<td>36</td>
<td>36.4%</td>
</tr>
<tr>
<td>Somewhat frequently</td>
<td>1</td>
<td>2.0%</td>
<td>10</td>
<td>10.1%</td>
</tr>
<tr>
<td>Hardly ever</td>
<td>0</td>
<td>0.0%</td>
<td>16</td>
<td>16.2%</td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>1.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>3</td>
<td>5.9%</td>
<td>6</td>
<td>6.1%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
<td>99</td>
<td>100%</td>
</tr>
</tbody>
</table>
Consistent with the frequency of use of technology, technology is important in the professional life of students and faculty. See Table 15.

**Table 15 Demographic Question 13**

<table>
<thead>
<tr>
<th>Question #13</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>How important is information technology in your professional life?</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>Extremely important</td>
<td>41</td>
<td>80.4%</td>
</tr>
<tr>
<td>Mildly important</td>
<td>6</td>
<td>11.8%</td>
</tr>
<tr>
<td>No difference</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Rather unimportant</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>Extremely unimportant</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>3</td>
<td>5.9%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
</tr>
</tbody>
</table>

Students and faculty overwhelmingly agree that distance education’s advantage either outweighs or is equals any of its disadvantages as reported in Table 16.

**Table 16 Demographic Question 14**

<table>
<thead>
<tr>
<th>Question #14</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which of the following best describes your view?</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>Distance education’s advantages outweigh its disadvantages</td>
<td>27</td>
<td>52.9%</td>
</tr>
<tr>
<td>Distance education’s disadvantages outweigh its advantages</td>
<td>2</td>
<td>3.9%</td>
</tr>
<tr>
<td>Distance education’s advantages are about equal to its disadvantages</td>
<td>18</td>
<td>35.3%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>4</td>
<td>7.8%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 17 reports the perception that distance education is appropriate at most levels of university instruction. The literature and practice within this university system are more likely to report that lower level courses are the primary targets of distance education deployment.

**Table 17 Demographic Question 15**

<table>
<thead>
<tr>
<th>Question #15</th>
<th>Faculty</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which category of student learning should be the primary focus for distance delivered instruction?</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>First two-years or community college or lower undergraduate</td>
<td>6</td>
<td>12.2%</td>
</tr>
<tr>
<td>Advanced undergraduate (3rd &amp; 4th year)</td>
<td>4</td>
<td>8.2%</td>
</tr>
<tr>
<td>Graduate Studies</td>
<td>5</td>
<td>10.2%</td>
</tr>
<tr>
<td>Appropriate at all levels</td>
<td>33</td>
<td>67.3%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>1</td>
<td>2.0%</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 18 demonstrates some interesting perceptions concerning workload. The question should have been unpacked to concentrate more sharply on its two dimensions – faculty and student perception of their personal workload in distance education and faculty and student perception of the other’s workload. In general, a theme appears emergent that each perceives that distance education demands more workload of them. Possibly, there is also the perception that the workload of the other – faculty of student and student of faculty – shows a belief of less effort. In other words, everyone believes distance education cause more work but each believes that work falls upon themselves rather than the other – faculty upon faculty, student upon student but not a belief by faculty of more student workload and conversely.

<table>
<thead>
<tr>
<th>Question #16</th>
<th>Faculty</th>
<th></th>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>When you consider the faculty or staff workload to create and deliver a</td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>distance education course or when, as a student, you consider the student</td>
<td>19</td>
<td>37.3%</td>
<td>66</td>
<td>66.7%</td>
</tr>
<tr>
<td>workload to take, participate, and study in a distance education course, you</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>believe that</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty and student workload is generally greater than traditional classes</td>
<td>8</td>
<td>15.7%</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>Faculty and student workload generally equal that of traditional classes.</td>
<td>0</td>
<td>0.0%</td>
<td>23</td>
<td>23.2%</td>
</tr>
<tr>
<td>Faculty and student workload are less than that of traditional classes.</td>
<td>20</td>
<td>39.2%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Faculty workload is generally greater but student workload is generally</td>
<td>1</td>
<td>2.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>unchanged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty workload is generally unchanged but student workload is generally</td>
<td>3</td>
<td>5.9%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>greater.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty and student workload are both generally the same for distance</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>1.0%</td>
</tr>
<tr>
<td>education and traditional classes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Applicable</td>
<td>51</td>
<td>100%</td>
<td>99</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 19 finds that students and faculty agree that the University does not incentivize faculty to deploy courses using distance education. This finding is interesting in the faculty category because the literature review would suggest most universities provide incentives via training, reduced load, or pay differentials for faculty participating in distance education.

<table>
<thead>
<tr>
<th>Question #17</th>
<th>Faculty</th>
<th>Percent</th>
<th>Student</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete the following statement: The University of Hawaii rewards and incentivizes faculty and students to use distance education above face-to-face, traditional classroom instruction</td>
<td>10</td>
<td>19.6%</td>
<td>45</td>
<td>45.5%</td>
</tr>
<tr>
<td>is neutral in rewarding and incentivizing faculty and students to use distance education</td>
<td>30</td>
<td>58.8%</td>
<td>25</td>
<td>25.3%</td>
</tr>
<tr>
<td>discourages faculty and students from the use distance education</td>
<td>5</td>
<td>9.8%</td>
<td>8</td>
<td>8.1%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>6</td>
<td>11.8%</td>
<td>21</td>
<td>21.2%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
<td>99</td>
<td>100%</td>
</tr>
</tbody>
</table>

Students and faculty in Table 20 indicate satisfaction with infrastructure support for distance education.

<table>
<thead>
<tr>
<th>Question #18</th>
<th>Faculty</th>
<th>Percent</th>
<th>Student</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support for distance education by providing software, training, development time, computers and other required technology, Internet access, television production, and staff support for Distance Education by the University of Hawaii is</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fantastic</td>
<td>6</td>
<td>11.8%</td>
<td>52</td>
<td>61.2%</td>
</tr>
<tr>
<td>OK</td>
<td>26</td>
<td>51.0%</td>
<td>25</td>
<td>29.4%</td>
</tr>
<tr>
<td>Poor</td>
<td>2</td>
<td>3.9%</td>
<td>4</td>
<td>4.7%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>17</td>
<td>33.3%</td>
<td>4</td>
<td>4.7%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100%</td>
<td>85</td>
<td>100%</td>
</tr>
</tbody>
</table>

Students and faculty report that distance education courses is either equal to or exceed the quality of traditional, on-campus courses. Table 21 shows well over 50% support for equality or superiority. Again, the people in this discussion self-select toward belief in distance education. The literature review would suggest far more skepticism by the faculty at large. There was a rather large non-response from students to this question.
Faculty are split on the issue of the match between their area and the use of distance education while students appear more convinced that distance education applies in their field as reported in Table 22. This is consistent with this dissertation's literature review. Nationally, there are variations in the penetration of distance education based on academic subject area.

Table 21 Demographic Question 19

<table>
<thead>
<tr>
<th>Question #19</th>
<th>Faculty</th>
<th></th>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As either a faculty member or a student who has taken distance education classes and compares the distance education experience to the face-to-face, instructor lead traditional classroom instruction, I believe the quality and content of Distance Education courses is excellent</td>
<td>3  5.9%</td>
<td>5  5.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the quality and content of Distance Education courses is above average</td>
<td>16 31.4%</td>
<td>67 69.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the quality and content of Distance Education courses is ok</td>
<td>16 31.4%</td>
<td>1  1.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the quality and content of Distance Education courses is below average</td>
<td>2  3.9%</td>
<td>0  0.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the quality and content of Distance Education courses is poor</td>
<td>1  2.0%</td>
<td>0  0.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Applicable</td>
<td>7  13.7%</td>
<td>3  3.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Response</td>
<td>6 11.8%</td>
<td>21 21.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51  100%</td>
<td>97  100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 22 Demographic Question 20

<table>
<thead>
<tr>
<th>Question #20</th>
<th>Faculty</th>
<th></th>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>You have evaluated Distance Education as it applies to college instruction in general. Now, think specifically to your own discipline as faculty or to your major as a student, to conclude that distance education fits as well as the majority means of delivery of instruction</td>
<td>21 41.2%</td>
<td>24 24.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance education could be used but only sparingly</td>
<td>24 47.1%</td>
<td>57 57.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distance education has no place in this discipline</td>
<td>0  0.0%</td>
<td>0  0.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0  0.0%</td>
<td>7  7.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Response</td>
<td>6 11.8%</td>
<td>11 11%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51  100%</td>
<td>99 100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As shown in Table 23, faculty and students are enthusiastic or at least not opposed to teaching or taking a course delivered by distance education. Although, again, there is sizeable contingent of faculty and students who do not want to teach or take courses by distance education.

<table>
<thead>
<tr>
<th>Question #21</th>
<th>Faculty</th>
<th></th>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Given the option to teach a course by distance education or to take a</td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>course delivered by distance education, I would enthusiastically teach or</td>
<td>20</td>
<td>39.2%</td>
<td>16</td>
<td>16.2%</td>
</tr>
<tr>
<td>take such a course</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would be OK - neither enthusiastic or opposed -- to teach or take such a</td>
<td>15</td>
<td>29.4%</td>
<td>31</td>
<td>31.3%</td>
</tr>
<tr>
<td>course</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I do not want to teach or take courses delivered by distance education</td>
<td>12</td>
<td>23.5%</td>
<td>38</td>
<td>38.4%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>5</td>
<td>5.1%</td>
</tr>
<tr>
<td>No Response</td>
<td>4</td>
<td>7.8%</td>
<td>9</td>
<td>9.1%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100.0%</td>
<td>99</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Convenience dominates the reason for distance education. As Table 24, students are not as likely to see the cost reduction. Students and faculty show are secondary belief that reduction in cost is an alternative reason. Students are unlikely to see an actual reduction in cost as there is no system wide differentiation in tuition. Faculties, consistent with national data, perceive distance education as response to budgetary constraints.

<table>
<thead>
<tr>
<th>Question #22</th>
<th>Faculty</th>
<th></th>
<th>Student</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Which of the following do you believe is the primary reason for distance</td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>education?</td>
<td>32</td>
<td>62.7%</td>
<td>54</td>
<td>54.5%</td>
</tr>
<tr>
<td>Convenience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduction in cost of instruction</td>
<td>11</td>
<td>21.6%</td>
<td>33</td>
<td>33.3%</td>
</tr>
<tr>
<td>More consistent with the &quot;net&quot; generations expectations</td>
<td>2</td>
<td>3.9%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
<td>0.0%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>No Response</td>
<td>6</td>
<td>11.8%</td>
<td>12</td>
<td>12.1%</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>100.0%</td>
<td>99</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
The remaining Table 25 to Table 27 shows that the faculties in this survey have taught a number of distance education courses, that they have gotten training, and that they are cautious about their expertise in using distance education.

**Table 25 Demographic Question 23**

<table>
<thead>
<tr>
<th>Question #23</th>
<th>Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
</tr>
<tr>
<td>None</td>
<td>33</td>
</tr>
<tr>
<td>1 to 5</td>
<td>12</td>
</tr>
<tr>
<td>6 to 10</td>
<td>2</td>
</tr>
<tr>
<td>10 to 15</td>
<td>0</td>
</tr>
<tr>
<td>16-20</td>
<td>0</td>
</tr>
<tr>
<td>More than 20</td>
<td>0</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>0</td>
</tr>
<tr>
<td>No Response</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 26 indicates that most faculty involved in the delivery of distance education has had little formal training in its deployment.

**Table 26 Demographic Question 24**

<table>
<thead>
<tr>
<th>Question #24</th>
<th>Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
</tr>
<tr>
<td>Obtained a degree in a field in which the pedagogical foundations were taught</td>
<td>5</td>
</tr>
<tr>
<td>Participated one or more seminars or workshops for content delivery via distance education</td>
<td>8</td>
</tr>
<tr>
<td>No substantial, formal exposure</td>
<td>22</td>
</tr>
<tr>
<td>Not Applicable</td>
<td>1</td>
</tr>
<tr>
<td>No Response</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
</tr>
</tbody>
</table>
Table 27 reports that few faculty consider themselves experts in the use of distance education as an instructional delivery tool. The vast majority of faculty appears to consider themselves knowledgeable or somewhat knowledgeable.

<table>
<thead>
<tr>
<th>Table 27 Demographic Question 25</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question #25</strong></td>
</tr>
<tr>
<td>Thinking about your personal skills in creating content and delivering a course by distance education</td>
</tr>
<tr>
<td>I consider myself an expert</td>
</tr>
<tr>
<td>I consider myself knowledgeable but not an expert</td>
</tr>
<tr>
<td>I consider myself somewhat versed</td>
</tr>
<tr>
<td>I consider myself a novice</td>
</tr>
<tr>
<td>I have no applicable knowledge or experience</td>
</tr>
<tr>
<td>Not Applicable</td>
</tr>
<tr>
<td>No Response</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

4.1.2 **Expertise Survey Recapitulation**

The tables Table 3 to Table 27 above provide detail from which to construct a sketch of the social environment and expectations of ICT and education interaction. In order to reasonably tie this summary to the underlying empirical data, the question number or numbers will appear after summary assertions. The form will be the letter “Q” followed by reference to the expertise survey instrument.

Faculty and students frequently use and depend upon information technology and they believe it important to their professional lives. (Q12, Q13) Faculty and students have either taught or taken none or 1 to 5 distance education courses. (Q05) This result is somewhat unexpected given that the initial seeded and targeted populations were faculty and students participating in or likely to participate in distance education. Faculty and students do not consider themselves experts in distance education’s use but they consider themselves distributed over the novice to well versed experience level. (Q06) Faculty is more likely to support or at least not oppose expanded use of distance education whereas students are more likely to be cautious about expansion or have no opinion. (Q09) This finding parallels the likelihood that faculty are enthusiastic advocates or fellow travellers in their overall evaluation of distance education but students are more strongly opposed to further expansion. (Q11)
Although students believe that distance education will be slowly increased in deployment from its current level, faculty tends to believe that distance education will be increasing significantly. The students are more likely to believe that the growth rate of distance education has already plateaued. (Q07) Faculty are approximately evenly split as to whether distance education is appropriate in their field and again student caution is reflected in their belief that distance education may be used in a field but only sparingly. (Q20) Faculty and students concur that in large part distance education is appropriate at all class levels (first two years, advanced under graduate, and graduate). (Q15).

Faculty is about evenly split as to the motivational source of increased use of distance education coming from themselves, from administration, or from students. Students do not view the requests for distance education as motivated by their needs but by the needs of faculty and administration. (Q10) Faculty and students largely agree that the primary motivation for increased deployment of distance education is student convenience with institutional cost savings as a secondary motivation. (Q22) When asked about their desire to teach in a distance education environment, the faculty was enthusiastic to approving of doing so yet one quarter of the faculty indicated opposition. Interestingly students show no great motivation to take courses via distance education with approximately one-half indicate they would be approve of or enthusiastically take a course via distance education but a strong 40% are opposed to taking a course via distance education. (Q21)

Society expects the University experience to be of high quality. Students tend toward agreeing that distance education is more effective than classroom instruction but a substantial minority view the opposite. Faculty is split in their beliefs between a belief that traditional instruction is superior or that there is little or no difference. The question was intentionally worded to ascertain perceptions of quality not actual measured quality. The difference between the student perception of distance educations equal or superior effectiveness to traditional classroom instruction as compared to faculty’s equal or less than effectiveness requires further investigation. (Q08) When asked to compare quality, strongly believe that distance education course quality is
above average in comparison to traditional classroom instruction whereas faculty split between above average and essentially equal. (Q19)

When all the advantages and disadvantages are considered in total, students overwhelmingly indicate that the positives of distance education outweigh its negatives. Faculty concur strongly positive or state that the advantages and disadvantages are about equal. (Q14)

The successful deployment and increased use of distance education depends largely upon the faculty regardless of other motivations. Both faculty and students report that the workload in distance education is generally greater. Faculty believes that in the workload difference it is the faculty who must put more work into a course than students must. (Q16) Faculty do not see increased rewards for taking on distance education instruction though approximately one-fourth of the respondents believe the University does incentivize distance education deployment. Students believe there are incentives being offered by the University for them to select distance education over traditional classroom instruction. (Q17).

Students find institutional support for distance education superior but faculty find it only average. (Q18) Faculty in this survey are more likely to have not taught distance education which is surprising again given the target audience. (Q23) Faculty report that for professional development and formal preparation in distance education that generally they have had little exposure. (Q24) Faculty also consider themselves knowledgeable or somewhat versed in the skills to create and deliver distance education. (Q25)

None of the above data are challengingly different from the review of the literature on distance education. The acceptance of the increased and important role of technology is strong. Faculty and students see increased ICT growth in their profession and the likelihood that distance education will be increasing. Yet, both view this change with caution and express concerns about quality. Increased workload in distance education is a concern of faculty and students but both are unlikely to oppose and more likely to support the continued deployment of distance education.

4.2 Delphi Item Analysis

The central focus of a Delphi investigation is the round-based ranking of an item list. The list is traditionally constructed by the investigator. The construction procedures are highly varied
with no standardization. The ranking process almost always accommodates commentary by the participants about an item’s content or the item’s rank from a previous round. Item ranking done by the participants is normally a simple process of evaluating an item based upon whatever criteria the participant’s expertise is appropriate. The participant then reports back to the coordinator the item number and its rank in highest to lowest based upon the professional judgment as just discussed. The investigator then consolidates the ranks to determine an overall rank for each item to be reported back. There is no standard for the metric used but generally it is a average of ranks by item.

The Policy Delphi literature as a special form of the Delphi technique, in specific the seminal paper article by Turoff (Turoff, 2002), suggests four attributes for each item. Three were used in this investigation as explained in the methodology section. In place of the pure ranking from highest to lowest, each item is evaluated on three attribute using a Likert style scale.

Figure 38 illustrates the attributes and their values as they were presented in the literature review. This figure replicates the display presented to the participants for each item in the item list. Participants responded with a choice for each attribute. Thus, each item in the item list receives three responses – one response for each of the three attributes.

![Figure 38 Example Attribute and Scale](image)

Ranking, as is common practice in Delphi, is not present and the participants make no trade offs that are inherent in a highest to lowest procedure. The responses shown in Figure 38 were encoded using the four levels for each attribute are very X, X, un-X, and very un-X. A numeric value 4 encodes Very X, 3 encodes X, 2 encodes un-X, and 1 encodes Very Un-X.
To consolidate three attribute scores, a frequency tally for each value of each attribute for each item was constructed. Thus, for any given item there were three sets of four frequency counts with the cells labeled 4, 3, 2, and 1 corresponding to the VeryX to Very Un-X as described as shown in Table 28. A weighted average for each each attribute was then constructed as the sum of the products of the scores 4, 3, 2, and 1 by the frequency in a cell divided by the cumulative frequency count. Thus, each attributes ranking reduces to a weighted average from 1 to 4 and each item had three such scores. The final item score is the simple sum of these three scores. Thus, the final index score used for each item became the sum of the three weighted averages. Three index scores were created for each item: (1) consolidated index score, (2) faculty index score, and (3) student index score. Table 28 illustrates the index score calculation.

Each of the three attributes has a frequency count for values 4 to 1. The "sum*" column is the sum product. The Weighted column is the sum product divided by the count. The index score is the sum of the three weighted scores. Three sets of these scores were calculated – consolidated index, faculty index, and student index.

### Table 28 Illustrative Index Score Calculation

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Desirability</th>
<th>Sum*</th>
<th>Weighted</th>
<th>Feasibility</th>
<th>Sum*</th>
<th>Weighted</th>
<th>Importance</th>
<th>Sum*</th>
<th>Weighted</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>64</td>
<td>2.7</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The reporting of the results will be presented subsequently in two sections: (1) the Consolidated Item Rank and (2) the Faculty and Student Item Rank Differences. The former reports the items ranked from highest to lowest on the consolidated index score while the later reports two items rankings one rank ordered list for faculty and one rank ordered list for students. The faculty and student index score are used to create the independent item ranking tables.

