SYST 542
Decision Support Systems Engineering

Instructor: Kathryn Blackmond Laskey

Fall Semester, 2006

Unit 8: Evaluation-Centered Design

Copyright © 2006, Kathryn Blackmond Laskey

Outline

- Theories and hypotheses
- The scientific method and evaluation of hypotheses
- Threats to validity
- Test and evaluation of DSS
  - Formative evaluation / Developmental T&E
  - Summative evaluation / Operational T&E
- Evaluation-Centered Design

Copyright © 2006, Kathryn Blackmond Laskey

Theory

- **Theory**
  - Systematically organized knowledge applicable in a wide variety of circumstances; a system of assumptions, principles, and rules of procedure devised to analyze, predict or explain a set of phenomena
- **Scientific theory**
  - Always provisional
  - Must be falsifiable
  - Can never be proven

Copyright © 2006, Kathryn Blackmond Laskey

Hypothesis

- **Hypothesis**
  - Statement about relationship between two or more measurable variables
  - Precise enough to be verified or falsified
  - Usually backed by theory
    - » Verification provides support for theory
    - » Falsification calls theory into question
- **Science makes progress through iterative process:**
  - Propose theory
  - Develop hypotheses to evaluate theory
    - » Hypothesis is likely to be true if theory is correct
    - » Hypothesis is likely to be false under some plausible alternative to the theory
  - Collect and analyze evidence that bears on the hypothesis
  - Refine theory in light of evidence

Copyright © 2006, Kathryn Blackmond Laskey

Evolution of Theories

- **We view the world through our theories**
  - **The variables that describe the world**
  - **Relationship of causes to observable outcomes**
  - **Actions we can take to change the world**
- **"There is always an easy solution to every human problem -- neat, plausible and wrong." *-Mencken***
- **Theories and observations "co-evolve"**
  - **Revise theory based on observation**
  - **Use theory to tell us what to observe**
- ***Post hoc* ("after this") theorizing**
  - **Constructed to explain results after they have been observed**
  - ***Proceed with caution* until further confirming evidence has been collected**

Copyright © 2006, Kathryn Blackmond Laskey

Engineering and the Scientific Method

- Scientific method: hypothesize, test, refine
- Science pursues knowledge for the sake of knowledge
  - Seek theories that accurately describe and predict how the world works
  - Seek evidence that definitively establishes correct theories and falsifies incorrect theories
- Engineering applies the scientific method *with the purpose of solving problems in the world*
  - Use theories to make predictions relevant to the problem
  - Seek adequacy for purpose, not ultimate correctness
  - "How would a scientist think about this?" can be a useful heuristic
  - In engineering we often cannot afford scientific purity, but it is important to understand the tradeoffs we are making

Copyright © 2006, Kathryn Blackmond Laskey

Prediction, Explanation, Intervention

- Explanatory studies identify factors that explain the causes of an outcome variable
- Predictive studies make predictions about the future value of an outcome variable
  - Example: race *predicts* academic performance but we don't think race *causes* academic performance
- Good predictors do not necessarily explain!
- Evaluating interventions requires explanation
  - We want to design a DSS that causes improvements in task performance
  - We want to evaluate the DSS to determine whether there are improvements attributable to the DSS and if so what they are
  - We want to evaluate the ability of our organization's DSS development and/or procurement process to improve organizational decisions

Copyright © 2006, Kathryn Blackmond Laskey

Sources of Empirical Evidence

- **Anecdotal or informal observation**
  - Good for suggesting hypotheses (we always start this way)
  - Poor for rigorous confirmation of hypotheses
  - Good formal study will include an anecdotal component
- **Formal observation**
  - Systematic observation of natural covariation in A, B and other plausible causes of B
  - If A and B co-vary when other causes of B are held fixed, infer that A causes B
  - Causal inferences are limited by the range of natural variation in A, B, and other potential causes of B
- **Direct manipulation**
  - Manipulate A (and possibly secondary variables) directly
  - If B changes when A is manipulated, infer that A causes B
  - Experiment: assign values of A randomly
  - Quasi-experiment: assign values of A non-randomly in naturalistic setting

Copyright © 2006, Kathryn Blackmond Laskey

Sources of Empirical Evidence

- **Opinion**
  - Survey opinions of experts, users, customers
  - Evidence is as good as knowledge of those surveyed
- **Modeling**
  - Execute simulation of system
  - Manipulate A and see whether B occurs
- **Related system**
  - Test feasibility of technical approach by developing a prototype
  - Evaluate proposed DSS by examining similarities to and differences from related DSS
- **Combinations**
  - Test real components combined with simulated components
  - Use expert opinion to adjust data from related system

Copyright © 2006, Kathryn Blackmond Laskey

Example

- **Hypothesis: Introducing DSS will cause productivity increase**
- **Anecdotal evidence:**
  - **Interview some DSS users.**
  - ***Result: Most report that the system helps them to do their job more effectively and efficiently.***
- **Observational evidence:**
  - **Collect data on productivity metrics and other variables from a sample of employees, some of whom are using DSS and some whom are not.**
  - ***Result: DSS users tend to be more productive.***
- **Direct manipulation (quasi-experiment):**
  - **Introduce DSS into organization. Collect before and after productivity data, as well as data on other variables.**
  - ***Result: productivity increases after DSS is introduced.***
- **Direct manipulation (controlled, randomized experiment):**
  - **Identify pairs of employees matched on characteristics thought to affect productivity. Give DSS to randomly chosen employee from each pair. Measure productivity & other variables before and after system is introduced.**
  - ***Result: engineers using DSS increased productivity, others did not.***

Copyright © 2006, Kathryn Blackmond Laskey

Hypotheses in DSS Evaluation

- What are the objectives of DSS design?
  - Should be identified as part of defining the operational concept
  - Examples of objectives:
    - » DSS users accomplish a given task (or part of task) more quickly / more accurately / in fewer steps … than unaided users
    - » New workers using DSS learn to perform task faster than new workers without DSS
    - » DSS lifecycle cost is justified by productivity improvements
    - » DSS enhances productive workflow
    - » DSS users find system natural to use
- Hypotheses for evaluation should relate to objectives
  - Is objective achieved? To what degree?
  - What are reasons objective is / is not achieved?

Copyright © 2006, Kathryn Blackmond Laskey

Variables and Hypotheses

- Dependent variable - measures the presumed effect
- Independent variable - measures the presumed cause
  - Manipulated in intervention studies
  - Measured but not manipulated in observational studies
- Auxiliary variables - neither causes nor effects (sometimes called nuisance variables)
  - may be controlled (explicitly considered in design and modeled in analysis)
  - when uncontrolled, may be confounded with explanatory (dependent and independent) variables

Copyright © 2006, Kathryn Blackmond Laskey

Validity

- An empirical study is *valid* if it provides definitive results for or against the hypotheses it was designed to test
- Internal validity
  - Can the result be attributed to the explanatory variables the researcher claims are operating?
  - Can we rule out confounding variables as explanation?
- External validity
  - Can the result be generalized from the sample to the sampled population?
  - Can the result be generalized to some other population?
- Including more factors tends to
  - Increase external validity
  - Decrease internal validity

Copyright © 2006, Kathryn Blackmond Laskey

Threats to Validity

- An empirical study is designed to provide evidence for or against a given hypothesis
- *Threats to validity* are reasons the results may not mean what the study designer intended them to mean
- Careful study design requires:
  - Anticipating threats to validity
  - Controlling for effects of threats to validity

Copyright © 2006, Kathryn Blackmond Laskey

Example Threat to Validity

- **Experimental test of prototype decision support system**
  - 200 subjects were tested on several sample scenarios
  - 100 aided subjects and 100 unaided subjects
  - Subjects were scored against "ground truth" answers
- **Results of experiment:**
  - Aided subjects averaged 63 out of 100
  - Unaided subjects averaged 64 out of 100
  - Client decided not to fund full-scale implementation
- **Was this decision justified?**

Copyright © 2006, Kathryn Blackmond Laskey

Results by Level of User Experience

	Expert Users	Inexperienced Users	Overall
Unaided	70	40	64
	(n=80)	(n=20)	(n=100)
Aided	90	60	63
	(n=10)	(n=90)	(n=100)

- System produced sizeable improvement for both subgroups
- Experiment and analysis appeared to support the opposite conclusion!
- The reason: selection effect
  - Groups were unbalanced on a characteristic related to performance
  - Analysis failed to control for the imbalance
- Selection effects are common with nonrandom assignment (e.g., inexperienced users are younger and more willing to try new technology)