4.2.1 Faculty and Student Item Rank Consolidated

Given that the consolidated item ranking includes all items with both faculty and student scores undistinguished, the item ranking is reported as Appendix 3 to make the argumentation and discussion of those results the focus and not the extreme detail of the rank order item list. This
section presents the items thematically as they relate by relative position in the consolidated rank table.

Items 446 and 448 ranked highest. These items both place emphasis on the importance of a support infrastructure. Item 446 calls for support of instructional delivery while 448 emphasize the need for a student-learning infrastructure including help desks and potentially tutorial support.

Item 444 remains a theme in this dissertation’s literature review and is again expressed in the high ranking. Item 444, simply put, raises the question of the quality of the distance education delivery and experience. The question was worded intentionally provocatively to stir discussion as recommended by Turoff. The provocative wording and the selection of the question remains a central issue for distance education. In contrast, the opportunity was given to the participants to further downgrade distance education in Item 443 that stated that distance education is a fad. This item received the lowest ranking among all the items. Thus, faculty and student concerns about quality exist but each seems to take seriously distance education as a sustained instructional delivery platform.

A number of items centered on cost to the University, cost to the student, and the potential for additional revenue from student enrollment in distance education delivered courses. Item 427 reflects a changing world of publication in that commercial textbooks should be replaced with electronic and free equivalent sources. Consortia of colleges and many websites like FlatworldKnowledge.com produce free for use under open license quality textbooks. Item 431 ranked fifth overall argues that all the fees for services charged to students who actually reside on campus should not be charged to distance education students if they do not have a physical presence on a campus. Item 430, again an intentionally provocative statement, argues for tuition of one-third the usual tuition to reflect repeated lectures, low infrastructure needs, and the like. Item 434 proposes the elimination of out-of-state tuition for distance education. The proposal ties with Item 437 that argued for aggressive outreach of the instructional mission to audiences beyond the traditional on campus student. Item 432 ranked moderately high and states the students in distance education should not be charged a surcharge for ICT support as those
resources are also used by non-distance education students. Though the there is a constant theme by the respondents that distance education should reduce financial demands on students and reduce demands on physical infrastructure, Item 433 that proposes a slowdown or halt of physical facilities ranked very low. Again, the balance between themes of distance education should reduce financial burden on tuition and infrastructure and the need to have some physical presence.

Item 421 states an expectation that faculty and students will all have their own devices reducing demands on University ICT public infrastructure. Further, Item 422 stated an expectation that both faculty and students would have an enumerated set of common skills and knowledge. These items reflect against the expertise survey that found ICT frequently used and critical to both students and faculty.

This theme has an organizational change focus. Item 436 states that there should be an organizational structure that is separate for distance education. Closely ranked to this item is Item 424 that presented the option for an entirely new structure external to and separate from the University to delivery distance education. Item 429 would eliminate summer school. Item 440 would eliminate maximum enrollments. Item 428 would eliminate the fixed semester concept. Item 445 would add experience and expertise in distance education delivery to the tenure and promotion process. Item 447 focused on the need for distance education to originate with faculty and instructional demands not on external factors such as budget efficiencies. Item 439 stated a claim that students in large courses are the best target audience for distance education. Each of these items fell relative close to the middle of all items and close together.

The last them is characterized as the learning environment. Item 438 gave the opportunity to express a rating on the learner sophistication and the use of distance education. The item argued for an elimination of distance education in the first year or for those whose academic skills may be considered challenged. Again, this was an intentionally provocative item aimed as engendering feedback because this issue appeared in the literature review. Item 442 made the same exclusionary argument for graduate courses. Item 438 appeared in the low middle ranking and 442 next to last.
As noted, a number of items were intentionally seeded to be provocative based upon the literature review that found a number of contentious issues in distance education. The operating assumption of placing these items in the list was to stir the pot and to cause increased interaction and feedback. Unfortunately, participants simply ranked the items and did not provide the hoped for increases in feedback.

4.2.2 Faculty and Student Item Rank Differences

Sufficient responses are available to proceed with an analysis of the differences in item ranking between faculty and students. A Wilcoxon Signed Rank Test will be applied to the item rank order of faculty and students. The null hypothesis assumption is that of no significant difference. Should the test refute the null hypothesis, then analysis will proceed. The difference analysis will be qualitative not statistical with care not to claim differences that would require further testing. This stance is consistent with the acceptance in this dissertation of care not to make claims of psychometric instrumentation as critiqued by Sackman in the literature review.

4.2.2.1 Wilcoxon Signed Rank Test for Ranking Difference

To determine whether the rankings of students versus faculty as they appear in Appendix 3 differ, a Wilcoxon Signed Rank non-parametric test was applied. This test may be used on two related samples to determine significance in rankings. The null hypothesis assumption is one of no difference in rankings of the master item list between students and faculty. The alternative hypothesis of asserts a difference exists. In summary,

\[ H_0: \text{There is no difference in faculty and student ranking of items} \]
\[ H_A: \text{Faculty and students differ in their ranking of items.} \]

The results appear in Table 29 and indicate a significant difference at \( p = .02 \) level. Faculty and students have statistically significant differences in their rankings of the Delphi items.
The results indicate that the faculty and students differ but the found difference does not explain the source of these differences nor does it indicate the strength of differences on individual items.

4.2.2.2 Faculty and Student Item Difference Interpretation

With a statistically significant difference established between the item rankings of faculty and students, the question arises as to the source of the differences. Given the literature review that included the Sackman cautions concerning the use of the item list as a psychometric instrument (Sackman, 1974, p 70), further statistical testing for differences is abandoned to be consistent with the design of the item list and constraints within this dissertation’s methodology section. To deal with the question of differences Table 31 appears below. Table 31 extracts those items most strongly favored by students. Table 32 lists items most strongly favored by faculty. Table 33 lists items for which the rankings of faculty and students tend to agree by showing little difference.

Before proceeding to the findings, there are general methodological commonalities in the interpretation requiring comment.

First, the index used to separate the three lists is the faculty index less the student index by item. The process of index construction was explained in the text surrounding Table 28 Illustrative Index Score Calculation. The index score that results is negative when a student index score on an item is less than that of the faculty index score on the same item. The index difference is then used to sort and separate the list into the three sections listed at the start of this
sub-section. Table 30 illustrates the component scores and the final column that is the faculty less student score. These calculations appear for each item. In this case, the faculty index score of 8.9 has the student index score of 8.8 subtracted to give a difference of 0.1.

<table>
<thead>
<tr>
<th>Item</th>
<th>Desir</th>
<th>Feasibl</th>
<th>Important</th>
<th>Stu</th>
<th>Desire</th>
<th>Feasible</th>
<th>Important</th>
<th>Fac</th>
<th>Desir</th>
<th>Feasible</th>
<th>Importa nt</th>
<th>All</th>
<th>Fac</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distancia education instruction must be designed so that it does not require any more examiner sophistication and self</td>
<td>3.1</td>
<td>2.7</td>
<td>3.0</td>
<td>3.8</td>
<td>3.2</td>
<td>2.6</td>
<td>3.1</td>
<td>8.9</td>
<td>3.2</td>
<td>2.7</td>
<td>3.1</td>
<td>9.0</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

Second, the difference of the index values obscures where an item ranked in priority within a group. For example, in Table 30 illustrated above, the difference of 0.1 gives no indication that within each group the indices of 8.8 and 8.9 are relatively high or important. All that is gained from this subtraction is difference not positional difference. From the tables in which there is disagreement, the stronger positive or stronger negative provide a relative difference of faculty and student rankings.

Third, care is required, as the identification of three categories based on more highly different or similarity is not being done using a statistical test of differences. The Wilcoxon test identifies the difference in rankings as real but the analysis of the source contributions to that established difference is not being statistically established in this review. Rather, a division akin to effect size with a difference generally different by more than ±0.3 is used. Differences less than negative 0.3 are student favored while those greater than positive 0.4 are faculty favored. Differences within the range -0.3 to 0.4 groups into those with reasonable agreement in rank.

4.2.2.3 Student Preferred Items

Table 31 indicates the items where students ranking of each item differed most strongly from those of faculty. There are some themes evident in the differences that are established by grouping similarly focused items.
Students see cost of education as an issue. Items 430, 431, 433, and 434 all relate to a theme that distance education courses should not be more expensive but more likely should be less expensive due to decreased demand on resources. Students also argue for use of text material that is online and potentially free in place of traditional textbooks as in item 427. Item 423 argues that academic programs must be available equally to distance education and on-campus students. Students also question enrollment capacity maximums on distance education courses as very different from faculty in item 440.

Students appear to be arguing for an expansion of distance education in items 424 and 436 whose thrust is the creation of alternative organizational structures if the current on campus
driven University is unable to respond. As the literature indicated, the faculty are usually concerned about the quality of the educational experience in distance education and more so when the offerings and access are enlarged. Yet, according to item 444, it is the students who question the quality of distance education. Students also argue for inclusion of skill in the presentation of instruction using distance education in item 445.

The themes from students thus include a belief that the distance education should not result in higher costs through assessments or other means but instead should be lower in cost. They appear to be arguing for expanded offerings in distance education to the point where the on campus experience and distance education experience are equalized even if that means creating a new organizational structure. Quality is an issue that is not well explored in the preset items presented and students appear concerned.

4.2.2.4 Faculty Preferred Items

Table 32 contains those items that faculty ranked substantially higher than did the students.

<table>
<thead>
<tr>
<th>ID</th>
<th>Item</th>
<th>Fac-Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>428</td>
<td>The traditional semester schedule will be gone or significantly reduced in importance. Learning schedules should reflect the demand, pull-driven model of the just-in-time education of distance learners.</td>
<td>0.3</td>
</tr>
<tr>
<td>448</td>
<td>Distance learners must be supported by an active infrastructure of help desks capable of solving computer and instructional material needs of learners including on-line tutors in most subject areas. Students will be able to call a help desk 24/7.</td>
<td>0.4</td>
</tr>
<tr>
<td>438</td>
<td>Distance education is inappropriate for non-sophisticated learners. No student should be allowed into distance education during their freshmen year. Further, students who might be characterized as academic risks should not be allowed to take distance education courses.</td>
<td>0.6</td>
</tr>
<tr>
<td>429</td>
<td>Summer school is eliminated. The summer program is merged into the regular schedule which itself has changed to a flexible, demand drive model.</td>
<td>0.8</td>
</tr>
<tr>
<td>447</td>
<td>Faculty must own the advocacy of distance education not the administration or support personnel. Faculty leadership and committees become the central driving force for distance education.</td>
<td>0.9</td>
</tr>
<tr>
<td>432</td>
<td>Distance education students should be surcharged to offset demands placed on the communication and information infrastructure.</td>
<td>1.0</td>
</tr>
</tbody>
</table>
There are fewer items in the table and each item is almost a standalone theme. Items 428 and 429 both provide the opportunity to provide a schedule different from the Fall-Spring-Summer schedule. Faculty seem to be arguing for a less rigid approach to study. In finding this theme, the possibility exists, that distance education faculty have structured instruction in such a way that arbitrary start and end dates are less important. Faculty do believe they are responsible for advocacy of distance education as item 447 suggests. Interestingly faculty did not agree with the student theme of cost containment for higher education as discussed in the previous section and as is further evidenced by item 432 – the strongest difference – that faculty believe students should pay a surcharge for the infrastructure to support distance education. Items 448 and 438 each demonstrate the concern of faculty for having the right infrastructure and the right student target audience for distance education.

4.2.2.5 Student and Faculty Concurrence Items

Table 33 lists those items for which the difference between the student and faculty scores show little, if any, difference.
### Table 33 Items with Student-Faculty Agreement

<table>
<thead>
<tr>
<th>ItemID</th>
<th>item</th>
<th>item</th>
<th>Student</th>
<th>Faculty</th>
<th>Faculty - Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>422</td>
<td>2</td>
<td>Every student will be expected to have as set of enumerated computer skills. Faculty and staff will be expected to have at least the same skills as students. The University will not provide training in these skills except through outreach or other remedial delivery.</td>
<td>7.9</td>
<td>7.6</td>
<td>-0.3</td>
</tr>
<tr>
<td>435</td>
<td>2</td>
<td>The University should provide increased tuition support for distance education students to encourage growth in distance delivered instruction.</td>
<td>8.1</td>
<td>7.9</td>
<td>-0.2</td>
</tr>
<tr>
<td>443</td>
<td>2</td>
<td>Distance education is a fad. The University must limit distance delivered education to no more than 10% of its instructional delivery. Distance education destroys the social fabric of the University.</td>
<td>6.2</td>
<td>6.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>421</td>
<td>2</td>
<td>Every student must have their own computer and other technology that meets University minimum requirements. The University will create the minimum requirement list. The University will not provide end user equipment.</td>
<td>9.7</td>
<td>9.6</td>
<td>-0.1</td>
</tr>
<tr>
<td>425</td>
<td>2</td>
<td>Every counseling or support service available to on-campus students must have an equivalent distance education service.</td>
<td>9.2</td>
<td>9.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>446</td>
<td>2</td>
<td>The creation of quality distance education depends upon specialized skills like curriculum specialists, graphics designers, and IT. The University must publish and provide a uniform set of services available to all faculties who convert to distance education or who need to maintain current distance education.</td>
<td>10.0</td>
<td>9.9</td>
<td>-0.1</td>
</tr>
<tr>
<td>437</td>
<td>2</td>
<td>Using aggressive promotion and marketing, the University should dramatically increase student enrollment in its Outreach College beyond the current student audience by using distance education. The objective is to increase revenue to the University.</td>
<td>8.3</td>
<td>8.3</td>
<td>0.0</td>
</tr>
<tr>
<td>441</td>
<td>2</td>
<td>Distance education instruction must be designed so that it does not require any more learner sophistication and self-reliance than would be required in on-campus, face-to-face instruction.</td>
<td>8.8</td>
<td>8.9</td>
<td>0.1</td>
</tr>
<tr>
<td>442</td>
<td>2</td>
<td>Distance education is inappropriate for graduate education. Distance education must be discouraged in graduate education.</td>
<td>6.3</td>
<td>6.4</td>
<td>0.1</td>
</tr>
</tbody>
</table>
As a reminder, care must be taken when reviewing this list as the small difference could be the result to two high or two low scores on the item ranking within the faculty or student group. To assist in the interpretation of rank and rank differences, this table has been enhanced with two additional columns – the faculty and the student rank score. Providing these scores allows mitigation of the difference issue and context for differences and the interpretation of the differences.

Faculty and students agree and show relatively strong support for the expectation that students should have a minimum set of enumerated computer skills that can depended upon throughout the University and that students should be expected to have their own computer equipment thus not relying on the University to provide labs or other open resources. See items #422 and 421. This is consistent with the demographic findings that showed a strong use of technology in everyday academic life and the view that such activity is highly important to completion of tasks.

Items 425 and 446 argue positively on behalf of faculty and students for support. In item 425, the support is general and related to normal student activities while item 446 argues for environment of specialists to assist in the production of quality instructional material for distance education.

Item 441 presents the case that if indeed there is a difference in workload or the expectation of more learner sophistication or self-reliance than might be expected from an on campus, face-to-face class, then such requirement should be inspected and eliminated. In the demographic survey, students and faculty both indicated that distance education is often more demanding. This item reconfirms that position and argues that the difference must be recognized and ameliorated.

One large national theme evident in the literature review is the use of distance education to broaden the audience of the on campus program with the supposed intent of increased revenue. Faculty and students agree that the University should dramatically increase student enrollment through outreach as reported in item 437. Revenue increase is a particular focus of
the item thus eliminating the effect of this item being a request by students to increase access as has been noted elsewhere in this study.

As might be expected of any question of increased financial support of student -- an assertion that is probably a constant across any context -- the faculty and students agree the incentives in terms of scholarships should be provided to distance education students to encourage growth of this instructional approach. This is reported in item 435 and is rather strongly advocated.

Though they agree, faculty and students do not appear to pan distance education. Item 443 argues against the view that distance education destroys the social fabric of the University while item 442 counters those who argue that upper division and graduate division courses are inappropriate for distance education. Though faculty and students agree, these items are of relative unimportance to them.

4.2.2.6 Item Voting Dimension Independence

The Turoff article that outlined the Policy Delphi methodology (Turoff, 2002, p. 86) identifies the four categories upon which this dissertation relies to construct the weighted rating scale used to rank items. Turoff identifies desirability, feasibility, importance, and confidence and notes that for these four voting dimensions they “seem to represent the minimum information that must be obtained if an adequate evaluation is to take place.” (Turoff, 2002, p. 86) He continues, “These types of items will usually induce a good deal of discussion among the respondents and may lead to the generation of new options.” (Turoff, 2002, p. 86) In this article, Turoff also argues, “no neutral answer is allowed other than No Judgment” (Turoff, 2002, p. 87) as was enforced by the web-based Policy Delphi and applied in constructing ranking scales. The use of the four voting dimensions was a supposition on Turoff’s part that these dimensions would generate discussion. There exists an unproven assumption that these dimensions are sufficiently independent in the participant’s evaluation process to provide additional information. If the participant fails to differentiate each dimensions’ focus by responding similarly across all the voting dimensions similarly than little information is gained. As was noted in the literature review,
design review meetings with Turoff resulted in the elimination of confidence as a voting dimension.

During one of the software reviews conducted after the Turoff restructuring of the voting dimensions, the behavior of the review participants indicated a pattern that the rank choice on the first presented voting dimension (desirability) became the default for the remaining two dimensions. Participants appeared to just copy their responses across the dimensions mechanically without regard to the actual intent of the dimension. This pattern appeared to grow stronger as the fatigue of additional items set in. The Ho`ailona software collects no verifiable data on this behavior and modifications to collect this behavior might have further delayed necessary testing on critical path activities to complete the software.

Though this dissertation's design focuses on other issues related to the web-based Policy Delphi, the opportunity to observe item rating differences is minimally available. From the summarized item rank table, the mean score of each item and the raw weighted score are available. An error measure of the absolute value of the difference of each voting dimension from the average score results in three difference scores. If the voting dimensions are indeed independent, the expectation should be a reasonably spread distribution rather than are relatively close distribution.

Figure 39 illustrates the distribution of the ranking differences. The differences appear small and tightly distributed.
The independence of the voting dimensions might be a critical need to establish in a much firmer way. There are at least two implications from this graphical challenge to the independence of the dimensions. First, if the dimensions are not independent, then the voting dimensions introduce unneeded tedium in the ranking process. Second, if the dimensions are not independent, then they might not be serving the purpose identified by Turoff to stir the pot and increase the discussion. Third, and most important for designers, the elimination of three voting dimension to one will greatly simplify the mechanics of the interaction of the participant with the system allowing for a more intuitive interface. For example, if only one voting dimension ranks items then a participant might be presented with a list that they could drag and drop into rank order position. This change would also allow the introduction of a tradeoff mechanism absent in the complexity of three voting dimensions. Such a mechanism might leverage the newly found drag and drop interaction with 100 points distributable over the list to get a sense of relative weight that a position only system does not provide.
This simplification would reduce the load to explain the system. This simplification would reduce the effort and time consumed by the participants.

Again, the lack of difference in response to the scales used in the voting dimensions is not being statistically denied but neither was the original assertion in the literature that the voting dimensions are indeed independent or most central to the point for which they were included—that are to “induce a good deal of discussion among the respondents and may lead to the generation of new options.” (Turoff, 2002, p. 86)

4.3 Web-based Policy Delphi Tool and Intervention

The web-based Policy Delphi tool and all the activities surrounding the use of the tool to accomplish the end goal emerged as an additional but highly important finding.

This research used the approach of driving functional requirements of the software design based upon the Policy Delphi and highly related literature. The literature that specifically spoke to the design of a Policy Delphi application is very thin. Though thin, there exist many articles the surmise or recommend capabilities. These recommendations are highly valuable to a comprehensive implementation. For example, many authors suggested the need to establish “expertise” of the participant as a potential modifier of the weighting that other participants would give to the perceived more expert commentary. Other authors suggested the need for factual or almanac queries to which neutral or fact based responses might be provided as clarification. Still others presumed that sheer interest as stakeholders in a policy would drive participants to recommend others in the network of experts to increase participation. This research would be remiss if it failed to comment on the observations about the utility of these presumed functional necessities.

The analysis that follows, as it relates to the outcomes of success or failure of various functional requirements, could very well be unique to this implementation or to the subject of this investigation. Separating these intertwined relationships and adding in the constraints of a qualitative project compound the need for caution and warn against generalization. Yet, the observations gleaned from this intervention may be highly valuable to those who might next attempt a web-based Policy Delphi investigation.
1.1.1 Results from this Delphi Intervention

This Policy Delphi investigation did not fail to produce evidence. This Policy Delphi did not fail to provide the empirical foundation to create policy statements. This intervention did not fail to produce evidence for use to clarify functionality of a Policy Delphi tool.

Regardless, the results obtained do represent the thoughts of the participants and larger n-size does not assure more richness of communication acts or the externally valid representation of the distance learning’s futures relationship by the item list. Delphi is qualitative research where the n-size of participant pool has weight. However, n-size alone does not dominate analysis. In this research, a web-based tool collects the empirical evidence. The use of a tool does not alter the fact that like other qualitative research this effort is a consolidation of the communication acts of those who agree to participate. The policy statements stemming from this investigation represent just what they are – the summary and extension of the communication of those who participated.

The dissertation’s expectations captured in the assertive title might have been well served to indicate more caution with less generality than is implied in the current wording of the title.

Many researchers, through professional training, focus on the sample selection and sample size. Supporting these expectations related to sampling is Delphi’s appearance to reviewers that Delphi is a survey instrument. Delphi is not survey.

Unlike the closed-item survey instrument, Delphi studies require participant generation of items included in the item list after initial the initial round. Item addition increasingly defines the construct. There is danger in the misinterpretation of reviewers that Delphi interventions are survey research. Far more critical is the same assumption by participants that the Delphi is a survey. The assumption by participants chills the need for item construction and the active creation of new or clarifying items contributed to the item list is one of the major contributing values of Delphi. This participant’s interpretation that Delphi is a survey is extremely difficult to dislodge as indicated in the comment by one participant in this research. The participant wrote,
“... having difficulty with the wording of the survey items.” The survey aura stifles item creation that in turn reduces the efficacy far more than does n-size or sample selection considerations.

1.1.2 Participation Challenge

A substantial risk to a Policy Delphi is highlighted by Turoff as stated, “The first and foremost problem in conducting a Policy Delphi occurs with the initial steps in the process. [As one must identify] the respondents that feel strongly about the issues.” (Turoff, 2002, p. 87) Turoff’s comments direct themselves to the item generation activity and not specifically to the issue of participation. As has been found in this research, getting participation even from those who one might suppose have a stake in a topic important to their future and their organization’s future may not be sufficient.

This project hoped to engage participants in two categories. The categories reflect participation categorizations necessary to ground in empirical evidence a policy that describes the futures interaction of higher education and information and communications technology. The two categories of potential participation are:

1. Agree to participate, complete the initial assessments, and rank and comment upon items in the item list.

   Participation here supplies data describing the current organizational context and at least one round of item rating.

2. Upon the decision to participate in category #1 activities above, participants would add items, comment on other items, contribute other members of the community (Wedemeyer’s Swarm Procedure), and re-rank or comment upon items as more participants added their responses and comments.

   Participation here adds items to the existing list that extend and refine the potential policy statements. Participation here also provides indications of usefulness of features added to the instrument driven by the literature review.

To determine overall successful participation, each category demands a separate review because the sophistication and engagement demands increase as a participant moves from category one to category two participation. The minimum assessment of category 1 is the n-size of those
agreeing to participate and their completion of the expertise summary survey coupled with at least the rating of the current item list. The assessment of category 2 participation considers not only the re-visit to the site but also the addition of new items, the use of comments, and the use of other features in the web-based Delphi extended feature set. Category 2 assumes category 1 participation.

As outlined in the launching of this project, the ideal target populations include all teaching faculty, specialists, students, and other stakeholders but the organizational and political difficulty in reaching the wide-spectrum population on initial solicitation blocked this approach. Access to official University email lists requires approvals of research process and content unacceptable and unobtainable. In general, the University is not encouraging in the use of master email lists for the obvious reason that the number of requests would be large, judgment of what is important research is required, and targeted recipients already received similar email that looks like “spam” solicitations. The limitations compelled a strategy change that recalls the literature findings that Delphi studies usually purposefully select participants with demonstrable “expertise” in the subject of the study. In a Policy Delphi, expertise translates to involved, concerned, and informed stakeholders expected to be sufficiently motivated and knowledgeable to understand the value of participation. Given the access limitations and the original Delphi argument for expert or stakeholder participation, the target population procedures required modification.

The modified targeted populations became:

1. Faculty throughout the University system identified on the Spring and Fall 2010 schedule of courses as teaching distance education courses. Other sources included public lists of faculty serving on faculty senate committees with distance education as part of their charter. Various committee reports and a large report by the community colleges on distance education spanning seven years provided additional potential participants.

2. Student participant selection focused on available distance learning courses in business administration at University of Hawai’i West O’ahu campus.
The sources of participants above remain target populations. Actual participants are those who decide to complete tasks once solicited. One addition of this Delphi study that is uncommon in most Delphi – the expertise questionnaire – provides demographics that facilitate characterization of those who do finally respond. Direct identification using the login to the system is not allowed under the announced human subject’s confidentiality. Participant assurances stated that no email address would be used to identify any part of the research.

Two points standout requiring further discussion.

First, an unknown and to be tested capability of this research’s web-based, literature driven Delphi system concerned the inclusion of the Wedemeyer Swarm (aka, snowball sampling or respondent data sampling). The reference to “literature driven Delphi system” demands attention. Software development usually originates from purposeful response to human or organizational driven requirements. In this research, the software requirements result from the lengthy literature review included herein. Refinements of functionality normally associated with the development feedback loop do not apply as they might in traditional development processes. There is no active, well-defined target audience specifying terminal objectives or performance criteria. The literature is what the literature is. Variation might originate from an interpretation of the literature different from the interpretation presented in chapter 3, but who would write that interpretation and do the associated software development? No claim exists that the interpretation done and the development goals stated by this researcher are those that are universally unique and represent absolute truth. Rather, these are the best for this researcher at this time. Opportunities do exist for functionality adjustments within the broad category of human computer interaction but limited here to actual interface and sequencing questions and not to larger considerations of design goals. Some follow on research might wish to consider more active treatment of broader HCI considerations but the goal of this research sought other ends and is not an HCI methodology investigation. It is impossible, within the scope of this research, to determine how the structure of the process as a web tool and its specific interaction influenced the decision to participate.
Potential participants in the swarm approach to population identification are seeds who should in turn identify other participants in the assumed network of distance learning stakeholders within the organization. In today’s web lexicon, the process approximates “going viral” within a target community of those who might be most engaged in discussing the future of distance learning. Numerous published studies successfully used this sampling approach as discussed in this dissertation’s literature review. This approach to identification makes no claims that it identifies fully the target population nor does this approach claim equivalence to the gold standard of random sampling. The inclusion of the Wedemeyer swarm, as an exemplar of the snowball genre of sample identification, is itself a risk due to its dependence on the assumptions just discussed. However, the inclusion is also an opportunity to gather evidence as to the soundness of this assumption in a Policy Delphi intervention. A literature review found no discussion of the success or failure of the swarm approach’s identification efficacy. Regardless, the approach remains central to this Policy Delphi study to identify a target population beyond that which the researcher seeds.

Second, the selected faculty and student participants identified as seeds in this research actively participate in distance instruction and learning. This seeding should increase the likelihood that the participant has formulated an understanding or opinion about distance learning.

The capability existed in the system to recommend other participants. A total of ten (10) additional participants were recommended by those faculty who agreed to participate in the Policy Delphi. Of those ten recommended, seven (7) of them were already identified as potential targets. Though the swarm capability functioned, the generation of new participants is small.

One issue requiring attention from a future developer relative to the swarm is the placement of the swarm in the sequence of activities comprising the Policy Delphi. The early drafts of the software provided the opportunity to nominate additional participants early in the sequence and prior to actually ranking items. Reviewers argued that not until they had actually seen the items and done a ranking would they be able to determine who might additionally participate. The option to include additional participants was then moved to a post item ranking process. In one test of this placement, reviewers simply closed the browser after completing the
item ranking and thus did not add additional potential participants. The final design includes the capability to nominate within the context of the main item ranking process.

4.3.1 Attrition

As discussed, the identification, solicitation, and actual participation of those believed to be potential stakeholders in distance education within the University system became an unexpectedly difficult challenge.

More than getting agreement to participate, the motivation to participate in subsequent rounds was even more challenged.

The dropout rate as measured by return to the system to review previous postings rendered the multiple rounds moot. Of the faculty, only twelve returned more than once. Of these, two returned because they had not completed the survey and three had had technical issues with their browser requiring them to return. The former were identified by SQL query and the later via email to the Delphi moderator. Of those that did return, few if any ranking changes were made.

As for students, twenty-eight returned but only three made changes. No reports of problems were received from students. Eight of the returnees returned because they did not complete the activity the first time.

The reasons for the attrition include so many possibilities. First, the participant’s actual attachment as stakeholders may have been weaker than believed or the subject just might not be that important. Second, the complexity of the activity dissuaded them from returning. Third, the lack of definitive rounds where they were specifically invited back to review on going ranking changes may have lessened the need of review.

4.3.2 Item List Generation

Delphi is structured communication between stakeholders in a designated topic area. The belief of Policy Delphi investigations is that participants will engage actively in the activities of the technique. Again, Turoff provides observations that assume participants are strongly engaged in the subject of the Policy Delphi,

[If strong engagement] should be the case, they will generate a large amount of written material. If they are provided a certain number of items to deal with on the first round then each of them will make approximately the same number of written comments or additions in response. These must be abstracted carefully
and duplications among the respondents eliminated. On the average, the written material in the questionnaire for the second round will be five to ten times that of the first round. (Turoff, 2002, p. 87)

The items list bounds the Policy Delphi discussion. Two issues arise with the item list. First, should the list be empty when initially circulated to the stakeholders? If the item list is not empty, then how should it be created? Second, once an initial list is constructed, do the participants then add items that they come to realize as important?

As explained in the phases of the development of the web-based Policy Delphi application, the feedback from reviewers of the software was very clear. If there were no initial list of items, there would be no responses. Reviewers were clear stated that they needed an item list for two reasons.

First, they needed the list because they expected a list. No matter how much one makes clear that Delphi is not a survey, there is an expectation that the Delphi will look like a survey. Expectations are strong framing forces.

The only alternative remaining was an investigator-generated list of pre-panel items. To create the list, additional literature searches used the library on-line tools to locate articles on “strengths” or “weaknesses” of distance education. Additional searches sought articles that project the growth or organizational structure of distance education. No single article appeared that accomplished the goal of identifying a list of items useful for this study. A mind map classifies, categorizes, and hierarchically organizes the source material. The process of creating the map took a number of iterations and restructurings. A series of draft items evolved from the mind map. A list of near 60 items appeared in the prototypes distributed for testing and review. Murray Turoff did a high-level review and suggested that 100 to 200 items should be on the list. The items should be very specific. Countering the Turoff view, the reviewers made clear that a long list – and in particular one that required three responses (feasibility, importance, and desirability) – was far too much to expect from their participation. The reviewers suggested no more than 25 items in the initial list.

As proposed toward the start of this subsection’s discussion, the role of the item list after panel initiation is also a subject rich for investigation. As with the issue of the creation of the initial
seeded item list, a discussion of adding items during the execution of the Delphi, is absent in the literature. One might expect this process modification to be common to Delphi because Delphi, as a set of communication acts, has to evoke new understandings from participants. These understandings were not previously contained in the seed item list. Of course, it is possible that the item list was perfect at initiation or those panelists have no new ideas. Yet, there does not appear to be literature on this subject.

The participants did add items to the master item list. The participants submitted thirteen items. Some of the submissions were not items per se. The request to add items appeared in sequence after the participant completed the ranking process but before signoff. The item inclusion web page was a free form textbox. A number of the participants used the textbox to make comments about this implementation, the assumptions buried in the items list, and the relationship of the attributes to the items.

Five of the items as rewritten to match the expectations and general structure of items were included back into the master item list. Those not included either were comments or did not allow for restructuring.

The results of this input indicate the need for a general comment capability. The results also indicate that a “query and clarify” capability must be present to follow up on suggested items. For example, one item suggested for inclusion mentioned “anything on virtual reality.” As it stands, this item is not salvageable but with clarification, there may be value behind the thought.

The complete list of items contributed is contained in Appendix 2.

Thirteen suggestions appeared for new items. Of the thirteen, five were salvageable as items to be added. By the time the solicitations and re-solicitations were sent, the items were not added because they would have gotten little response. A simple matter of timing and the effect of attrition made adding the items to the master item list a null as these items would be unlikely to gain a reasonable number of votes.

4.3.3 Items Morphing into Comments

The design of the web-based tool had not identified the need for a tool-wide comment capability. Feedback in the form of ranking, comments, and research are all item centric. There is
no option to provide feedback outside of the context of an item. A number of the items were not truly items but were instead more general comments about distance education, the Delphi implementation, or other issues. Three exemplars are below:

Some of the items are so inclusive as to be unanswerable with the responses provided. Many of the final ranking statements are far too extreme to warrant input, for example: All publisher texts should be dropped in favor of Open Content? I am a huge Open Content proponent, but I can't support (or even respond to) the statement as it's posed.

Distance learning should only be given to faculty who cares about their students. They tend to spend more time on their students' progress because they care. Those who do not care only do so little because they are only doing it as a standard procedure only that is why they do not go the extra mile to reach out to their students.

I hope you hear from more people like myself, who have been educators for decades, and who have a great deal of experience teaching graduate students. (I've been the graduate chair of my department for more than twelve years.) Only then will the pursuit of graduate studies (and even upper level undergraduate studies) by means of distance-learning begin to seem problematic (if then!) to someone who could have formulated this kind of questionnaire. Lord help us if the assumptions that shaped this questionnaire ever begin to influence the direction of higher education!

A future design must accommodate a communication on the target subject of the discussion of the Delphi. There may in fact be themes that could be pre-structured for this feedback including item structure, Policy Delphi web-tool procedures, or the panel focus, herein Distance Education, as a general topic.

4.3.4 Comments

Delphi assumes that participants utilize feedback to clarify, refine, and change their item rankings. Classic Delphi emphasized the ranking results as the primary source of feedback. Later work indicates that free form comments by the participants are another source of feedback.

Comment input in Ho'a'iona is limited to an item, not to the discussion of Distance Education or the Delphi process in total. Comments originate with a focus on the specific item. The input mechanism for comments appears when an item is exposed.

Participants used the comment capability. Thirty-five comments were received over a number of different items.
All comments appeared to be one-way communication. Participant response to already posted comments appears non-existent. No statement can be made as to whether or not discussion threads might appear if interest increased and attrition decreased.

There is a design issue with comments that is fundamental to the structure of the system. Comments are item context related. Thus, a comment must be focused on the current. This constraint has a limiting effect upon participation and makes display of comments and items complex. The interplay of the ranking categories and their display concurrently with the option to display or enter comments may have had other deleterious effects upon generation. Even though the system had an icon to indicate that something new or changed had taken place to an item since last reviewed, that change might have been one of three rankings or a comment or a reply to a comment. Differentiating these again increases complexity of the interaction.

The comments make some excellent points. These points should be the source of additional items or restructuring of current items. Further, many of these comments should trigger further response. Faculty tended to engage more on issues related to instruction. For example, the following are taken from the comments made by faculty

Again, student is the consumer of the university setting. If they want to bypass the social fabric, it does not mean they cannot learn what is required for a degree. There are ways of socializing via the internet for things like group projects as needed, and presence in a classroom does not mean that a survey can’t be conducted, it just means that a population of college students can’t be required, other populations are available for surveying.

At a university, all faculty are specialists in one thing or another. To suppose that all faculty must be qualified and competent at distance-learning is to grossly overestimate the significance of that form of education.

in my opinion distance education is ideal for classes that typically have large lectures simply because when in this setting its hard to ask questions during lecture because it could set the class back. in both online and traditional setting questions can be asked after the lecture and appointments can be made to meet the professor and further discuss issues....presence in the classroom is not neccessary. i feel video lectures as done in my busa320 class are the ideal lecture tool

Distance education is here to stay. That said, there are specific courses that require at least some Face-2-Face. In my field, it is important that students manipulate laboratory equipment in a group setting and learn how to interact in this environment.
The complete list of comments provided is available in Appendix 1. The comments were for the use of the participants. No analysis appears here.

4.3.5 Requesting Research
There were no requests for research. This feature stemmed from the literature indications that participants of discuss points not founded on a fact-based position. The argument in the literature suggested that participants would make clarifying requests.

4.3.6 Swarm
The participants did suggest others to participate. Participants suggested twenty-two other candidates. These suggestions were compared to those already existing. Ten were already present. Thus, twelve new solicitations were included in a future email. None of the additional members participated.

The swarm email originated from a generic GMail account. It is quite possible that spam scanners removed the solicitation. Interpreting the literature, if the motivation to participate is related to the personal nature of the appeal, then the appeal should originate from an email account of the sender and not from a system account.

4.4 Summary
This chapter addressed the two groups of results and their subsections as described at the start of the chapter. The results focused on the motivation for this study, which was the conduct of a Policy Delphi that would provide results to be used in the construction of a futures policy for distance education. As the research progressed, the Delphi tool and the methodological choices made to implement and utilize the tool became an increasingly important and intertwined activity of the research. The results of observations about the tool and the behaviors surrounding the investigation were reported above.

The responses from the Delphi activity itself were presented in detail. The choice was made to separate the construction of the policy statements from the more empirically grounded results of this section. The policy recommendations appear instead in the conclusions chapter. The results from the Delphi intervention were reported in two sections major sections – the results of the demographic and expertise survey and the item ranking. In the original description of this
research, the demographic results played a minor role in that their purpose was not independent of the ranking process. The expertise survey was intended to inform the decision-making of the participants relative to the expertise, skill, or biases of other participants as they provided feedback on a given item. As the results developed, the expertise survey also informed the decisions of this research and in particular the policy conclusions that are to be presented. Thus, more detail was presented from the expertise survey to paint the picture of the participants.

The second major section dealt with a review of the item rankings. This section was divided into two subsections. The first presented was the thematic gathering of items based upon similarity of content and ranking. This result ranking was based upon the consolidated index score from both faculty and students. The second section began with a Wilcoxon test to determine the ability to refute the null hypothesis of no difference in item ranking. The test returned a result indicating a statistically significant difference between item ranking of faculty and students. With the difference grounded, the analysis proceeded to divide the item list into those most strongly favored by faculty or by students. The third subdivision of the list looked at those items for which faculty and student differences in rank were small. The items were grouped and a narrative constructed.