This phenomenon is called Simpson's Paradox

Copyright © 2006, Kathryn Blackmond Laskey

Control

- Isolating the effect of variables of interest may be difficult
  - Dependent variables measure phenomenon of interest
  - Independent variables measure hypothesized causes
  - Extraneous variables introduce additional variation and interfere with ability to explain
- Control means incorporating a variable into the design in a way that its impact on conclusions can be assessed scientifically
- Uncontrolled variables may be confounded with explanatory variables
  - Can't tell variation due to explanatory variables from variation due to confounding variables
  - Confounding variables threaten validity

Copyright © 2006, Kathryn Blackmond Laskey

Controlling for Confounding Variables

- **Eliminate variables**
  - Set value to constant -- e.g., study only males or only a single geographic region
  - Limits ability to generalize
  - Increases power of statistical tests
- **Include variables as factors in design**
  - Statistical control -- adjust results to account for included factors
  - Including too many factors leads to design with low statistical power
- ***Hierarchical Bayesian inference* can include more variables than traditional statistical methods with less loss of statistical power**
  - But there are no free lunches

Copyright © 2006, Kathryn Blackmond Laskey

Manipulation

- **Experiment manipulates independent variable using randomization**
  - Controls statistically for variables not included in design
- **Quasi-experiment manipulates the independent variable but does not randomize**
  - Confounding can be a problem
  - Validity of statistical control depends on *ignorability* of assignment of values of independent variable
- **Observational study examines variation in dependent variable under natural variation in independent variable**
  - There are many sources of confounding in observational studies
  - Statistical control is often used, but ignorability of assignment is often suspect
  - We may need to explicit model for assignment of independent variable

Copyright © 2006, Kathryn Blackmond Laskey

A Note on Randomization

- **Uses of randomization**
  - Obtain samples representative of a population
  - Control for variables not included in design
- **Balance (or representativeness) is more important than randomization *per se***
  - restricted randomization
  - post-hoc adjustment for unbalanced samples
- ***Ignorability***
  - A design is ignorable if inferences do not depend on how the sample was obtained
  - Randomization is a special case of ignorability
  - Sometimes we can achieve ignorability even when it is infeasible to randomize
  - Concerns with quasi-experiments relate to deviations from ignorability

Copyright © 2006, Kathryn Blackmond Laskey

Controlling for Selection Effects

- **Random assignment**
  - Not a cure-all: random assignment can be unbalanced by luck!
  - Not always possible
- **Matching**
  - Match units on important characteristics thought to be related to outcome of interest
  - One from each matched pair assigned to each treatment group
  - Can be used with random assignment
- **After-the-fact adjustment**
  - Measure characteristics thought to be related to outcome of interest
  - Model relationship between outcome and characteristics
  - Reduces statistical power (sometimes very much because you are estimating many parameters)

Copyright © 2006, Kathryn Blackmond Laskey

Threats to Internal Validity

- *History* - unmeasured things that happened prior to study
- *Maturation* - subjects change during study
- *Testing* - subjects learn after multiple measures are taken
- *Instrumentation* - problems with measuring (nonequivalent measures, researcher learning)
- *Regression toward the mean* - extreme values of a measure tend not to repeat
- *Selection* - how individuals are assigned to treatments
- *Mortality* - subjects drop out of study
- *Diffusion or imitation of treatments* - those not in treatment group get similar treatment
- *Emotional effects* - compensatory rivalry, resentful demoralization

Copyright © 2006, Kathryn Blackmond Laskey

Threats to External Validity

- Interactions between treatments and variables not included in study
- Interaction between treatments (effect of earlier treatments on later outcome; “research-wise” subjects)
- Pretest and posttest sensitization - pretest affects subjects; effect may be due to test not treatment
- Changes in the world that affect validity of study (this can be viewed as an example of treatment/variable interaction)
- System as used in the real world is not the same as system as implemented in test environment
- Subjects tested in the evaluation study are not typical of operational users (“golden crew”)

Copyright © 2006, Kathryn Blackmond Laskey

Researcher Bias

- **Expectations and attitudes may affect outcome and interpretation of results**
  - Subjects studied
  - What is measured
  - Subtle effects in measurement
  - Exclusion of "outliers" or "problem cases"
  - Computational errors in direction of expectations
  - Reporting only analyses / visual displays that confirm researcher's point of view
  - <u>These biases may be entirely unconscious!</u>
- **Protecting against researcher bias**
  - Maintain multiple hypotheses
  - Be aware of potential biases
  - Involve impartial parties

Copyright © 2006, Kathryn Blackmond Laskey

Iterative Evaluation Centered DSS Design

1. **Define what is meant by "quality of decision"**
2. **Develop measures of decision quality**
3. **Develop hypotheses of factors contributing to decision quality**
   - characteristics of users
   - characteristics of tasks
   - characteristics of environment
   - characteristics of decision making process
4. **Develop explanation of status-quo decision quality**
5. **Generalize to theory which predicts decision quality under alternatives including status quo**
6. **Use theory to guide design of new or enhanced DSS**
7. **Implement and evaluate theory by testing it on this case**
   - Measure decision quality <u>and</u> features of DSS related (by theory) to decision quality
8. **Refine theory in light of evaluation and iterate**
9. **Apply same approach to organization's DSS lifecycle process**

Copyright © 2006, Kathryn Blackmond Laskey

The fundamental objective for a decision support system development effort is to achieve the highest possible effectiveness and productivity for the users within reasonable cost limits

Copyright © 2006, Kathryn Blackmond Laskey

DSS Evaluation Supports DSS Lifecycle Process

- The DSS development process involves a number of decision points
  - Build prototype?
  - Move to full-scale development?
  - What level of resources to commit?
  - Add functionality?
- DSS evaluation provides information to support these decisions
- DSS evaluation also provides information to community knowledge base about DSS
- Evaluation strategy is itself a decision
  - Resources to commit to evaluation
  - What to measure
  - How to measure it

Copyright © 2006, Kathryn Blackmond Laskey

Why Evaluation is Difficult

- Costs of evolutionary development are hard to identify
- Benefits are often "soft"
  - Hard to relate to "bottom line"
  - Hard to measure
- Rigorous empirical test may be prohibitively expensive

Copyright © 2006, Kathryn Blackmond Laskey

Formative & Summative Evaluation

- Formative evaluation evaluates design as it is being developed
- Summative evaluation evaluates complete or nearly complete product

Copyright © 2006, Kathryn Blackmond Laskey

Formative Evaluation

- **Purpose:**
  - Identify weaknesses and priorities for improvement
- **Decisions supported:**
  - Should development effort continue or be terminated?
  - Is the system ready to move to the next milestone?
  - Which new features should be included in next prototype?
  - Should requirements be modified? How?
  - What problems should be fixed? In what order?
- **What to measure:**
  - Concern is less with statistical validity than with understanding strengths and weaknesses of system
  - Collect process data, user anecdotes and impressions along with summary decision quality metrics

Copyright © 2006, Kathryn Blackmond Laskey

Summative Evaluation

- **Purpose:**
  - Measure overall quality of substantially finished product
- **Decisions supported:**
  - Evaluate which of several competing products to purchase
  - Compare internally developed product with others for use in marketing
  - Evaluate quality of finished DSS
    - » provide feedback to improve development process
    - » decide whether, when & how to implement next version
    - » evaluate impact of DSS on organizational process
- **What to measure:**
  - More emphasis on valid measure of overall decision quality
  - Process data and anecdotes are useful for later product enhancements but are not the main focus

Copyright © 2006, Kathryn Blackmond Laskey

MOEs and Formative Evaluation

- **Use MOEs to identify areas for improvement and generate options**
  - **"System scored low on usability. Users thought placement of icons was inconsistent from screen to screen. An option for fixing this problem is to have users help us to redesign screens. Estimated cost: 2 person-weeks of programmer time, 1 person-week of user time."**
  - **"System did not meet time constraint on generating a schedule. We need to analyze the software and algorithms and determine whether there is a way to speed up the run time."**
- **Use aggregate measure of value to choose among clearly defined options**
  - **Continue development?**
  - **Fix this problem or that one?**
  - **Which new functionality to implement first?**

Copyright © 2006, Kathryn Blackmond Laskey

How Much Formative Evaluation?

- **Rule of thumb (Hicks & Hartson): for each significant version of an interaction design**
  - Three major cycles of formative evaluation
  - Each followed by redesign
  - Followed by summative evaluation
- **10% rule:**
  - Development effort should have something to evaluate by the time 10% of the resources are expended

Copyright © 2006, Kathryn Blackmond Laskey

Types of Measure

- **Objective**
  - Directly observed measures
  - Typically user performance using interface to perform benchmark tasks
- **Subjective**
  - Opinion / judgment of users, experts, engineers
- **Qualitative**
  - Non-numeric data and results (e.g., lists of problems users had)
  - Useful for developing hypotheses about what features of system are associated with measured usability problems
- **Quantitative**
  - Numeric data and results
  - Can be objective (timing on tasks) or subjective (percentages of different responses to questionnaire items)