The second major division of the results dealt with the Policy Delphi tool. Comments concerning the tool appeared in this section and will again appear in the conclusions. The intent of this section was to document those observable activities from the tool. For example, how well utilized were functions like the "swarm", the feedback, the review of expertise, the contribution of additional items, or the almanac research request. The conclusions chapter will provide context and observations beyond the empirical.
Chapter 5. Conclusions, Discussion, and Future Research

This section presents conclusions and discussion focused on the two primary findings of distance education policy and the construction of a web-based Policy Delphi tool. The section concludes with a discussion of contributions and some indicators of potential future research.

5.1 Distance Education and the Futures Instructional Mission

At the inception of this investigation, the primary goal sought to describe the futures relationship between information and communications technology and the instructional delivery mission of the University. What did current stakeholders desire the interaction to be? What should the relationship be?

The construction of the narrative rested upon the hoped for rich communication and participation of stakeholders in a futures’ state discussion of the interaction of education and ICT using a web-based Policy Delphi.

Faculty and students were solicited to participate and a number agreed to participate. As with all Delphi, the solicitation of participants likens to a convenience sample. In traditional Delphi the constructed sample is one of experts identified by the investigator as having requisite knowledge. In a Policy Delphi, experts are instead stakeholders who should have motivation to participate based on the importance of the topic. Stakeholders are expected to have varying degrees of attachment to the proposed discussion topic and even more widely varying expertise.

The structure of the investigation never assumed that faculty participation was a measure of success as the actual participant count depended upon the unknown efficacy of the swarm process and the motivation to participate. The numeric goal served only as a goal to generate a reasonably rich discussion within the context of the Policy Delphi technique.

The narrative derived from the intervention never terminates. The narrative intended to be a launching point from which further investigation, clarification, and review ensues. As was argued in the literature review, the Policy Delphi does not substitute for the actual decision-making process. The relationship of a Policy Delphi to decision-making is like that of evidence at
trial to the jury deliberations. The jury remains the formal mechanism of decision-making but the evidence presented should persuade, guide, and inform the decision-making process.

This work failed to achieve its desired participation goal. The difficulty of securing participation from each of the categories of stakeholders was decidedly more difficult than was anticipated. The failure to achieve a desired number of participants does not negate the fact that the participants who did interact using the Ho’ailona system provided insights. Generalization was never the goal of the intervention, though, like any research, the application of the findings would create value. A larger participation might generate more weight although Delphi methodology largely relies upon convenience sampling negating many concerns about audience size.

5.2 Policy Recommendations

The results of this Delphi Policy investigation are cautiously reported within epistemological and methodological limitations discussed previously. Policy statements are generalizations but Delphi, as an instrument in the qualitative genre, strongly grounds its internal validity but claims only moderate external validity. The challenge originating from cautious, restrained claims to external validity limits the assertiveness and breadth that statements of policy from the Delphi intervention allow.

Two themes require reiteration here. First, the statements of conclusions from the Delphi exercise are policy recommendations of the desired state of higher education and information and communications technology some 15 years into the future. These statements are not statements of what will be but those of desires. Statements of desired outcomes contrast significantly with most Delphi interventions used to estimate future outcomes. Second, as carefully discussed in the literature review there must be a relationship between the empirical data collected by the intervention and the policy derived. The relationship need not be one-to-one as but the inferences drawn and policy posited must have roots in the observed results. Maxwell, as presented in the literature review, explained the issue of constructing conclusions within the qualitative context.

Discussed previously was the push and pull concerning the number and content of items within the Delphi item list and the effect that length, comprehensiveness, and completeness has upon participation. If this research elected a full, rich item list constructed by this researcher
based upon the literature review of the current and futures state of distance education, then a Delphi intervention morphs into a survey. The decision in this research avoided the temptation to create a survey-like instrument. Instead, the Delphi item list presents some seeded items derived by creating a concept map of major themes from the distance learning literature review. The risk in this decision relates to participant behavior and to the richness and breadth of intervention data. If participants continue to perceive the Delphi as a survey because it provides a lengthy initial seeded item list, then participants may decrease their propensity to create and contribute new and differing items to the original list. To the contrary, the shorter and potentially more controversial item list increases the likelihood of contribution due to its perceived incompleteness on the part of stakeholders. The tug-of-war between a comprehensive and a minimalist item list directly influences richness of the participant communication that in turn directly enriches or limits policy construction derived from the communication acts.

5.2.1 Policy 1 – Need For Policy of Commitment

Respondents indicate the need for a future-state policy of the interaction of higher education and information and communications technology. The demographic expertise survey made clear that ICT is heavily used and is of vital importance to all stakeholders to complete their multiplicity of daily tasks of which instruction and learning is one.

Official, current Board of Regents policy allows distance learning for “clientele other than regular on-campus students and/or deliver such courses and programs in ways and/or at times and locations suited to the needs and convenience of those to be served.” [emphasis added] (University of Hawaii Board of Regents, 2002). The policy seems initially restrictive but then the policy allows for “convenience of those served.” (University of Hawaii Board of Regents, 2002). Since it is unlikely that any course is intentionally inconvenient, the policy appears to allow all courses to be delivered by distance learning. The purpose of the initially restrictive language contrasts with the open invitation. The policy additionally indicates the desire to extend “instructional resources to students anywhere in the state who are committed to higher education but are constrained by diverse factors which prevent their participation as regular on-campus students at the UH campus.” (University of Hawaii Board of Regents, 2002). This sentence
seems to contradict or limit the previous approval by emphasizing the supplementary nature of distance learning when a student is unable to attend on-campus. The policy goes on to indicate “distance learning is an integral part of the mission and a primary responsibility of every campus” but then the policy warns that quality in “learning outcomes appropriate to the degree or certificate to be granted and are of equal rigor and breadth as those required of on-campus.” (University of Hawaii Board of Regents, 2002).

The Distance Learning Action Plan states, “distance learning remains a critical imperative” and that “distance learning is not about technology; they are about academic planning” across all units of the university requiring “collaboration to maximize opportunities for students by leveraging the offerings available” (Distance Learning Committee, 2003). The report notes, “a strong commitment to distributed learning pervades the new University of Hawai‘i Strategic Plan.” (Distance Learning Committee, 2003). The report summarizes the Plan to include “mainstream institutional services”, “engage, develop, and support the University’s entire faculty and staff to create pervasive technology-rich environment”, “develop a coherent pricing and funding model”, and "resolve issues that create barriers to faculty participation.” (Distance Learning Committee, 2003).

There appear to be contradictory themes in official policy and the findings of committees charged with original guidance on distance education. The University might consider a discussion leading to better the alignment of the Board of Regents Policy and the various strategic plans. One question asked prior to attempting alignment is, “Why is there a distance learning policy different from any other learning?” Distance learning is learning with distance separation, time separation, or mediated communications channel delivery playing the same role as in any other part of the instructional mission.

5.2.2 Policy 2 - Workload
Participants identify increased workload in distance learning. Faculties indicate that distance learning preparation and delivery take more time and resources than a traditional, face-to-face course. They indicate inadequate recognition provided in comparison with normal workload. Students concur with the faculty belief. Faculties indicate little workload difference for
the student but students disagree. Students find workload increases for them in distance learning environments in comparison to on-campus conducted instruction.

Workload remains an issue unresolved. Faculty workload policy might evolve through normal negotiations processes. Absent and not considered is a similar policy for student workload. A reasonable policy on student workload might argue that no more effort should be required of a student in the identical course delivered in the traditional face-to-face manner than the same course delivered in distance education. Whether this is a reasonable expectation or not will be the result of further investigation.

5.2.3 Policy 3 – Misunderstanding Distance in Distance Learning

University policy contains procedures that differentiate distance delivered education from on-campus instruction. These procedures originated at the inception of distance education based upon a multiplicity of concerns including staffing, quality, control, and learning outcomes. The literature and the results of this dissertation’s demographic questions indicate a continuing concern about quality but concurrently call for an expanded and regularized distance education program.

Existing policy on distance learning, specifically as such policy defines distance learning and requires differentiated approval and review processes, appears to require revision to synchronize with current and emerging practice. The fixation of most policy is distance and outreach but as argued in the literature review physical distance is a factor but not the only factor in “distance.” A mission of this University is recognition of the island nature of the State that reasonably demands equalized access to higher education without the concomitant necessity to be present at the location from which the instruction originates.

Current University policy unnecessarily differentiates and places requirements upon distance education proposals not required of on-campus instruction. Existent policy calls for distance learning curriculum review separate and in addition to that required of on-campus, face-to-face, lecture instruction. An effectiveness review of this policy appears in order.
5.2.4 Policy 4 - Growth

Students believe distance learning will expand across the curriculum and expand in offering frequency. The faculty is more circumspect. National data indicates that expansion is subject area differentiated with some subject areas far more likely to deploy distance learning than other subject areas. The participant responses indicate differences in expansion for lower level courses as opposed to more advanced courses. Differences present roadblocks to students pursuing studies through distance learning.

Expectations of growth and the reality of growth differ.

A growth plan might consider the rapid expansion of Internet technology and the wide diversity of devices. Expectations of increased access rise as students, raised in the ubiquitous Internet communications generation, begin their University experience.

Expectations of increased availability of distance learning by students contrast with institutional directions. If the demand increases for distance learning by students, then policy response might consider the inclusion of proactive goals and activities to deliver distance learning. Budgets and hiring policy alignment facilitates increased deployment of distance learning.

5.2.5 Policy 5 – Consistency in Programs

Students, more than faculty, indicate inconsistency in the commitment to distance learning at a programmatic level. The issue is not particular courses offered but integration and consistency of the courses within a comprehensive course of study. The indication from students is a lack of consistency in availability of distance instruction courses and coordination with a complete program. Students find difficulty in scheduling a degree through distance education.

The literature indicated that many institution’s faculty believe that residence is an important aspect and thus compelling campus attendance and limiting options is laudatory.

In many early reviews of distance learning, for example the comprehensive report of the community colleges, many programs appear with commitments made to deliver rich, complete, and regular courses to distance learners. This research did not investigate the alignment of
original statements and actual delivery. It is beyond the scope of this research to assess the actual difficulty of completion against the planned programs within the various colleges.

A review of programs to determine the consistency and comprehensiveness of distance delivered programs appears desirable. A program is not one or two courses. A program is the larger effort leading to certifications and degrees. Nothing in this discussion argues that an entirely distance delivered education program is desirable or required. The question raised as a priority for students is the consistency of the communication and structure.

5.2.6 Policy 6 – Support

Participants indicate support as a necessity to effective distance learning. For faculty, support is training, equipment, professional services, time, and resources. Support for faculty also includes recognition in the tenure and promotion process of the commitment to distance learning. For students, support includes availability of technical staff on line using tools common to help desks. Help for students includes not only IT technical support. Many indicated the need for “support groups" and “course tutoring or homework assistance.” The literature review indicated that private, web-based services are emerging. Many book vendors provided tutoring support allied with textbook purchase. Social networks of content support are also emerging.

Technical and content support in a distance education context is new to the University setting. At present, an exploration of support to determine what types of commitment can be made seems most reasonable.

5.2.7 Policy 7 – Academic Institutional Support Services

Students indicate the need for support services beyond distance learning. Conceptually, they argue for “distance student services.” Transactional services like registration, transcripts, and program planning already appear online. Students seek all other services equally deployed with on-campus students.

A review of the role of student services in a distance education context appears needed. How services are delivered equally in the on-campus and distance education mode might reveal opportunities to better serve the emerging need.
5.2.8 Policy 8 – Loci of Control

All participants agree that control of distance learning content is a faculty responsibility. Policy recognizes that course and program offerings serve the University community but the decisions on delivery methodology and content lie within the faculty. Distance learning is not an exception. Literature review demonstrates that a major contributing motivator for distance learning originates from the IT staff within organizations. The relationship between instruction control and IT motivation is symbiotic. Current Board policy states, “It is the responsibility of faculty to establish or approve standards that apply to all instructional programs of their campus, and to apply those standards to on and off-campus programs in the process of program review.” (University of Hawaii Board of Regents, 2002).

The demographic and item list reports indicated some clarity issues about the source of motivation for distance education that are resolvable with clearer policy statements.

5.2.9 Policy 9 -- Quality

Quality of distance learning remains suspect. National data indicates that both students and faculty raise concerns. The literature review demonstrated multiple instances of institutions and degrees offered of lesser quality associated with distance learning. The concern always existed but grew with the broad reach of the Internet.

The Board of Regents Policy on distance learning reinforces the concern.

A fundamental requirement for distance learning is that the quality and standards of such courses and programs be comparable to other instructional programs of the University. Distance learning regular credit courses and programs result in learning outcomes appropriate to the degree or certificate to be granted and are of equal rigor and breadth as those required of on-campus programs. Mechanisms for the planning, delivery, and evaluation of distance learning programs should assure the maintenance of standards and quality. It is the responsibility of faculty to establish or approve standards that apply to all instructional programs of their campus, and to apply those standards to on and off-campus programs in the process of program review. (University of Hawaii Board of Regents, 2002)

This policy implicitly assumes that on-campus courses represent a gold standard against which to measure distance learning. A not so hidden assumption of the language of the text demands monitoring of distance learning courses not required of non-distance learning courses. The literature review made clear that the assumption of superiority of either on-campus or
distance learning remains impossible to determine. The most supportable hypothesis is that no significant difference exists between the approaches.

A review and refocusing of policy might eliminate the policy difference. Leaving the inference that distance delivered education is suspect tarnishes courses and programs. The University is required to deliver quality in all it does. Thus, a resolution of the quality variance question appears necessary.

5.2.10 Policy 10 – Fees and Costs

Current policy tends to indicate that distance learning increases cost of delivery. National data argues that distance learning will decrease the cost of instructional delivery. The outcome depends on implementation.

The Board of Regents Policy states:

The University’s on-going goal is to offer distance learning credit instruction in Hawai‘i at a cost to the student similar [emphasis added] to charges for regular on-campus instruction. Applicable nonresident tuition applies for courses delivered outside the State unless superseded or supplemented by special credit course or other fees approved by the President. (University of Hawaii Board of Regents, 2002)

Popular belief supported in the literature argues that distance learning is less costly to an institution with the reduction reflected in student cost. Other documents advocate distance-learning surcharges to recoup IT expenses.

Students participating in this study support the idea of lowered tuition and expenses for distance learning delivered instruction.

The Board’s policy continues to outline requirements for non-credit instruction and the fee structure for non-credit.

Non-credit instruction, professional workshops, and similar instructional services should not be subsidized by fees generated by credit instruction, but rather, at a minimum, should be fully self-supporting by fees and tuition charged, plus any requested general fund support appropriated, allocated, and budgeted to support specific non-credit offerings. (University of Hawaii Board of Regents, 2002)

The discussion of fee structure for non-credit in the context of a Board policy on distance learning is somewhat inconsistent as the method of delivery of non-credit is not necessarily distance
learning related. The sentence following the discussion of non-credit then asserts demands for cost recovery outside of the State. The

Similarly, at a minimum, cost recovery is expected of courses and programs delivered to students outside the State of Hawai‘i. (University of Hawaii Board of Regents, 2002).

The discussion of cost recovery and equivalence appears to need review. Students indicate that distance-learning courses should have lower tuition. In some sense, such a policy becomes consistent with national data indicating increased efficiency of distance learning.

5.2.11 Policy 11 – Textbook and Materials

Technology is radically changing the textbook and media industry. Book purveyors like Amazon now provide eBooks right along with printed text. Book vendors like Prentice Hall, Wiley, and Course Technologies are all offering eTextbooks as alternatives to printed textbooks. Vendors deeply discount eTextbooks in comparison to printed texts.

Like open source software, consortiaums of faculty now produce textbook and supporting instructional resources. Many disciplines have the option to use a free, open source textbook.

A policy encouraging the increased use of electronic resources – regardless of distance learning – is increasingly seen in higher education.

5.2.12 Policy 12 – Skill Level and Target Audience

The literature review suggests that the most unprepared target audience for distance learning is lower level or less sophisticated learners. The literature argues that lower level courses have the least sophisticated learners whose study skills and self-discipline are inconsistent with the expectations of distance learning that demand good study skills and self-discipline. Regardless of educational expectations, the literature also suggests that the favorite target of distance learning is the lower level, larger, and more numerous in section count courses. The literature suggests that community colleges are more likely to deploy distance learning whereas professional schools, graduate division, and upper level courses are less likely. The differentiation continues, as certain disciplines are more likely to support distance learning while others characteristically eschew distance learning. The reasons for these differences relate to cost savings, seniority of faculty, flexibility, and many other factors.
No particular policy currently exists. Whether the literature is true in the instance of this system is unknown. The assumptions about learner preparation may also have changed. The literature did indicate that community college and first or second years courses are the most frequently targeted population.

The absence of policy may be the right standard. Yet, the absence also appears to beg a question of the appropriate use of distance education on communities of learners.

5.2.13 Policy 13 – IT Infrastructure Expectations
Participants expect decreasing support for University provided end-user equipment except for specialized equipment used in specific coursework or areas of study. The participants see reduced requirements for general laboratories or computers in open spaces such as libraries.
Participants believe individuals must provide their own equipment.

The seeded item list did not include statements of expected IT infrastructure and participants did not add expectations.

Participants expect individuals to provide their own end-use equipment. This view has benefit to the University as it plans. Computer labs and public resources may be reduced.

Expectations of basic infrastructure like networks, access points, and the like did not receive policy consideration in this study.
5.3 Web-based Policy Delphi Tool

The software developed to conduct the Policy Delphi intervention increasingly became important within the intervention. The Delphi literature and literature closely related to Delphi literature drove the functional requirements of the overall design. Various preliminary designs tested by volunteers provided functional feedback causing changes in interaction and sequencing of events within the system but were not critical in the actual functional design.

The design and other allied issues are now an important result of this intervention as they contribute to the knowledge that others who might pursue a web-based Policy Delphi should heed. The conclusions reached herein must be tempered with two constraints. First, this is a specific implementation driven by the perceived requirements of the literature as developed. No claim to some universally faithful representation of the literature is suggested. A different implementation might well emphasize a different set of functional requirements extracted from the literature and have different results. Although this is the case, it seems likely that the conclusions drawn are reasonable and sufficient to suggest to others limitations that require consideration prior to design or implementation. Second, distance education may not have been as compelling a subject matter to engage a variety of dedicated participants who then utilize the tool to complete the tasks. The impossible entanglement of the user base and its uses of the tool against the backdrop of the very tool itself confound the assignment of clear and unequivocal statements to outcomes.

Regardless of these limitations, the conclusions still have merit.

1.1.3 Participation Challenge

Policy Delphi, unlike traditional Delphi, targets stakeholders who are believed to have a heightened involvement and drive to participate in the investigation’s topic. The results of this intervention question the assumption that the draw of the topic is sufficient to entice and support continued participation through the Policy Delphi rounds.

As learned here, the draw of a seemingly seductive and important topic with an appeal directed at those most likely to be stakeholders is insufficient to motivate substantial participation. There are so many possible reasons for the non-response. First, regardless of how well written
and enticing the appeal to participate is the use of email may simply be ineffective. Today’s flood of messages and appeals may simply drown out the appeal in a sea of spam. There is also the possibility that filters blocked receipt of email to the targeted audience. Second, Policy Delphi looks like a survey but with considerably more effort required – a complex interface, expertise studies, feedback, item ranking on multiple voting criteria, and multiple rounds. These attributes may contribute to a judgment that too much time and commitment are required. Third, the use of human subject policies that do not allow remuneration to those who actually participate versus requiring that all solicited be equally rewarded challenges the resources available to nearly all doctoral researchers not independently wealthy. The only appeal to participation is reduced to either one of importance of the subject or some general reward to the institution for the participation of the individual. The financial remuneration structure runs counter to the “what is in it for me” response. Fourth, the assumption that a swarm or snowball approach to selection and participation requests failed to generate interest either on the part of the seeded participants or for those recommended by the seeded participants. Had this research relied upon the assumptions of the swarm and initiated itself with a very small initial seeding then it is quite possible that the results would have been so small as to be completely without merit. In particular, no student made a recommendation of any other student to participate. Thus, had not significant efforts been made through the cooperation of instructors to seek participation in their classes, and then few, if any, students would be included.

Participation is vital to any meaningful result. For any of the above reasons and possibly for some not recognized, the targeted solicitation of stakeholders does not appear to be a reasonable approach to generating interest. During the development of the seeds, Turoff suggested that getting official participation and the stamp of approval from the University would improve participation. Inquiries were made but the indications are that the University is not actively supportive of appeals using its mail lists or of sanctioning or expressing support of any particular institutional research. This indication is completely reasonable to avoid the inundation of the University with study after study with appeal after appeal to finish participation. In a recent Manoa campus survey that required far less work than a Policy Delphi and for which there was
official approval and support, the participation rates achieved a reasonable but relative small participation base. The creation of an overall direction for the campus must be assumed important to all the stakeholders – faculty, administration, students, alumni, and support personnel – yet the responses were not overwhelming.

Thus, a critical consideration for any future researcher considering a Policy Delphi is participation. Literature is available that indicates that surveys in general, and web surveys in particular, suffer from a high non-response rates. These instruments are decidedly less complex than a Policy Delphi. In addition, many of these surveys provide individual remuneration or enter participants into a pool from which winners are selected but neither of these meet human subjects constraints that limit enticement options.

5.3.1 Item List Generation

The purpose of this discussion must focus back on the original goal of the dissertation study question as below:

How do these items reflect the wider issues of organizational change that distance education is likely to encourage? How well does the list of items – once cleaned and consolidated for round one of the Policy Delphi – reflect a list generated from the literature review of this dissertation that may be wider in scope given the breadth and depth of the review?

The intent of this section was the generation of a list that would represent the view of the panelist as to the future interaction of distance education and the University’s instructional mission. How good the list is – in a universal sense – is beyond the scope of this work. Within the limits of the process described, an initial seed list of items was constructed from the literature. The panelists will judge its usefulness.

As proposed toward the start of this subsection’s discussion, the role of the item list after panel initiation is also a subject rich for investigation. As with the issue of the creation of the initial seeded item list, a discussion of adding items during the execution of the Delphi, is absent in the literature. One might expect this process modification to be common to Delphi because Delphi, as a set of communication acts, has to evoke new understandings from participants. These understandings were not previously contained in the seed item list. Of course, it is possible that
the item list was perfect at initiation or those panelists have no new ideas. Yet, there does not to appear to be literature on this subject.

The implementation of Delphi in this research allowed the introduction of new items. The results appear elsewhere.

The potential for further investigation is ripe.

1. Do panels regularly generate new items? Is there an expectation of the number of items that might be generated?
2. What is the role of the length and thoroughness of the initial seeded item list on the discussion and the generation of alternatives? Does a comprehensive list – or simply a long list – reduce or stimulate item generation?
3. What quality control will be placed on items? How are duplicates eliminated? How is wording clarified?
4. When an item is introduced, how does it garner votes? How does a new topic compete with topics that have been previously included?
5. The item list can grow long. Some believed that categorization of the item list was critical to weighing items or at least critical to contrasting similar items with each other. However, how is this ontology to be imposed? What constitutes a category? Who will make this category? How will items be assigned to a category? What affects will a category have on how an item is treated?
6. Can items be written for a Policy Delphi that span the potentially very divergent target population without losing so much precision that they become bland tautologies?
7. How stringent a set of rules similar to those used in the construction of survey items should apply to Delphi items? How much rewording of participant proposed items should be done? By whom?
8. Turoff suggested that some of the items not be neutral – that some items should be intentionally provocative to stir discussion. What is the effect on the discussion of seeding such items?

Delphi is a meaning-making collection of communication acts. The artifact of the communication acts includes the item list, rankings, and comments. Little attention shown to the some of the issues identified above leading to a need of further research.

5.3.2 Construction of a Delphi Web Tool

The details of the construction of the web tool were thoroughly discussed. The original research question 3 raised the question as to whether a Policy Delphi web-based implementation consistent with the research could be constructed such that participants would be able to use it. The question did not suggest a full review of the success of the implementation. The question did not suggest that a comprehensive human factors review be conducted. The expectation was
minimal. Could a web-based application driven by the functional requirements as identified in the literature be constructed that Policy Panelist could use to participate in the process.

As documented, the process of constructing a functioning web-based Policy Delphi tool consistent with the literature driven requirements was challenging. The working tool could become the basis for further significant refinements.

“Could a tool be built” was the question. However, that question raised a host of additional issues. Many of these are discussed in the limitations section.

Two major themes appear. First, the Policy Delphi instrument is a complex implementation driven by the literature. A complex construct challenges expert and well-trained consumers; but in this case, relative novices use the complex construct. It is not just the functionality of the Policy Delphi web-based tool that is part of the complexity. The construct of Distance Education itself is complex. The understandings of the participants is limited by experience rather than broadened out by investigation and literature. The combination of a complex instrument and a complex subject compounds the issues. For example, Amazon’s website is well engineered from the human interaction perspective. The complexity of the construct – purchasing something – is rather well understood by the public. However, Amazon is not asking for evaluative functions, suppositions, or clarification of the process of purchasing like those required dealing with a futures state of distance education and University instruction.

Second, to be clear, this investigation was never intended to be principally focused on the process of developing a web-based software tool. Such a dissertation would have spent considerably more time focused on software, infrastructure, and classic design methodology. The web-based tool became a participant in the discussion of a Policy Delphi intervention. The tool acted to force refinement of those functional requirements drawn from the literature that are so easy to describe but so difficult to implement. The conversation between the tool and the literature produced a very rich set of observations capture in the limitations of this study and the areas in which further research is needed. Unfortunately this conversation tends to change the appearance of the focus of this work away from the Delphi towards that of software development.
5.3.3 Role of Feedback and Feedback Information and feedback’s role is at the very center of the justification for using Delphi as opposed to a normal survey.

Delphi is fundamentally a formal, structured communication method emphasizing feedback’s role in deepening and altering a participant’s response to an item under consideration. Delphi’s earliest implementations took for granted the role of feedback to change a panelists’ assessment. The change should be directional and toward the consensus of the group. Information compelled convergence – information reduced variance. Most Delphi research proclaimed the convergence to the unified position tantamount to reaching a “correct” answer. Through feedback and reasoning, convergence not only reduced variance but like the confidence interval, the process captured the true population parameter. As the literature review established, the support for the conclusion that consensus yields the “correct” answer is at best weak but more likely unsubstantiated. Delphi consensus yields an answer – not necessarily the correct one.

Why are Delphi’s results unsubstantiated? Delphi interventions typically predict events in the future. According to a number of articles in the literature, little motivation exists tens of years into the future to assess the success of very old, previous predictions. Even the assumption that researchers ten years or more after a prediction would know of the prediction’s existence challenges reality. For the few post-predictive investigations Delphi that have been done, the results remain unclear due to wording that may have been sufficiently nebulous in the original project. Data are not available to make the call of Delphi as a valid tool or not valid tool. Further, little if any interest exists to collect the data to make the determination.

Feedback’s effect on Delphi panelists’ activities remains unclear. Some Delphi research demonstrates that feedback, presented as summary rankings by other panelists, has little to no effect upon an individual panelist. Rather, the participant is just fine-tuning his or her own ranking based on re-reading or further consideration of the question.

The critical issue for Delphi is: “Does feedback matter?” If feedback does matter, how does it operate? What kind of feedback matters? For example, this study presented the results of
ranking graphically. Is graphic presentation effective? Is it too effective and too influential? Knowing how charts might overly influence a decision in a particular direction, is such influence a problem in Delphi? One might expect expert panelists to analyze effectively and be cautious in their use of ranking results. However, do differences appear when stakeholders, who might have highly diverse analytic skills, must complete a similar analysis?

A number of articles and reviewers including Dr. Turoff felt that presenting a summary data about the voting patterns of the panel before a panelist made their first vote created too strong an influence to think independently. The relationship between the first impression gained from some summary data and the voting pattern is an area for investigation. Most simple ranking systems show the current voting immediately while others demand your vote. The rationale for some systems that demand your vote first is the enticement to vote before the reward of seeing how others vote. The promise of the data are the enticement to participate. This enticement does not answer the influence question except that such sites have not addressed the issue, as it is moot to them.

Originally, the Delphi implementation in this research showed summary ranking data after the first vote is recorded. Many reviewers specifically mentioned that the lack of summary data created confusion. The reviewers asked that the summary data be displayed at all times. The final implementation shows summary rankings during the initial ranking process. The implementation decision to display or not display summary data prior to the first vote remains an open issue.

The questions remain open. Does the requirement to rank first before showing data entice more participation? Does the display of summary data before the first vote overweight and pre-prejudice future evaluation behavior?
How does the volume of comments matter? Suppose one item receives a substantial number of comments but another receives few. How is this interpreted? Is frequency domination tantamount to importance? How is the content of the feedback that may be contradictory digested in the decision-making process?

In this research, the ranking results appear categorized by faculty, student, staff, administrator, and other as shown in Figure 40. Is there a particular categorization that is important? Do these categories matter?

Turoff had suggested this categorization to be important. This returns to the participation argument and the idea of providing a sense of how the various major actors in the system are constructing their understanding.

Is the categorization by attributes of faculty, student, staff, administration, and non-UH valuable as shown in Figure 40 Delphi Voting? Instead, is the proper presentation the simple tally based upon the actual response categories of the attribute as shown in Figure 41? Does the selection of categories influence the voting? For example, when a student sees that faculty overwhelmingly support the need for skilled personnel to create distance education does this carry more, equal, or less weight than the same or dissimilar voting pattern by students? Do the categories carry unequal influence?

On the other hand, in this study there are twenty-four other categorical classifications possible from the demographic instrument used. Instead of a cross tabulation as done with faculty, student, staff, etc., suppose instead the categorical classification variable used was the
participant’s view of quality of distance education. What makes any cross tabulation more meaningful than the next?

This research introduced an additional functionality suggested in the literature— the ability to request research. Research plays a role like feedback but many identified it more like feeding – almost like unsolicited introduction of supposed “almanac” facts. What are the proper mechanisms of this process? How different is the impact upon the panel from the normal feedback of ranking and comments? Does the super role of research carry more or less weight? Is it fair and reasonable to introduce extraneous and supposedly authoritative material into the discussions of the panel?

Eventually, this discussion boils down to the question – Do panelists change their vote and is the feedback or feeding flow influential in the decision to change?

Ancillary questions also arise. If panelists do change responses, which of the influences is the dominant influence? How many times do they change? Many suggest the first decision is the final decision and few move from that position. When they change, do they change toward the consensus?

All of the questions above appear in threads throughout much of the Delphi literature and remain open and important.

5.3.4 Consensus and Concluding the Delphi

When does a Delphi panel conclude? As discussed in its original construction, a Delphi panel concludes when consensus emerges. Consensus operationalized as a decrease of variance between the ranking responses of the individuals or a marked decrease in changes of ranking. Sometimes the panel just expresses disinterest and the Delphi ceases. Participant mortality often determines termination. Many Delphi facilitators use an arbitrary cut off x-number of rounds with x usually varying from three to five. No single approach dominates the Delphi literature. As important as this issue of termination is, the literature is, at best, scarce.

What is the equivalent process to terminate a Policy Delphi? Turoff and others have suggested about five rounds. However, Turoff has also suggested the elimination of rounds. One possible measure is lack of interest. If no changes in rankings occur, no addition of participants,
or no addition of items, then the Delphi concludes. While the Delphi remains active, it is possible that the results and understanding could be changing.

A literature search failed to discover literature on the termination of a round-less Delphi. The possibility exists that a consensus measure might indicate decreasing interest. Termination is an unknown.

There appears to be a related consequence of the decision to go without rounds that has a major impact on the performance of the intervention. Since there is no specific time demarcation, it becomes unclear as to when appeals should be made to participants to revisit the Delphi tool to review rankings and comments of others. This lack of specific timing also tends to allow participants to reduce necessary tension to complete initial participation, to provide feedback, or to revisit the site. Roundlessness requires frequent reminders to participate and review and this action only compounds the flood of email that may be ignored.

5.3.5 Item Attributes and Cumulative Ranking

The usual ranking process in a Delphi panel uses a single attribute or single scale. Alternative approaches use no scale but call for the panelist to rank items from highest to lowest using whatever means the panelist feels appropriate and whatever attribute satisfies their understanding.

Previous Delphi investigations used a single ranking attribute for each item. Turoff’s research on Policy Delphi suggested a different approach. In place of a single attribute, each item’s review used four attributes: feasibility, importance, desirability, and confidence. Turoff reviewed this dissertation work and suggested a reduction two attributes: feasibility and desirability. He made specific comments about confidence as an attribute being inappropriate for a Policy Delphi. Other reviewers suggested that the importance attribute must remain. The final implementation uses feasibility, desirability, and importance. Confidence was eliminated.

Though the three attributes appear in literature, no literature appears justifying these three measures as the most important measures. No literature appears discussing how completely these attributes capture all the issues associated with future policy items. This leads to the obvious shortcoming and research need: What are the “correct” attributes for a Policy
Delphi? Is there a sufficiently robust and differentiating set with universal applicability? The extension of this argument then asks – are the attributes of an item used during the ranking process contextual to the subject matter of the Delphi?

If more than one attribute is appropriate, how is overall ranking achieved? Prioritization of items in the Delphi is a fundamental expectation achieved with relative ease on a single attribute but what score for each item consolidates the various dimensions represented when multiple attributes per item apply? Is the consolidation methodologically valid? If no consolidation passes methodological validity requirements, does the list display ordered by each attribute? That is, should the software provide the user the ability to select the attribute for ranking and display the list according to the choice? The multiple display option is relatively simple to implement but how will it be used? Does a multiple ranking confuse or clarify? Does one of the attributes become the dominant selected viewing attribute? For example, do most participants request “importance” as the display order? If participants’ default item list display utilizes a dominant ordering, should we bother with the other attributes even if they are well grounded in theory? Should the practical trump the theoretically identified attributes?

Suppose that the current three-attribute scale is sufficient. Each panelist ranks each item on a Likert scale for each attribute. Suppose further, that a weighted measure can be constructed to consolidate the three scales. Essentially, eliminate the previously identified issues. Now, this elimination introduces an even more interesting set of questions. The Policy Delphi is ultimately about tradeoff between alternatives. The ranking of items by individual attributes forces no tradeoff between items. The only tradeoff would be the tradeoff within the item within the attribute’s scaled values and not even between attributes of the same item. Simply put, if the panelist is considering the “feasibility” of requiring students to all have their own computers, the only focus demanded is on the scale of “very feasible” to “not very feasible.” A tradeoff between feasibility, importance, and desirability is unneeded. Nothing forces interplay. Even more interesting, nothing forces interplay and tradeoff in the total scheme of items. Every item could be highly rated on every scale without regard to how the item interacts with other items. If you suppose that every item uniquely and independently provides a dimensional perspective of the
future state, this issue is not important. However, most items interact yet no tradeoff is required. Should tradeoff be forced on the overall item list? For example, should there be 100 points in total for importance distributed over the items list? This distribution of voting points mimics the market-based approach where items compete for resources. The result is a weighted tradeoff. The complexity of this model of inter-item weighting with each item having three attributes could very well exceed the ability of the panelist to comprehend or even further discourage participation. Inter-item tradeoff is important. The mechanism to be used requires further investigation.

Again, in classic Delphi clear rounds exist. Most uses clear the results of the previous round as the succeeding round commences though this is not always the case. In a continuous Delphi, no rounds exist. All ranking is cumulative. Where one might assume that clearing the previous might compel reasoned change in classic Delphi, does that cumulative, unchanging ranking of roundless Delphi introduce stagnation? Are participants more or less likely to make changes? This discussion raises another serious issue related to retention and involvement. Rounds forced the participants to “get back into the game” – they had to rank and participate to register their ideas. In continuous Delphi, no such compulsion exists. Does the lack of definitive rounds increase mortality? If mortality increases, should rounds be imposed? Are rounds and mortality related? Are rounds and participant re-evaluation of item ranking related?

5.3.6 Change during the Interaction

Most classic Delphi begins with a fixed item list constructed by the facilitator and a fixed participant list selected by the facilitator.

This implementation released both of these fundamental constraints. Though the item list and the panel are seeded, mechanisms in the software allow the addition of either items or participants as the panel continues its deliberations.

Little research on Delphi discusses the issues that are introduced with these procedural changes.

Do panelists introduce items and members? Is there a point in the conduct of the process where introduction of either is more likely?
Previously argued, Delphi is a meaning-making activity. The meaning constructed is usually considered fixed to the panel doing the meaning-making. How do the dynamics of adding items and adding members change the Delphi meaning-making process and outcomes?

As the composition of the panel might change with time as original members drop and newer members join, what are the implications? For example, if a member “disappears” should their previous votes be removed as they are no longer part of the process? If rankings start to change due to these dynamics, when is convergence or agreement reached? Who is agreeing? Can one claim that at some point, the results represent everyone who participated or should those claims be limited to only those currently participating?

Allowing change in item and panelist as implemented in this research thus triggers a wide array of new opportunities. Potentially, investigations in the social media arena, crowd sourcing, and other fields can contribute to the Delphi.

5.3.7 Requesting Research

The capability to request almanac research on a particular item may not be needed. The literature indicated this as a valuable function but it went unused. The judgment to eliminate the functionality might be premature at this point because a sufficiently rich interaction of participants during which a research question might surface was not available. As previously note, the complexity of the interface to support ranking across three voting categories, creating and reading comments, and other function may have obscured this need.
5.4 Significance and Potential Contributions

These two sections attempt to place this research in the panoply of Delphi investigations. Further, the research hoped to provide a more serious discussion of the interaction of ICT and University instruction.

5.4.1 Significance for Policy

As was argued, this research is clearly in the qualitative genre. Claims to generalization would be inconsistent with the known and accepted limitations.

What was learned from this intervention for those who participated was captured in the themes built in the conclusions. These themes indicate that a much more serious planning effort is needed to outline the future state of the relationship of the instructional mission of the University and distance education. However, what became clear in the investigation of distance education is that distance education is a poor labeling of the interaction of ICT and higher education. The interaction is far more complex. The issues are not unlike those with which the private sector has dealt with.

The pressures on resources will not cease. The demands for access to higher education throughout one’s life are expected to grow. The drumbeat of politicians for a more educated workforce only increase. With these demands comes the reality that higher education costs are increasing at a rate that exceeds the CPI and are demanding a larger and larger share of severely strained government resources. The private sector has harnessed technology to deliver significant productivity gains. The private sector recognized an average over 3% per year increase in productivity in the last decades. ICT played a major role. Eventually, Universities will have to constrain their demands on public finance and one way to save the essence of the institution is to make more and better use of ICT.

5.4.2 Significance for Practice

Futures foresighting remains a vital tool for decision makers to allocate current resources for future benefit. As outlined in the Armstrong guide to the selection of a foresighting methodology, the Delphi is a key technique for painting a picture of futures outcomes. The Policy Delphi extends the classic Delphi by replacing a panel of experts with a panel of stakeholders.
This dissertation comprehensively reviewed the literature on Delphi. The literature review drove the functionality of a web-based tool.

The functioning of the tool itself during this research was interesting and did produce very limited results but it was the demand for clarification of assumptions of the Delphi technique that the tool required that are the significant contributions to practice. These issues and limitations are discussed in a detailed fashion in the sections that follow.