Copyright © 2006, Kathryn Blackmond Laskey

Measuring Achievement of Objectives

- **Requirements for Measures of Effectiveness (MOEs):**
  - Measurable (subjectively or objectively)
  - Correlated with (some aspect of) DSS performance
- **Ways to measure**
  - Objective measurement
  - Expert observation
  - Subjective judgment
- **Identify dimensions of value first, then consider how to measure**
  - Common error: if we don't know how to measure it then it doesn't matter
- **Hierarchical decomposition**
  - Group attributes into categories
  - 7 or fewer attributes in each group
  - Highest level is fundamental objective
  - Try to measure all important dimensions of value

Copyright © 2006, Kathryn Blackmond Laskey

Types of Attributes

- **Natural attribute**

  Directly measures attribute in question. Common interpretation and generally understood. [*cost in dollars*]

- **Constructed attribute**

  Developed specifically for a given problem context; defines what is meant by attribute. Usually a numerical scale is developed to designate different impact levels. Over time some constructed attributes come to be treated as natural attributes. [*GNP*]

- **Proxy attribute**

  An indirect measure of some attribute valued for its perceived relationship to the corresponding objective. A proxy attribute for a fundamental objective may be a natural attribute for a means objective [SO_2 *emissions as proxy for "stone disfiguration"*]

Copyright © 2006, Kathryn Blackmond Laskey

Categories of Value

(Keen, 1981)

Benefit	Easy to measure?	"Bottom Line"?
1. Increase in number of alternatives examined	N	N
2. Better understanding of the business	N	N
3. Fast response to unexpected situations	Y	N
4. Ability to carry out *ad hoc* analysis	Y	N
5. New insights and learning	N	N
6. Improved communication	N	N
7. Control	N	N
8. Cost savings	Y	Y
9. Better decisions	N	N
10. More effective teamwork	N	N
11. Time savings	Y	Y
12. Making better use of data resource	Y	N

Copyright © 2006, Kathryn Blackmond Laskey

Dimensions of Evaluation

(Sage 1991)

- **Structural** - did you implement a correct instantiation of the system design?
- **Functional** - does the system produce the right results?
- **Purposive** - does the system fulfill the purpose for which it was designed?

Copyright © 2006, Kathryn Blackmond Laskey

Types of Evaluation

(Adelman, 1992)

Technical evaluation - apply technical criteria to evaluate correctness (algorithmic and systemic)

Empirical evaluation - collect objective data on DSS performance

Subjective evaluation - evaluate DSS from perspective of potential users

Copyright © 2006, Kathryn Blackmond Laskey

Example MOEs for Each Type of Evaluation

Technical

- errors identified in code walk-through
- expert evaluation of design; method/task match
- results of running modules on benchmark cases

Empirical (field observation, quasi-experiment, experiment)

- time to decision
- quality of decision (expert judgment or objective measure)
- covariates: experience (problem/computer) ; problem type; aided/unaided

Subjective

- questionnaires
- interviews
- Delphi or other group facilitation process
- field observation

Copyright © 2006, Kathryn Blackmond Laskey

Interfaces to be Evaluated

- **DSS <--> User**
  - **Match between DSS and user** (background, workstyle, operational needs)
  - **Adequacy of DSS** (ease of use, adequacy of features, …)
- **DSS + User <--> Organization**
  - **Organizational efficiency**
  - **Fit into organization**
- **Organization <--> Environment**
  - **Quality of decisions**
  - **Match of technical approach to task requirements**
  - **Quality of decision making process**

Copyright © 2006, Kathryn Blackmond Laskey

Organizing Data Collection

- **Test plan should include:**
  - Measure to be taken
  - Requirement(s) to which it is relevant
  - Expected outcomes categorized as successful / acceptable / problematic / unacceptable
- **Test log should include:**
  - Measure
  - Relevant requirement(s)
  - Categorization of outcome as successful / acceptable / problematic / unacceptable
  - Action to be taken as result of test
  - Priority
  - Status (updated as solutions are implemented)
- **Test report should summarize this information in a form that supports decision making**

Copyright © 2006, Kathryn Blackmond Laskey

Experimental Design

- **Usually many factors may influence the value of a dependent variable**
  - Characteristics of users
  - Characteristics of the problem
  - Characteristics of the experimental set-up
- **Goal: design an experiment to estimate *main effects* of factors and *interactions b*etween factors**
  - An interaction means the effect of a factor depends on the values of other factors
  - Larger sample sizes are needed to estimate interactions than main effects
- **Experimental design lays out a plan for how subjects will be assigned to experimental conditions**
  - Goal: sufficient statistical power to achieve valid result