5.5 Limitations and Suggestions for Future Research

In the eyes of this researcher, this investigation has more than likely yielded far more questions than answers. Maybe the result of an intense period of investigation should always yield new opportunities. The limitations and suggestions for further research are highly interrelated because the limitations provide the opportunity for more investigation.

5.5.1 Future Technical web tool Design and Implementation

The tool used in the Delphi is no more neutral to the process then are the other areas of concern raised above. The tool reflects the functional requirements and as such must await exploration of the issues of Delphi itself.

Delphi is a complex system. Delphi could be made simpler once decisions are made on functionality including ranking attributes, identification of participants, threaded, or non-threaded comments, consolidation, and each of the areas of further research on Delphi previously suggested. Once functionality stabilizes, the initially important issue that develops is how to communicate the functionality of the complex model to a target population. The population engagement in the object of the Delphi might be assumed but the tool itself might be a hindrance of effective participation. Yet, if the simplification takes place that so reduces the Delphi just to make interaction more palatable then what value does the technique have? The interplay of functionality demanded by the methodology and the limitations of pushing the functionality to the target population is another area of needed research.

Consistent with the exploration of the tool is a discussion of the potential differences in implementation between target audiences. If the target audience is a set of experts as in classic Delphi, is more of the functionality of feedback, complex ranking, data exploration, and so on
more likely to be included and be of benefit? However, if the audience is more novice then would far less functionality be included? Is there a universally applicable Delphi implementation that crosses from the classic expert model to the Policy Delphi?

This researcher became more and more aware of the need to increase the sophistication of the client-side code to create an increasingly user-friendly environment. The purpose of this code is to reduce the mental load of understanding the tool in contrast with the subject of the panel discussion. As noted in the discussion of all the changes to the interface, the ranking process for an item does an automatic post back without the user being required to hit submit buttons. Everywhere in the interface, there are opportunities to improve interaction. What is unclear is how much more work must be done on the interface functionality versus the total complexity of the system and how these two trade off. Can interface code mask the complexity issues?

The training and tutorial capabilities of the system need review. Beyond the fundamental issue of length and detail of the videos as a reason to delete these tutorials, a number of technical issues appeared with playback. Some users were unable to view videos. Some users did not have audio capability. Some users had both but neither worked. The attempt to use video and audio met with many technical challenges. Multiple paths must be available meaning more development time. The research question is: Does it make a difference? Video is clearly more intuitively appealing but is it necessary? When doing training and dissemination to a wide audience, what is the simplest, most assured approach to gain understanding?

A number of other issues logged below remain interesting for future development.

1. Creation of the comprehensive system that includes an over-arching maintenance capability to allow for multiple instances of Delphi. This must include configuration options to allow varying functionality once the functions are identified as discussed previously.

2. The insertion of a comprehensive functionality monitoring capability. User clicks, menu selections, and use of extended capabilities must be recordable. Many of the questions concerning functionality may be answered through use or non-use. Some of these event tracking capabilities should focus on these actions.

3. Tracking changes between logins of item rankings. Changes while in same session do not count
   a. Track use of detailed survey/expertise
   b. Track requests for comments at all

260
c. Track requests to read research
d. Track the use of tutorials

4. The entire interface requires the review of professional HCI designers. The resolution of some system complexity might be resolved through more intuitive and more flexible interface design.

5. Turoff suggested that panelists should be allowed to edit or delete comments. Though not difficult to implement, does the ability to change a comment after the fact change the history and thought processes. What does editing a comment that drew the comments of others do to the entire train of thought. From the perspective of the originator the ability to edit might be important but how does this change the meaning-making and use of the comment by others?

5.5.2 Richness of Possible Future Research

Delphi is an important methodology frequently utilized in futures research though it is increasingly appearing in the literature as a team consensus-building tool not specifically applied to futures research. Many health care teams use Delphi to determine the best protocol and intervention for a given health issue. Delphi remains the most frequently used methodology for gathering and consolidating expert opinion. Delphi appeals to policy makers who need a set of plausible futures to allocate resources today. Many examples appear in the literature where Delphi applications aided governments to plan current investment in infrastructure for future societal benefit. In the previous sections a plethora of potential research issues appeared. Each of these issues calls to those interested in futures’ research to investigate and clarify our understanding of each of the issues.

This limitations and opportunities for research section largely ignores the vast field of distance education. The complexity and richness of distance education must be obvious to all. For this researcher, distance education motivated the interest in Delphi but Delphi methodology and all its attendant issues became the true centerpiece of the research. Distance education was the vehicle that lead to the interesting problems of how do we talk about futures of which distance education is one? This in no way decreases the importance of the actual issues of distance education. For this researcher, the futures impact of distance education on higher education appears to have the potential for some revolutionary change. The economic fact of the near 6% compound rate of growth in the cost of higher education coupled with the increasing realization by the public that so much government is not affordable are likely to compel a serious look at
distance education as an efficiency and cost savings approach. One only looks at the arguments made for information technology and medical records in the recently passed health care bill. Efficiency, effectiveness, and cost reduction appear throughout the argument. Health care cost increases are less than those of education so soon the arguments will transfer. This is not a bleak, doom and gloom thought but one of hope. We can appropriately plan and harness the capabilities that information technology brings to higher education to do the most cost effect and quality yielding job possible.

5.6 Lessons Learned

The formal process now completed, this section allows one last look at what was learned in this investigation from the viewpoint of the researcher. As has been the case throughout this entire investigation a bifurcated comment again arises to capture the dissertation’s fundamental focus – the construction of policy recommendations – and the construction and utilization of a web-based tool based upon the literature to collect data. Though these are lessons learned by this researcher that provide the re-emphasis of the value of this exercise, these lessons are also aimed at those who might decide to conduct a policy Delphi similarly structured or build their own web based tool.

5.6.1 Distance Education

Distance education, as the realization of the intersection of instruction and information with communications technology will not disappear and for a large number of reasons is likely to grow. If one reads the literature from key advocacy organizations like EDUCAUSE the argument for the expansion is monitored and the trend made clear. Whether it is student convenience in this learning on demand society or egalitarian extension to the masses of humanity unable to attend formal institutional learning or the reduction of the ever spiraling costs of higher education, distance education will increase penetration and application in higher education for these and so many other motivations.

The learning gained in this dissertation from the literature review and from the actual investigation follows. First, there are the issues of organizational change. Second, the nature of research found on distance education.
5.6.1.1 Distance Education and Organizational Change

Central to this Policy Delphi investigation is the issue of higher education and its interaction with ICT. There is nothing given or sure about this interaction. Technologists tend to see revolutionary trends that uproot the current higher education. Many others view the change as adaptive and evolutionary.

One of the extreme views of the interaction of higher education and ICT is that of Peter Drucker who predicts, “Traditional universities as we know them will become a big wasteland in the next 25 years. The future is outside the traditional campus, outside the traditional classroom. Distance learning is coming on fast.” (Drucker, 1997, p 1)

Clayton Christensen, the well-known business author who focuses on innovation, ventures into education reform in his very recently published Disrupting Class: How Disruptive Innovation will Change the Way the World Learns. Christensen believes that ICT will so customize instruction and focus on the individual learner at a cheaper cost and more effective delivery that it will greatly influence the stultified, unionized, failing school system. He writes,

The result of these four factors – technological improvements that make learning more engaging; research advances that enable the design of student-centric software for each type of learner; and inexorable cost pressures – is that 10 years from the publication of this book (2008), computer-base, student-centric learning will account for 50 percent of the “seat miles” in US secondary schools. Given the current trajectory of substitution, about 80% of the courses in 2024 will be taught online in a student-centric way. Given how long some have been in the trenches of the school reform, this will be quite a “flip.” (Christensen, 2008, p100)

Roadblocks like organizations, unions, and other value chain delivery systems are circumvented. The classroom will be global – networked, distance delivered, shared, and correcting all the previous failures.

Privateer’s comments take on overambitious comments that change comes from technological intervention. He does believe that change is needed and that fundamental reform is required. He writes,

The belief that technology is a panacea or that computers are universal change agents is evident in several cultural venues. In academia, there is a tacit assumption that instructional technologies can spearhead serious institutional reform because they create real change, especially in the area of course content and delivery. Nothing, however, could be closer to or farther from the truth. The implications of this naïve infatuation with automated learning are significant,
especially to the future of higher education as distance learning agenda continues to grow. Innovational use of technology, grounded in traditional notions of teaching, will not bring about important reforms needed to reduce the cost of higher education and create first-rate learning environments. In essence, higher education must understand and act. (Privateer, 1999, p. 62)

The reader cannot be left with the overestimate of change. Hamilton and Feenberg challenge the big bang change asking if “the rush to a technological revolution was all about money – money to be saved by substituting capital for labor on campus along lines familiar from many earlier de-skillings of crafts and professions with only secondary consideration given to the pedagogical and professional concerns that guided early experimentation and innovation.” (Hamilton & Feenberg, 2005, p97). For Hamilton and Feenberg the period of the 1990’s saw online education become “embedded in the rhetoric of reform” with the “virtual university the logical replacement for the cumbersome, rigid, and anachronistic ‘traditional’ institution.” (Hamilton & Feenberg, 2005, p 97).

Evangelists claim online education to be the transformative replacement and “brick-and-mortar universities would vanish to be replaced by the effervescent movement of digital information in global communication networks.” (Hamilton & Feenberg, 2005, p98). Networks supplant physical existence and university content would be unbundled and distributed in components. The lines of debate sharpen and as expected, the technology divides.

In the wake of the general disappointment with the exaggerated claims made for online education, there is now wide latitude for faculty intervention and participation in shaping the terms on which it will impact the academic labour process, the division of academic labour, and ownership of intellectual resources. It is now clear that online education will not destroy the university as we know it. What it will become will be determined ultimately by the politics of the very institution it promised to replace only a few years ago. (Hamilton & Feenberg, 2005,p 118)

The interaction of higher education and ICT is completely unsettled and will provide a rich platform for years to come. This researcher was at one time a proponent and then cautiously pessimistic during the overhype phase. Now, having read the research and seen what use is being made of ICT, the belief is one that the institutions will muddle through this interaction with unclear goals and wasteful experimentation. Ultimately, the trend for this researcher is clear that distance education in 15 years will be a major delivery mechanism for higher education.
5.6.1.2 Distance Education Research

This researcher spent months reviewing distance education research though the reduction of that work does not appear in this final document. The researcher felt that understanding the topic was important so that decisions on structure and implementation would be informed by research. Of course, the purist might argue that the only skill needed by a Delphi investigator is the Delphi process and as written into this document the focus became the Delphi technique and its results. Yet, had this research not been done, the construction of an item list would have been impossible. Thus, this researcher believes that Delphi investigators should have some basis in the subject of the investigation and that the literature review should contain the review of the subject. The problem for the final write-up is how much of each – the Delphi or the target of the Delphi – should appear. The choice made herein rests on the researcher’s belief that the contributions to future research that is most pressing and of value are those about the Policy Delphi intervention. The results and policy recommendations are not unimportant but the instrument issues appeared more important.

This researcher was struck by the volume of research in widely varying journals. The introduction of electronic publishing in an area that is itself largely dependent upon the Internet appears to multiply the number of articles and journals appearing on the broad subject of distance learning. This researcher observed a number of challenges in reading the literature. First, so much of the literature is written to compare and contrast distance education with traditional face-to-face education across hundreds of techniques, assumptions, and processes. Yet, all of this research fails to take into account one of the most basic findings about educational technology by Clarke. He states, “The evidence is overwhelming that media do not influence educational achievement” (R. E. Clark, 1983,p 485). Yet, researchers ignore this finding that remains basically unchallenged and proceed to attempt to show difference in achievement. When achievement fails, many turn to attitudinal measures. Second, researchers fail to report the corollary of the Clarke hypothesis of the “no significant difference” syndrome. Excellent reports and web sites report study after study that show no significant difference in attempts to influence achievement by any number of interventions – most based upon the assumptions that technology
will make a difference. Third, studies often fail on research design issues. Convenience samples seem to be carelessly used to extrapolate to populations without sufficient constraint on generalization necessitated by the research model. Fourth, few studies that report significance go further to talk about effect or total cost or total result. The significance, if found, fails to show real or practical difference. Further, these studies do not take into account larger issues of total cost of deployment outside the laboratory setting or of longitudinally lasting effect. Fifth, when all else fails, meta-analytic studies on a given topic appear and, as so many meta-analytic studies do, find that the intent of the investigation is supported. But, as this researcher has learned the issues with meta-analytic studies are numerous chief amongst are what is being consolidated as representative of the final construct and the well-known statistical result of significance increase as sample size is dramatically increased.

This researcher has concluded that for him that it is unlikely that we will ever know based upon current research models, whether distance education is more effective or efficient than other methods. The policy decision will as always be made with partial and incomplete information.

5.6.2 Policy Delphi

The Policy Delphi is unique in the matrix of extensions of the Delphi technique. This researcher experienced a challenge to find literature that evaluated many of the assumptions and procedures that extend traditional Delphi to Policy Delphi as proposed by Turoff in “Policy Delphi.” (Turoff, 2002). This investigation attempted to align as closely as possible with the methodological directives laid out by Turoff. In addition, the implementation-integrated capability in the web-based instrument that originated from an investigation into the literature that naturally extends the Delphi technique. These findings were some of the hard won lessons that accrue value to this investigation and provide guidance to those who might attempt additional investigations. One salient point must be made early. The results that follow are based upon the particular implementation and the particular circumstances of the organization and topic selected. The dissertation laid out the detail of implementation and circumstances so re-introductions or explanations of learnings will not appear in the below except as they clarify previous material.
5.6.2.1 Participant Expectations of Delphi as a Survey

The Delphi technique cannot escape its resemblance to a survey regardless of other claims. Potentially, if the Delphi matched more closely to the traditional technique where a limited number of experts participated then the survey-like structure would less limiting. In the Policy Delphi where participants are stakeholders, the framework of the instrument appears to constrain feedback because the respondents see their responsibility as a survey activity with one round of responding to items. The iterative feedback nature is highly challenged by the imposition of external expectations of the participants. This survey expectation is like that of juries today that having seen criminal investigative television programs where DNA is always available and conclusive expect such evidence and discount other forms like direct witness testimony. Expectations can overrule the best explanation and enticement to participate consistent with the goals of the Delphi technique. This researcher was surprised at the strength of preconceived notions brought forward into the research setting. Future researchers are cautioned that expectations that a Delphi is a survey may override the desire for a richer set of actions.

5.6.2.2 Item List Construction

Directly related to the survey expectations of participants is the role of items in the Delphi. The item list frames the topic and this framing contradicts with the need for a futures Policy Delphi in which the control of the discussion should originate from the participants whose understanding of the futures options are the desired outcome. The ideal approach is an initially empty item list that is built by the participant submissions. Unfortunately, a blank list circulated appears to get little if any response and the evaluators in this research made clear that something must appear. Those software professionals who previewed the Delphi tool argued that regardless of instructions or task definitional statements that some kind of list was necessary to provide focus. The problem with the a list constructed external to the actual participants—regardless of the process of construction by the researcher or by a small group of pre-participants— is that such a list can limit the futures options put forward by actual participants. Again, the Delphi is being made into a survey.
The absence of an item list in the initial solicitation may also decrease the decision to participate.

A long item list may limit participant initiated contribution. Both of these extremes – no list or exhaustive list -- appear to have effects that are not yet well understood.

5.6.2.3 Creating the Initial Item List

Constructing the initial item list was a large challenge. This researcher did not want an item list for all the reasons above. Unfortunately, as reported in the methodology section on the development of the instrument, the pre-participant reviewers disagreed. They argued that in the absence of a list to focus the discussion, that they believed future participants would disregard the entire solicitation. Constructing the list demanded a base and analysis of the options, which meant that this researcher had to go beyond the task of observing and understanding actions to that of a participant by way of the construction of the item list. This intervention to create the list is seldom addressed in any of the Delphi literature. No guidance is given. Sackman does criticize the treatment of the item list as a psychologically valid instrument but that criticism is post hoc. The creation of the list and its influence on framing the debate and outcomes is seldom addressed in any Delphi. This researcher attempted to construct a minimum number of items based on the literature using a mind map. There are also items in the list that are intentionally controversial or extreme positions as recommended by Turoff to stir the pot. The literature is silent on this proposition.

Those who might develop their own Delphi – policy or otherwise – are cautioned by this researcher’s experience to pay specific attention to the issues of the item list construction. To pay attention to the issue of how a long, comprehensive list over frames the discussion and short list might cause excessive dropout or lack of focus. The writing of the items, which to include, and so many other issues surround this vital step yet, as noted, the extensive review of Delphi interventions seem to miss the criticality of this issue. The difficulty of constructing the items appears to be more challenging in a futures policy Delphi as opposed to a much more targeted Delphi.
5.6.2.4 Stakeholders, Experts, and Motivation

The traditional Delphi is a round-based ranking or rating of items by select experts whereas a Policy Delphi replaces experts with stakeholders. An expert is presumed to have deep knowledge and deep attachment to the area of expertise that the Delphi is examining. A stakeholder is a person who has a concern, interest, or involvement in the area of examination but opinion – hopefully reasonably informed opinion – replaces expertise. The change from deep attachment and expertise to involvement and variable combinations of expertise and opinion more dramatically change the process than was originally thought. Distance education was thought to be a driving force and particularly so for those who are actually participating in it. From the levels of participation and involvement, distance education is far less a driving force. Those who might apparently been stakeholders have far less attachment to the process to motivate highly active participation in this research. All Delphi, and in fact all surveys, suffer from non-response and mortality. Both of these were highly evident. As involved as the targeted faculty was in distance education, the choice to participate was not stellar. In addition, even more so, the willingness to provide feedback, to re-rank, and to use the facilities of the system were simply not evident. Thus, researchers are cautioned that use of stakeholders in place of experts in the Delphi intervention is extremely challenging. The subject area may be important but it is not necessarily the case that it is sufficient to engender a potential participant to take action. One must suppose that this response would be even more challenging if the solicitation were aimed more broadly and to a less potentially involved target audience. For example, attempting to get faculty to respond who had no involvement with distance education would probably yield little return even though all faculty members should be concerned about the interaction. Futures researchers are cautioned to think long and hard about targets and responses and to make few assumptions about how actively responsive stakeholders might be. Further, to give thought to non-response bias introduced at all levels. The expertise model with a limited subset of active participants may yield richer results.
5.6.2.5 Roundless Delphi

Turoff proposed a modification to the original web-based Delphi collection instrument to eliminate rounds. Instead, the Policy Delphi would become a continuous discussion vehicle where participants were viewed as actively involved and would return regularly to revise their thoughts. In traditional Delphi, the clear demarcation of a round allowed the investigator to tally results and provide ranking feedback to the participants. There is clearly recognizable new feedback. Activity could be scheduled and enticements to consider the feedback could be made. In the continuous Delphi, the lack of rounds appears to reduce the value of the summary results that became a continuously changing status report. There is no point at which reminders to review rankings because there is no indication that significant, coordinated activity had concluded. The roundless Delphi does not appear to be a good choice to encourage participation or to provide regularized feedback that should drive revisions of item ranking by participants. The roundless Delphi may not be a good vehicle in a Policy Delphi environment.

5.6.2.6 New Participants, the Swarm, and New Items

As was described above, the purpose of the item list was initiation of the discussion on the futures topic. The software system provided the ability for participants to add new items based upon the participant's analysis of the current item list. Moreover, the system allowed new entrants as a change to traditional Delphi in which all participants are identified prior to the first voting rounds. This change was made because of the inclusion of the swarm participant nomination procedure that also continued throughout the Delphi's lifetime. If strict rounds had been used, the option to allow inclusion of new items and new participants may have been less likely but the structural changes made these options appealing. The results from this investigation were not encouraging for either of these two options. Though participants did propose additional items and did suggest inclusions of additional members, the count of each of these categories was small. One has to reserve judgment as to the actual effectiveness of the conclusion because the actual return after initial ranking by participants was so low that one cannot conclude if the options would have been better exercised. It is impossible also to separate the roundless Delphi effect and the stakeholder involvement effect. These operational assumptions also introduced a
new complexity beyond that already recognized above that rankings are constantly changing. Whether the ranking results or the feedback, the inclusion of new items and new participants change the construct being negotiated by the group. The mechanics of new items versus old items is never discussed in the literature. For example, an old item may get an index score of 2.5 out of 4 from all previous votes, which for argument assume is 100. A new item is introduced and gets 2 votes each of 4. The new item immediately ranks above the previous item and appears high in priority yet has not been considered in the same process. In a round-based Delphi, each item begins again its competition for ranking but in a continuous Delphi, the data is simply updated with previous round data simply melded into the ranking. This researcher cannot make the call on all these issues. Future researchers should consider carefully the roundless approach, the inclusion of participants, and the inclusion of new items. One should not simply say no to these modifications because doing so then makes the item list a closed end ranking survey exercise.

5.6.2.7 Feedback and Decision Making

A fundamental tenet of Delphi is the use of feedback from the group to make a more reasoned decision. Feedback should cause reconsideration. Either feedback comes to the participant as the ranking or the comments attached to an item. This implementation went further to provide almanac research requests that allowed participants to request fact based, neutral commentary on an item. For example, a participant might ask for budgetary data or the history of tuition increases when considering an item related to cost of higher education. The goal of each of these categories of feedback within the Delphi technique is a movement toward group consensus on item rank. Consensus is often inferred to be tantamount to the correct decision. Many Delphi studies implicitly make such assumptions. Many cite the popular press notion contained in works like the Wisdom of the Crowds by James Surowiecki. Yet, there is an amazing dearth of well-founded research that establishes this proposition in any generalized way. Many examples of crowd consensus leading to a correct conclusion do not establish the propositions validity. This researcher has learned that so many Delphi studies accept this implicit assumption of consensus is tantamount to correctness. A number of studies reported in the literature review
indicate that the primary feedback used by participants was not feedback, discussion, or any broader fact base but instead was simply the participant's previous ranking. That is, participants used their own ranking as the primary motivator to change a current rank.

The role of feedback in a Delphi remains an open question. This researcher had hoped to shed some light on the interaction of change in item ranking and feedback but the lack of successive participation prevented observation. The role of feedback to item ranking is a fundamental issue in Delphi. This researcher would strongly suggest that future research address this issue directly – how does feedback, feed in, and all other forms of data that might flow in and through a Delphi influence ranking? If it cannot be established that feedback influences ranking then this researcher believes the Delphi to be nothing more than a survey.

5.6.2.8 Multiple Attributes per Item in Ranking

The original Turoff article (Turoff, 2002) proposed that four attributes (importance, confidence, desirability, feasibility) be used to assess each item in a Policy Delphi. As explained, this research removed the confidence attribute though Turoff in review sessions suggested removing both confidence and importance. This researcher decided to include importance. Regardless of removing one or two items, the complexity of using multiple attributes in a ranking process is unwise. First, deciding the value of each attribute on a Likert-like scale that creates an index score that modifies a ranking is too complex for the participants. Participants cannot see the direct relationship of their action to the rank location. Some expressed simply being tired of considering so many options. Second, participants do not differentiate between the attributes and it appears that the value for one becomes the values for the others even though the attributes are different. Too many items and too many choices seem to confuse the participant. Third, the indirectness of the rating by attribute to arrive at an index and thus a ranking does not encourage tradeoff. The participant is not actively and visually making tradeoffs in an item rank. An additional to be considered is what ranking should the participant see from which to make decisions? Should the item list be displayed in rank order as the group result? as the individual's previous rank which would provide no further information to cause change? Alternatively, should the display provide the ability to perform ranking with feedback from both sources? This researcher
strongly recommends that during the active Delphi in which items are and participants are added that the only attribute be rank. The use of additional attributes seems to confuse and participants do not appear to differentiate. In a completely revised Delphi, a post process might consider the additional attributes.

5.6.2.9 Demographic and Expertise Survey

A unique feature of this Delphi intervention was the use of the expertise and demographic survey presented to participants before they entered the ranking process. This survey provided valuable contextual data about the participants. As was reported in the literature review, failure to provide context for a participant allows other respondents free reign to construct what image they choose of others in the conversation. This construction, of course, requires that some significant and regular dialog occur and that an identification of the participant be available. The decision to have only anonymous responses or those tagged with “handles” or pseudonyms or with actual names is an area of future research. Some insist that increasing identification limits comment while others argue identification increases responsibility. This research provided an identification number with all responses and the ability to recall the demographic data tied to the respondent. This choice preserves anonymity but allows easier identification of participant demographic and expertise characteristics. Like the construction of the item list, the construction of the demographic and expertise survey requires careful thought. The items in this investigation came from the literature review and distance education surveys of attitude of faculty and students. For some, the survey was too long but for this researcher the contextual data became increasingly important. This was an unanticipated benefit and an extension not described in the literature for a Delphi. The original intent of this research never called for the survey to be used in analysis. The survey was designed solely for the consumption of the participants themselves.

The length of the survey is an interesting point and is driven by establishing a review of the issues identified for the object of the Delphi intervention. If the volume of data were rich, then the demographic data would also provide interesting tabulations that crosscut the ranking data.

Does the belief that distance education is a fad influence the way ranking is done? For this
researcher, the availability of the context data was vital to the final construction of the policy. Any future Delphi investigations should always include the expertise and demographic survey.

5.6.2.10 Software Complexity

A Delphi exercise appears to be a relatively simple procedure when done with paper and pencil with only ranking and written feedback. This researcher and colleagues conducted Delphi investigations and found little difficulty (Gilbert & Remus, 1976; Remus, Sprague, & Gilbert, 1976). The intuitively simple procedure when translated into software is far less simple. It is the case that this investigation may have had unique requirements driven from the literature that in hindsight may have exceeded a minimum necessary to accomplish the data gathering requirements. Choices were made to implement a number of features and capabilities consistent with the researcher’s understanding of the literature. These features and capabilities had to be tested to determine usefulness as part of the investigation. At this point, it is not possible to separate these requirements and their encapsulation in the software implementation except to note that options added complexity that largely went unused. This researcher will not jump to the conclusion that lack of use in this instance is a systemic negation of the value of these capabilities. It is not beyond reason to claim systemic negation but it is also not beyond reason to say the interaction design factors of the interface played a significant role. The methodology chapter demonstrated the difficulty of evolving sequences, capabilities, interfaces, and literature expectations. This researcher has already made the argument that the multi-attribute ranking of each item must go in favor of a simpler, direct trade off ranking of an item. With changes, there are opportunities to retain some of the feedback mechanisms central to the Delphi and its assumption of reasoned modification of a position based upon feedback. Given a chance to do over, these interaction scenarios would be investigated and each of these tests may themselves be the subject of significant research.

5.6.2.11 Writing the Conclusions

In a traditional Delphi, the items are likely to be estimates of a proposed, specific event. For example, in how many years will 90% of all college students have a table or laptop computer? Items constructed with specific outputs provide a natural mapping of the item list to the results
portion of the research. Most examples of this style of Delphi research use a closed item list. Further, many uses of Delphi today – in particular, those examples from healthcare where the participants are developing a best practices protocol – are specific in context. Policy Delphi’s value is its attempt to address policy yet this is also its biggest weakness because policy development is broad, contentious, and futures oriented. The challenge is compounded further by the new items that have the capability to take the discussion in entirely unanticipated directions.

The result of the Delphi activity, like all research, is the results, findings, and conclusions. Already discussed were the issues of the tripartite attribute set and their consolidation into an index. From index scores, certain findings may be reported like the rank order of the items in the item list but rank order is probably the least interesting result. The individual items scale from facts with limited scope to broad, sweeping statements each of which must be woven into the pattern that is a policy statement. Policy statements are consolidations of observations written into a holistic format that can be operationalized in an organization. The mapping of items to policy is at best indirect and subject to any number of issues of validity, relationship to the item list, and bias of the writer. Recognizing, as this research did, that the Policy Delphi is not a conclusion but input into the larger decision-making process of an organization the fears are mediated because the review and consideration like that of a jury given evidence will result in the actionable decision. Regardless, going from the final item list, ranking, feedback, and demographic and expertise survey to a set of potential policy conclusions was a significant challenge for this researcher. Attempts to find procedures in the literature generally failed except to see examples of other research and the results of those investigations. Certainly, no algorithm exists to go from the empirical results to the policy recommendations. Future investigators should carefully consider this procedure because it will be the paramount challenge as one struggles with your own bias that has been informed by your literature investigation, your Delphi results, and your personal views. This researcher spent sleepless nights thinking about the policy construction.
5.6.3  **In Conclusion**

What appeared to be a reasonable project using a well-defined tool on a topic of importance turned out to be a challenge at every phase of the investigation. Nothing was simple. Nothing was well defined. Everything was always in flux.

But then, the fun of the research and the completion of this work was the turbulence and realizing both the end of this investigation and the challenge for those who might try another approach.
## Appendix 1 User Comments

<table>
<thead>
<tr>
<th>Item #</th>
<th>Item</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>421</td>
<td>Every student must have their own computer and other technology that meets University minimum requirements. The University will create the minimum requirement list. The University will not provide end user equipment.</td>
<td>university should provide access to university system computers (like community college or university computer labs) to distance education students as well as traditional student. its likely that not many students will use this but it should be available. if someone is an exceptional student they could do distance education using outdated software/hardware in combination with university provided equipment (like if they use their computer to watch lectures, read the book on their own time but complete assignments at a computer lab) this should especially be an option if the tuition for distance education is kept equal to traditional costs.</td>
</tr>
<tr>
<td>422</td>
<td>Every student will be expected to have a set of enumerated computer skills. Faculty and staff will be expected to have at least the same skills as students. The University will not provide training in these skills except through outreach or other remedial delivery.</td>
<td>in busa 320 the professor goes through what i need to know in excel step by step, granted not at a exception beginner level, but with enough explanation that my basic computer skills from high school and ics 101 allow me to keep up. the faculty needs to be able to do everything that the student will do throughout the course and they should be fairly good with the program that allows you to video your computer screen while recording the lecture (i think its white board but im not sure). requiring ics 101 for students is feasible for classes that will use a more than the basics but i feel that if you are in distance education from the beginning then you will learn everything you need to know over time....but if you get into distance classes later (300 and 400 lvl) then ics 101 or maybe an online test of some sort should be used to qualify students so they know they have the skills required for class (just like any remedial class or placement test)</td>
</tr>
<tr>
<td>423</td>
<td>Every degree program and all its courses must be available via both on-campus and distance education. Distance learners must not experience roadblocks in pursuing their academic objectives.</td>
<td>As the desire to attend school increases there should be both options as more and more working people are returning to school to get degrees and it is not possible without distance education. I feel that if there was a way to get your degree purely via distance education more people would be enrolling.</td>
</tr>
<tr>
<td>425</td>
<td>Every counseling or support service available to on-campus students must have an equivalent distance education service.</td>
<td>Current Staff should be trained in both face-to-face and DL issues.</td>
</tr>
<tr>
<td>Item #</td>
<td>Item</td>
<td>Comment</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>426</td>
<td>The University shall eliminate all printed textbooks and instructional material. All material must be available in digital, on-line form regardless of whether or not students are distance education learners.</td>
<td>not everyone can stare at a screen all day, students should be given an option between printed and online material. I personally find it very hard to read a textbook on the computer, I can only do it for 20 or so minutes at a time, where as I can read a physical textbook for hours</td>
</tr>
<tr>
<td>427</td>
<td>Commercial textbooks should be eliminated. In their place, the University must use open source and collaboratively generated material that is free of charge to the student.</td>
<td>Agree if the quality of free textbooks are those of commercial ones.</td>
</tr>
<tr>
<td>428</td>
<td>The traditional semester schedule will be gone or significantly reduced in importance. Learning schedules should reflect the demand, pull-driven model of the just-in-time education of distance learners.</td>
<td>This question seems to presume that universities are about nothing more than teaching. When exactly are faculty expected to do their own research?</td>
</tr>
<tr>
<td>429</td>
<td>Summer school is eliminated. The summer program is merged into the regular schedule which itself has changed to a flexible, demand drive model.</td>
<td>Why eliminate face-to-face instruction during the summer terms? This is a better form of education, and for those who wish to pursue it on-campus, there would seem to be no reason not to offer the traditional list of summer courses.</td>
</tr>
<tr>
<td>429</td>
<td>Summer school is eliminated. The summer program is merged into the regular schedule which itself has changed to a flexible, demand drive model.</td>
<td>I do not understand the question.</td>
</tr>
<tr>
<td>430</td>
<td>Tuition for distance education delivered courses should be set at one-third (1/3) the equivalent on-campus courses. On-campus courses demand expensive physical infrastructure that must be recaptured via higher fees. Distance education students benefit from their reduced demand on physical infrastructure.</td>
<td>I feel that tuition could probably go to about half of what it is now for distance education students-however I would not mind paying the same amount if we had access to better software through the university. My boyfriend gets free software (like windows 7 upgrade and anti virus) through his university. I feel that if UH supplied us with the tools we needed (like microsoft office and OS upgrades as needed) then this would offset the cost difference caused by the unused physical infrastructure</td>
</tr>
<tr>
<td>431</td>
<td>Distance education students should not pay fees assessed of on-campus students unless those services are actually provided to the distance education student. For example, a 100% distance education student would not use a bus pass. Fees relate to services actually provided.</td>
<td>On-campus students are regularly charged fees for services that they often do not use. For example, just this year a mandatory fee supporting athletics has been required of all students. Why should distance learners be exempt from this (unfair) practice when on-campus students are not?</td>
</tr>
<tr>
<td>432</td>
<td>Distance education students should be surcharged to offset demands placed on the communication and information infrastructure.</td>
<td>at most tuition should be the same as traditional classes, not more</td>
</tr>
<tr>
<td>Item #</td>
<td>Item</td>
<td>Comment</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>433</td>
<td>Because distance education will reduce the need for increased</td>
<td>of course cutting back on university spending would be good, but i dont think that construction should stop if there is still a need for better classrooms for traditional students.....i believe distance education should increase enrollment, not shift a lot of enrollment from traditional to distance. i feel distance education removes barriers that make it impossible for people who want to go back to school (like a disabled child who needs constant attention or odd work hours that make it hard to schedule classes)</td>
</tr>
<tr>
<td></td>
<td>University physical infrastructure, all construction of new classrooms and other on-campus infrastructure including faculty offices, meeting rooms, laboratories, and the like will have been stopped.</td>
<td></td>
</tr>
<tr>
<td>433</td>
<td>Because distance education will reduce the need for increased</td>
<td>You can go to a dentist that has only seen teeth on-line, but I prefer one with hands-on experience.</td>
</tr>
<tr>
<td></td>
<td>University physical infrastructure, all construction of new classrooms and other on-campus infrastructure including faculty offices, meeting rooms, laboratories, and the like will have been stopped.</td>
<td></td>
</tr>
<tr>
<td>434</td>
<td>Distance education learners should not pay out-of-state tuition.</td>
<td>Ill Phrased question. You need to distinguish between in-state students and out of state students.</td>
</tr>
<tr>
<td>434</td>
<td>Distance education learners should not pay out-of-state tuition.</td>
<td>if they still live in state and they pay hawaii taxes then why should they pay more for tuition? if its an out of state student taking distance courses then there should be a separate tuition chart for them because they wont make use of any university lab equipment, at most it will be phone bills, but they are out of state so they should pay a higher tuition, but possibly less than the out of state tuition costs for traditional students</td>
</tr>
<tr>
<td>434</td>
<td>Distance education learners should not pay out-of-state tuition.</td>
<td>Or perhaps there could be a &quot;distance&quot; rate for out-of state that was more than resident, but less than out of state.</td>
</tr>
<tr>
<td>435</td>
<td>The University should provide increased tuition support for distance education students to encourage growth in distance delivered instruction.</td>
<td>i dont think increased tuition support in terms of scholarships should be given...but if tuition on distance education is discounted because of the disuse of infrastructure and lack of software provided, then it might help to get student who wouldnt usually try distance education in the program</td>
</tr>
<tr>
<td>436</td>
<td>A separate organizational structure (university, college, campus) should be constructed for distance education so that its rules, regulations, expectations, and costs are physically separated from the on-campus model.</td>
<td>This proposal presumes that distance-learning can genuinely be an effective substitute for on-campus education in all of its facets and at all levels. This is an assumption that I think is utterly wrong.</td>
</tr>
<tr>
<td>Item #</td>
<td>Item</td>
<td>Comment</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>436</td>
<td>A separate organizational structure (university, college, campus) should be constructed for distance education so that its rules, regulations, expectations, and costs are physically separated from the on-campus model.</td>
<td>i feel all rules being implemented now are functional in both traditional and distance settings, the only difference that i feel would enhance the program is a discount on tuition for distance education (to improve enrollment, and possibly get students who might not otherwise try the program)</td>
</tr>
<tr>
<td>437</td>
<td>Using aggressive promotion and marketing, the University should dramatically increase student enrollment in its Outreach College beyond the current student audience by using distance education. The objective is to increase revenue to the University.</td>
<td>This policy would carry education in the wrong direction in this State.</td>
</tr>
<tr>
<td>438</td>
<td>Distance education is inappropriate for non-sophisticated learners. No student should be allowed into distance education during their freshmen year. Further, students who might be characterized as academic risks should not be allowed to take distance education courses.</td>
<td>In my experience, 18-20 year-olds are more capable than older students.</td>
</tr>
<tr>
<td>438</td>
<td>Distance education is inappropriate for non-sophisticated learners. No student should be allowed into distance education during their freshmen year. Further, students who might be characterized as academic risks should not be allowed to take distance education courses.</td>
<td>Too many of our on-campus students arrive today inadequately prepared to meet the demands of university-level education. Students who are not highly motivated, and properly prepared, are liable to do very poorly in distance-learning, given the even-lower levels of direct one-on-one contact with professors.</td>
</tr>
<tr>
<td>439</td>
<td>Distance education will be the preferred and encouraged mode of instructional delivery for large, undergraduate courses.</td>
<td>In my opinion distance education is ideal for classes that typically have large lectures simply because when in this setting its hard to ask questions during lecture because it could set the class back. in both online and traditional setting questions can be asked after the lecture and appointments can be made to meet the professor and further discuss issues....presence in the classroom is not neccessary. i feel video lectures as done in my busa320 class are the ideal lecture tool</td>
</tr>
<tr>
<td>440</td>
<td>All maximum enrollment limits should be eliminated for distance education courses.</td>
<td>I think it should be on a case basis as some class could handle more than current limit.</td>
</tr>
<tr>
<td>440</td>
<td>All maximum enrollment limits should be eliminated for distance education courses.</td>
<td>I still want to hear from my instructor when I ask him/her a question!</td>
</tr>
<tr>
<td>Item #</td>
<td>Item</td>
<td>Comment</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>441</td>
<td>Distance education instruction must be designed so that it does not require any more learner sophistication and self-reliance than would be required in on-campus, face-to-face instruction.</td>
<td>If by the question “how important is this item”, you mean how much does this subject matter, then yes! It matters. Students just cant expect to get away with the same things they got away with before. Just like a college degree used to matter, and now it really only means something if you have a grad. degree or a PHD. Learning how to become independent and study or do research because you need to and not b/c someone is reminding you to get it done, is what will get that student a job after college.</td>
</tr>
<tr>
<td>442</td>
<td>Distance education is inappropriate for graduate education. Distance education must be discouraged in graduate education.</td>
<td>Distance education methods can be adapted to support graduate students doing field research off-campus.</td>
</tr>
<tr>
<td>442</td>
<td>Distance education is inappropriate for graduate education. Distance education must be discouraged in graduate education.</td>
<td>i dont have a good grasp of what “graduate education” entails so i dont know if distance education is appropriate, sorry.</td>
</tr>
<tr>
<td>442</td>
<td>Distance education is inappropriate for graduate education. Distance education must be discouraged in graduate education.</td>
<td>So then are we saying that people who have worked toward their degrees in this medium has a less valuable degree? I think that there are multiple ways to learn and to work toward a graduate degree. Like all things, some programs are of high rigor and others are not...this is the case for traditional programs as well.</td>
</tr>
<tr>
<td>443</td>
<td>Distance education is a fad. The University must limit distance delivered education to no more than 10% of its instructional delivery. Distance education destroys the social fabric of the University.</td>
<td>Again, student is the consumer of the university setting. If they want to bypass the social fabric it doesn't mean they can't learn what is required for a degree. There are ways of socializing via the internet for things like group projects as needed. And presence in a classroom does not mean that a survey cant be conducted, it just means that a population of college students cant be required, other populations are available for surveying.</td>
</tr>
<tr>
<td>443</td>
<td>Distance education is a fad. The University must limit distance delivered education to no more than 10% of its instructional delivery. Distance education destroys the social fabric of the University.</td>
<td>Distance education is here to stay. That said, there are specific courses that require at least some Face-2-Face. In my field, it is important that students manipulate laboratory equipment in a group setting and learn how to interact in this environment.</td>
</tr>
<tr>
<td>445</td>
<td>Tenure and promotion guidelines must reflect the demonstrated ability of a faculty member to teach using distance education. No faculty should received tenure or a promotion if the only demonstration of teaching is via on-campus, face-to-face delivery.</td>
<td>At a university, all faculty are specialists in one thing or another. To suppose that all faculty must be qualified and competent at distance-learning is to grossly overestimate the significance of that form of education.</td>
</tr>
<tr>
<td>Item #</td>
<td>Item</td>
<td>Comment</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>447</td>
<td>Faculty must own the advocacy of distance education not the administration or support personnel. Faculty leadership and committees become the central driving force for distance education.</td>
<td>I feel that both faculty and staff should work together on advocating distance education, after all students are the consumers in the university setting, however support from administration and other personnel should not be discouraged.</td>
</tr>
<tr>
<td>447</td>
<td>Faculty must own the advocacy of distance education not the administration or support personnel. Faculty leadership and committees become the central driving force for distance education.</td>
<td>Instruction is clearly the kuleana of the faculty.</td>
</tr>
</tbody>
</table>
### Appendix 2 User Item Contribution