(references: Campbell, 1986; Montgomery, 2004)

Copyright © 2006, Kathryn Blackmond Laskey

Factorial Design

- **Factorial design with *k*-fold replication - each combination of dependent variables is covered *k* times**
  - Allows estimation of main effects and all interactions
  - Required sample sizes are often prohibitive
  - 5 factors, 2 levels each, 2 replications: 2^5x2 = 64 runs are needed
- **Fractional factorial design - fewer runs than full factorial design**
  - Allows estimation of main effects and some interactions (depending on the design)

Copyright © 2006, Kathryn Blackmond Laskey

Developing MOEs for Successful Evaluation

- Effective feedback and redesign requires *sensitive*, *discriminating* and *convincing* measures of major dimensions of value
- <u>Sensitive</u> - measure will identify good quality result:

  P(M = HIGH I Q = HIGH) = LARGE (small probability of "miss")
- <u>Discriminating</u> - can rule out other explanations for good result:

  P(M = HIGH I Q = LOW) = SMALL (small probability of "false alarm")
- <u>Convincing</u> - relevant community will agree to the above statements

Copyright © 2006, Kathryn Blackmond Laskey

Making Do With Weak Measures

- **Measure multiple dimensions of DSS quality**
- **Tradeoffs**
  - For a given measure, decreasing miss rate will increase false alarm rate
  - Improving both miss & false alarm rate usually increases cost and may be infeasible
- **Use multiple measures**
  - Measures may be unconvincing alone, but persuasive when all agree
  - Combining multiple *independent* measures can improve miss/false alarm rates
  - Look for measures with no common source of bias

Copyright © 2006, Kathryn Blackmond Laskey

Dimensions of Testing Environment

- Fidelity of environment
- Fidelity of organization
- Representativeness of problems
- Cost
- Degree of control

Copyright © 2006, Kathryn Blackmond Laskey

Installation and Beyond

- Field testing is an important component of evolutionary design
  - Identify priorities for enhancements and new versions
  - Learn what worked and what didn't
  - Apply lessons learned to future DSS development efforts
- Consider designing DSS to support data capture for field testing

Copyright © 2006, Kathryn Blackmond Laskey

Summary: DSS Development and Evaluation

- Evolutionary development is usually necessary
  - dimensions of evolution
  - managing evolution
- Evolution involves cycles of formative evaluation
  - evaluate on multiple dimensions
  - identify & prioritize areas for improvement
  - periodically make go / no-go decisions
- Summative evaluation at end of project helps incrementally improve DSS development capability
- Field testing can help to prioritize enhancements and provide "lessons learned" to improve organization's DSS development process

Copyright © 2006, Kathryn Blackmond Laskey

In Summary...

Copyright © 2006, Kathryn Blackmond Laskey

References for Unit 8

Adelman, L. *Evaluating Decision Support Systems*, Wiley, 1994.

Boar, B. *Application Prototyping: A Requirements Definition Strategy for the 80s.* New York: Wiley Interscience, 1984.

Boehm, B.W. A Spiral Model of Software Development and Enhancement. IEEE Computer, 21(5), 61-72, 1988.

Boehm, B.W., Gray, T.E. and Seewaldt, T. Prototyping vs. Specification: A Multi-Project Experiment. Proceedings of Seventh International Conference on Software Engineering, New York: ACM and IEEE, 473-484, 1984.

Campbell, D. T. and Stanley, J.C. (1986) *Experimental and Quasi-Experimental Design for Research*, Rand McNally, Chicago.

Hicks, D. and Hartson, H.R. Developing User Interfaces: Ensuring Usability Through Product and Process. Wiley, 1993.

Keen, P.G.W., Value Analysis: Justifying Decision Support Systems, MIS Quarterly, 6(1), 1981.

Keeney, R., Value-Focused Thinking, Cambridge, MA: Harvard University Press1992.

Montgomery, D., *Design and Analysis of Experiments*, Wiley, 2004.

Rouse, W.B. Design and evaluation of computer-based decision support systems. In S.J. Andriole, ed., *Microcomputer Decision Support Systems*. Wellesley, MA: QED Information Sciences Inc., 1985.

Sage, A. Decision Support Systems Engineering, New York: Wiley, 1991.

Copyright © 2006, Kathryn Blackmond Laskey