<table>
<thead>
<tr>
<th>Item#</th>
<th>Item</th>
<th>status</th>
<th>Rewording</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>anything on virtual reality?</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I feel that more and more people need the option of on-campus as well as online classes. As a working person, I would not have been able to attend college without the availability of online classes. Please keep and increase these, as this is the way of the future.</td>
<td>P</td>
<td>Distance education courses are readily available to allow non-traditional students to complete their college career.</td>
</tr>
<tr>
<td>3</td>
<td>Some of the items are so inclusive as to be unanswerable with the responses provided. Many of the final ranking statements are far too extreme to warrant input, for example: All publisher texts should be dropped in favor of Open Content? I am a huge Open Content proponent, but I can't support (or even respond to) the statement as it's posed.</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Distance education should be data and assessment driven. Completion rates, performance in follow-on courses, and assessment of achievement of learning objectives should all be tracked.</td>
<td>P</td>
<td>Distance education should be data and assessment driven. Completion rates, performance in follow-on courses, and assessment of achievement of learning objectives should all be tracked.</td>
</tr>
<tr>
<td>5</td>
<td>Distance learning should only be given to faculty who cares about their students. They tend to spend more time on their students progress because they care. Those who doesn't care only do so little because they're only doing it as a standard procedure only that's why they don't go the extra mile to reach out to their students.</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Distance Education should be offered to Oahu students just the same as other island students</td>
<td>P</td>
<td>Distance education is available equally to students throughout the State with no preference given to more remote students over those closer to campus instruction.</td>
</tr>
<tr>
<td>Item#</td>
<td>item</td>
<td>status</td>
<td>Rewording</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>--------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8</td>
<td>I hope you hear from more people like myself, who have been educators for decades, and who have a great deal of experience teaching graduate students. (I've been the graduate chair of my department for more than twelve years.) Only then will the pursuit of graduate studies (and even upper level undergraduate studies) by means of distance-learning begin to seem problematic (if then!) to someone who could have formulated this kind of questionnaire. Lord help us if the assumptions that shaped this questionnaire ever begin to influence the direction of higher education!</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Hybrid models should be developed in which combinations of distance learning and face-to-face learning are used within the same course or program</td>
<td>P</td>
<td>Courses that combine on-campus and distance education will be as available as either pure face-to-face or pure distance education.</td>
</tr>
<tr>
<td>10</td>
<td>There were no questions about what makes distance delivery different. The question regarding whether distance vs. traditional is better / worse / the same is misleading. I think for some things one is better and for other things the other. Questions about the specificity of distance delivery would be appropriate.</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I want to recommend John Casken <a href="mailto:casken@hawaii.edu">casken@hawaii.edu</a> and Lois Magnussen <a href="mailto:magnusse@hawaii.edu">magnusse@hawaii.edu</a> as super users of distance learning in our MASTers and PhD programs in nursing</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>I have limited interest and knowledge of distance education at Hawaii, and therefore no motivation to wade through the multiple options of the previous page.</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>A uniform platform (such as Laulima) and a minimally uniform approach or standard shared by online faculty is the key to ensuring some baseline consistency in quality and accessibility. At the moment there is a cavalier, &quot;Wild West&quot; range of online architectures.</td>
<td>P</td>
<td>The University faculty established a minimally uniform or shared standard to ensure baseline consistency in quality and accessibility.</td>
</tr>
</tbody>
</table>
### Appendix 3 Item Ranking

<table>
<thead>
<tr>
<th>Item</th>
<th>Desire</th>
<th>Feasible</th>
<th>Important</th>
<th>Stu</th>
<th>Desire</th>
<th>Feasible</th>
<th>Important</th>
<th>Fac</th>
<th>Desire</th>
<th>Feasible</th>
<th>Important</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>The creation of quality distance education depends upon specialized skills like curriculum specialists, graphics designers, and IT. The University must publish and provide a uniform set of services available to all faculties who convert to distance education or who need to maintain current distance education.</td>
<td>3.4</td>
<td>3.3</td>
<td>3.3</td>
<td>10.0</td>
<td>3.4</td>
<td>3.2</td>
<td>3.3</td>
<td>9.9</td>
<td>3.4</td>
<td>3.2</td>
<td>3.3</td>
<td>9.9</td>
</tr>
<tr>
<td>Distance learners must be supported by an active infrastructure of help desks capable of solving computer and instructional material needs of learners including on-line tutors in most subject areas. Students will be able to call a help desk 24/7.</td>
<td>3.5</td>
<td>2.9</td>
<td>3.3</td>
<td>9.7</td>
<td>3.6</td>
<td>3.0</td>
<td>3.5</td>
<td>10.1</td>
<td>3.5</td>
<td>3.0</td>
<td>3.4</td>
<td>9.9</td>
</tr>
<tr>
<td>The quality of distance education is highly suspect. The University must maintain regular controls on distance delivered instruction to insure that quality is not compromised in the rush to provide distance delivered education.</td>
<td>3.4</td>
<td>3.4</td>
<td>3.5</td>
<td>10.3</td>
<td>2.9</td>
<td>2.8</td>
<td>3.2</td>
<td>8.9</td>
<td>3.2</td>
<td>3.1</td>
<td>3.4</td>
<td>9.7</td>
</tr>
<tr>
<td>Every student must have their own computer and other technology that meets University minimum requirements. The University will create the minimum requirement list. The University will not provide end user equipment.</td>
<td>3.1</td>
<td>3.1</td>
<td>3.5</td>
<td>9.7</td>
<td>3.3</td>
<td>3.0</td>
<td>3.3</td>
<td>9.6</td>
<td>3.2</td>
<td>3.0</td>
<td>3.4</td>
<td>9.6</td>
</tr>
<tr>
<td>Distance education students should not pay fees assessed of on-campus students unless those services are actually provided to the distance education student. For example, a 100% distance</td>
<td>3.4</td>
<td>3.4</td>
<td>3.2</td>
<td>10.0</td>
<td>2.9</td>
<td>2.8</td>
<td>2.8</td>
<td>8.5</td>
<td>3.2</td>
<td>3.1</td>
<td>3.0</td>
<td>9.3</td>
</tr>
<tr>
<td>Item</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
<td>Stu</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
<td>Fac</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
<td>All</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td>education student would not use a bus pass. Fees relate to services actually provided.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faculty must own the advocacy of distance education not the administration or support personnel. Faculty leadership and committees become the central driving force for distance education.</td>
<td>3.0</td>
<td>2.9</td>
<td>2.9</td>
<td>8.8</td>
<td>3.4</td>
<td>2.9</td>
<td>3.4</td>
<td>9.7</td>
<td>3.2</td>
<td>2.9</td>
<td>3.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Every counseling or support service available to on-campus students must have an equivalent distance education service.</td>
<td>3.1</td>
<td>3.0</td>
<td>3.1</td>
<td>9.2</td>
<td>3.0</td>
<td>3.0</td>
<td>3.1</td>
<td>9.1</td>
<td>3.1</td>
<td>3.0</td>
<td>3.1</td>
<td>9.2</td>
</tr>
<tr>
<td>Every degree program and all its courses must be available via both on-campus and distance education. Distance learners must not experience roadblocks in pursuing their academic objectives.</td>
<td>3.5</td>
<td>3.1</td>
<td>3.6</td>
<td>10.2</td>
<td>2.4</td>
<td>2.3</td>
<td>2.8</td>
<td>7.5</td>
<td>3.0</td>
<td>2.8</td>
<td>3.2</td>
<td>9.0</td>
</tr>
<tr>
<td>Distance education instruction must be designed so that it does not require any more learner sophistication and self-reliance than would be required in on-campus, face-to-face instruction.</td>
<td>3.1</td>
<td>2.7</td>
<td>3.0</td>
<td>8.8</td>
<td>3.2</td>
<td>2.6</td>
<td>3.1</td>
<td>8.9</td>
<td>3.2</td>
<td>2.7</td>
<td>3.1</td>
<td>9.0</td>
</tr>
<tr>
<td>Commercial textbooks should be eliminated. In their place, the University must use open source and collaboratively generated material that is free of charge to the student.</td>
<td>3.4</td>
<td>2.7</td>
<td>3.1</td>
<td>9.2</td>
<td>2.8</td>
<td>2.3</td>
<td>2.8</td>
<td>7.9</td>
<td>3.1</td>
<td>2.5</td>
<td>3.0</td>
<td>8.6</td>
</tr>
<tr>
<td>Distance education will be the preferred and encouraged mode of instructional delivery for large, undergraduate courses.</td>
<td>3.0</td>
<td>3.2</td>
<td>3.1</td>
<td>9.3</td>
<td>2.2</td>
<td>2.7</td>
<td>2.7</td>
<td>7.6</td>
<td>2.6</td>
<td>3.0</td>
<td>2.9</td>
<td>8.5</td>
</tr>
<tr>
<td>Item</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fac</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----</td>
<td>----------</td>
<td>----------</td>
<td>---</td>
</tr>
<tr>
<td>Using aggressive promotion and marketing, the University should</td>
<td>2.7</td>
<td>2.8</td>
<td>2.8</td>
<td>8.3</td>
<td>2.6</td>
<td>2.9</td>
<td>2.8</td>
<td>8.3</td>
<td>2.7</td>
<td>2.9</td>
<td>2.8</td>
<td>8.4</td>
</tr>
<tr>
<td>dramatically increase student enrollment in its Outreach College</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beyond the current student audience by using distance education.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The objective is to increase revenue to the University.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuition for distance education delivered courses should be set at</td>
<td>3.4</td>
<td>2.8</td>
<td>3.1</td>
<td>9.3</td>
<td>2.2</td>
<td>2.3</td>
<td>2.7</td>
<td>7.2</td>
<td>2.6</td>
<td>2.9</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>one-third (1/3) the equivalent on-campus courses. On-campus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>courses demand expensive physical infrastructure that must</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>be recaptured via higher fees. Distance education students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>benefit from their reduced demand on physical infrastructure.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The traditional semester schedule will be gone or significantly</td>
<td>2.6</td>
<td>2.5</td>
<td>2.6</td>
<td>7.7</td>
<td>2.4</td>
<td>2.7</td>
<td>2.9</td>
<td>8.0</td>
<td>2.5</td>
<td>2.6</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>reduced in importance. Learning schedules should reflect the demand,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pull-driven model of the just-in-time education of distance learners.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance education learners should not pay out-of-state tuition.</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
<td>8.1</td>
<td>2.3</td>
<td>2.7</td>
<td>2.7</td>
<td>7.7</td>
<td>2.5</td>
<td>2.7</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>The University should provide increased tuition support for distance</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
<td>8.1</td>
<td>2.5</td>
<td>2.6</td>
<td>2.8</td>
<td>7.9</td>
<td>2.6</td>
<td>2.6</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>education students to encourage growth in distance delivered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>instruction.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Every student will be expected to have as set of enumerated computer</td>
<td>2.5</td>
<td>2.6</td>
<td>2.8</td>
<td>7.9</td>
<td>2.5</td>
<td>2.4</td>
<td>2.7</td>
<td>7.6</td>
<td>2.5</td>
<td>2.5</td>
<td>8.8</td>
<td></td>
</tr>
<tr>
<td>skills. Faculty and staff will be expected to have at least the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>same skills as students. The University will not provide training in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>these skills except through outreach or other remedial delivery.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

288
<table>
<thead>
<tr>
<th>Item</th>
<th>Item Description</th>
<th>Desire</th>
<th>Feasible</th>
<th>Important</th>
<th>Stu</th>
<th>Desire</th>
<th>Feasible</th>
<th>Important</th>
<th>Fac</th>
<th>Desire</th>
<th>Feasible</th>
<th>Important</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>436</td>
<td>A separate organizational structure (university, college, campus) should be constructed for distance education so that its rules, regulations, expectations, and costs are physically separated from the on-campus model.</td>
<td>2.7</td>
<td>2.5</td>
<td>2.5</td>
<td>7.7</td>
<td>2.2</td>
<td>2.5</td>
<td>2.6</td>
<td>7.3</td>
<td>2.5</td>
<td>2.5</td>
<td>2.6</td>
<td>7.6</td>
</tr>
<tr>
<td>438</td>
<td>Distance education is inappropriate for non-sophisticated learners. No student should be allowed into distance education during their freshmen year. Further, students who might be characterized as academic risks should not be allowed to take distance education courses.</td>
<td>2.3</td>
<td>2.4</td>
<td>2.6</td>
<td>7.3</td>
<td>2.5</td>
<td>2.6</td>
<td>2.8</td>
<td>7.9</td>
<td>2.4</td>
<td>2.5</td>
<td>2.7</td>
<td>7.6</td>
</tr>
<tr>
<td>432</td>
<td>Distance education students should be surcharged to offset demands placed on the communication and information infrastructure.</td>
<td>2.0</td>
<td>2.4</td>
<td>2.4</td>
<td>6.8</td>
<td>2.2</td>
<td>2.8</td>
<td>2.8</td>
<td>7.8</td>
<td>2.1</td>
<td>2.6</td>
<td>2.6</td>
<td>7.3</td>
</tr>
<tr>
<td>424</td>
<td>If the current University structure is unable to commit to academic programs that are 100% distance education with all courses and requirements available without roadblocks, then the State must create a separate institution parallel to the on-campus model.</td>
<td>2.8</td>
<td>2.6</td>
<td>2.8</td>
<td>8.2</td>
<td>1.9</td>
<td>2.0</td>
<td>2.6</td>
<td>6.5</td>
<td>2.3</td>
<td>2.3</td>
<td>2.7</td>
<td>7.3</td>
</tr>
<tr>
<td>429</td>
<td>Summer school is eliminated. The summer program is merged into the regular schedule which itself has changed to a flexible, demand drive model.</td>
<td>2.2</td>
<td>2.2</td>
<td>2.6</td>
<td>7.0</td>
<td>2.4</td>
<td>2.7</td>
<td>2.7</td>
<td>7.8</td>
<td>2.3</td>
<td>2.4</td>
<td>2.6</td>
<td>7.3</td>
</tr>
<tr>
<td>440</td>
<td>All maximum enrollment limits should be eliminated for distance education courses.</td>
<td>2.6</td>
<td>2.4</td>
<td>3.1</td>
<td>8.1</td>
<td>1.3</td>
<td>1.7</td>
<td>2.9</td>
<td>5.9</td>
<td>2.0</td>
<td>2.1</td>
<td>3.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Item</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
<td>Stu</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
<td>Fac</td>
<td>Desire</td>
<td>Feasible</td>
<td>Important</td>
<td>All</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>445</td>
<td>2.2</td>
<td>2.5</td>
<td>2.5</td>
<td>7.2</td>
<td>1.9</td>
<td>2.1</td>
<td>2.7</td>
<td>6.7</td>
<td>2.0</td>
<td>2.3</td>
<td>2.6</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tenure and promotion guidelines must reflect the demonstrated ability of a faculty member to teach using distance education. No faculty should received tenure or a promotion if the only demonstration of teaching is via on-campus, face-to-face delivery.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>433</td>
<td>2.1</td>
<td>2.3</td>
<td>3.1</td>
<td>7.5</td>
<td>1.5</td>
<td>2.7</td>
<td>2.7</td>
<td>6.1</td>
<td>1.8</td>
<td>2.1</td>
<td>2.9</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Because distance education will reduce the need for increased University physical infrastructure, all construction of new classrooms and other on-campus infrastructure including faculty offices, meeting rooms, laboratories, and the like will have been stopped.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>442</td>
<td>1.9</td>
<td>2.0</td>
<td>2.4</td>
<td>6.3</td>
<td>1.9</td>
<td>2.1</td>
<td>2.4</td>
<td>6.4</td>
<td>1.9</td>
<td>2.0</td>
<td>2.4</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance education is inappropriate for graduate education. Distance education must be discouraged in graduate education.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>443</td>
<td>1.8</td>
<td>2.1</td>
<td>2.3</td>
<td>6.2</td>
<td>1.8</td>
<td>2.0</td>
<td>2.3</td>
<td>6.1</td>
<td>1.8</td>
<td>2.0</td>
<td>2.3</td>
<td>6.1</td>
<td></td>
</tr>
</tbody>
</table>
Bibliography


Anonymous. (circa 2010). About Us Page - University of the People, from [http://www.uopeople.org/groups/tuition-free-education](http://www.uopeople.org/groups/tuition-free-education)


Cannell, L. (1999). A Review of the Literature on Distance Education. Theological Education. Retrieved from mysma.saintmeinrad.edu/faculty/ehensell/Cannell.pdf


Drucker, P. (1997). I got my degree through E-mail Forbes. As the late Peter Drucker succinctly put it, "Universities won't survive. The future is outside the traditional campus, outside the traditional classroom. Distance learning is coming on fast."


297


Smith, E. *American Criminal Law Review*.


Board of Regent's Policy, Section 5.10  Distance Learning (2002).


Weatherall, A., & Hailstones, F. (2002). *Risk identification and analysis using a group support system (GSS)*.


Wilson, B. (2001). Trends and Futures of Education: Implications for Distance Education. *Quarterly Review of Distance Education, Special Issue*(October).


