CIFE CENTER FOR INTEGRATED FACILITY ENGINEERING

Design Exploration Assessment Methodology: Testing the Guidance of Design Processes

By

Caroline Clevenger, John Haymaker and Andrew Ehrich

CIFE Technical Report #TR192
June 2010

STANFORD UNIVERSITY

If you would like to contact the authors, please write to:

c/o CIFE, Civil and Environmental Engineering Dept.,
Stanford University
The Jerry Yang & Akiko Yamazaki Environment & Energy Building
473 Via Ortega, Room 292, Mail Code: 4020
Stanford, CA 94305-4020

Design Exploration Assessment Methodology: Testing the Guidance of Design Processes

This paper introduces the Design Exploration Assessment Methodology (DEAM) for comparing the impact of differences in process on outcome for given design problems. Current practice fails to reliably generate high performing alternatives in part because it lacks systematic means to compare existing or emerging design processes. Researchers lack empirical methods and data to evaluate design challenges and the strategies available to address them. In this paper we document, and then apply DEAM to professional implementation of six design strategies across two design challenges using the charrette test method. Results compare strategies according to the performance of the solution(s) generated. For the strategies and challenges investigated, more information during design does not always assist the designer to produce better performing alternatives. We discuss possible explanations, and conclude with a discussion of the strengths and weaknesses of DEAM as an evaluation method. Initial findings demonstrate that DEAM is a method capable of providing meaningful comparison of strategies in the domain of energy efficient design challenges.

Keywords: design theory, strategy, challenge, exploration, guidance, creativity

Terms:

Components

Variable: a design choice to be made. A *variable* can be discreet (i.e., number of windows) or continuous (i.e., building length).

Option: individual *variable* input(s) (i.e., number of windows = {1, 2, or 3}; building length = 10-20 meters).

Decision: the selection of an *option* (i.e., a number of windows = 2; building length = 12.75 meters).

Alternative: a combination of *decisions* about *options*.

Stakeholder: a party with a stake in the selection of *alternatives*.

Goal: declaration of intended properties of *alternatives*.

Preference: weight assigned to a *goal* by a *stakeholder*.

Constraint: limit placed on *options*.

Impact: *alternative's* estimated performance according to a specified *goal*.

Requirement: limit placed on *impacts.*

Objective: union of *stakeholders*, *goals*, *preferences* and *constraints*.

Value: net performance of an *alternative* relative to all *objectives*.

Dimensions

Challenge: a set of *decisions* to be made ranging from simple to complex.

Strategy: a procedure to generate *decisions* ranging from none to advanced.

Exploration: a history of *decisions* made ranging from misled to guided.

Design Process: implementation of a *strategy* to a *challenge* resulting in an *exploration*.

Guidance: variation in *exploration* produced by applying different *strategies* to a given *challenge*.

Spaces

Objective space: set of *stakeholders, goals, preferences* and *constraints.*

Alternative space: feasible (explored or unexplored) *alternative*s for a given *challenge*.

Impact space: analyzed *impacts* of *alternative*s relative to *goals*, determined to be acceptable or unacceptable according to *requirements*.
Value space: *values* of the set of *alternatives* generated during an *exploration*.

We use italics throughout this paper to indicate explicit reference to these definitions.

Introduction

Performance-based design consists of *explorations* supported by *strategies* to generate and analyze *alternatives* that address *challenges* with explicit *objectives*. *Strategies* used by designers range from none to advanced. As new *strategies* emerge, designers lack a method to assess the *guidance* provided. To assess *guidance* designers need to compare the *explorations* afforded by available *strategies* on potential *challenges*. This paper asks: what is a method for evaluating how well *strategies* address *challenges*? Metrics and a framework relating these process components are important for process improvement (Dorst, 2008). Clevenger and Haymaker, 2011 proposes a framework to describe design as *exploration* through *objective, alternative, impact,* and *value spaces*. It proposes the following metrics, categorized as to whether they describe *challenge*, *strategy* or *exploration*:

Challenge Metrics

Objective space size (OSS): the number of *objectives* considered in the challenge.
Alternative space interdependence (ASI) - the number of first order interactions among *variables* divided by total number of *variable* combinations. ASI represents the extent to which interactive affects impact *value.* In the synthetic experiment performed for this research, it is calculated using built-in capabilities of existing process integration design optimization (PIDO) software. In general, the higher the ASI is, the more complex the *challenge*.
Impact space complexity (ISC): the number of *variables* found to result in performance trade-offs (divergent *impacts*) divided by total number of *variables*. ISC represents the percentage of *variables* with competing *objectives*. In the synthetic experiment performed for this research, ISC is observable using built-in capabilities of existing PIDO software. The higher the ISC is, the more complex the *challenge*.
Value space dominance (VSD): the extent to which *value* is dominated by individual *variables* calculated using sensitivity analyses. VSD represents the importance of individual design *decisions*. In the synthetic experiment performed for this research, it is calculated using built-in capabilities of existing PIDO software. Because the lower the VSD, the more complex the

challenge, VSD is presented as its reciprocal (1-importance) in our metric assessment.

Strategy Metrics

Objective space quality (OSQ): a scalar number (0 to 1) that measures the extent to which the *objectives* analyzed using a particular *strategy* match the *objectives* proposed for a *challenge*.

Alternative space sampling (ASS): the number of *alternatives* generated divided by the number of *alternatives* required for a “significant sampling” of all feasible *alternatives*. It measures the extent to which a sampling is “representative” of all *alternatives*. Significant sampling can be determined mathematically using standard statistical techniques to calculate “sample size.” For comparative purposes when the statistically significant sample size is unknown, we use the total population of *alternatives*.

Alternative *space flexibility* (ASF): the average number of *option* changes between any two *alternatives* divided by the number of *variables* modeled. ASF indicates the variety in *alternatives* generated in a given *exploration*.

Exploration Metrics

Value *space average* (VSA): the mean *value* for the set of *alternatives* analyzed. This metric characterizes the average performance of *alternatives* generated in an *exploration*.

Value *space range* (VSR): the standard deviation in *value* for the set of *alternatives* analyzed. This metric characterizes the dispersion of *alternatives* generated in an *exploration*.

Value *space iterations* (VSI): the number of *alternatives* generated before the highest *value* is reached. This metrics characterizes the efficiency of an *exploration*.

Value *space maximum* (VSM): the top *value* calculated for *alternatives* generated in a given *exploration*. This metric characterizes the maximum *value* generated.

Existing literature provides metrics to evaluate individual *design process* dimensions. For example, signal-to-noise ratios have been used to evaluate the robustness of a design *challenge* (Phadke & Taguchi, 1987). Flexibility, robustness, and survivability have been used to evaluate design *strategy* (McManus et al., 2007). Design knowledge and design freedom have been used to evaluate the flexibility of a design *exploration* (Simpson et al., 1996). Our research focuses on collecting data to compare the *guidance* provided by a given *strategy* relative to a specific *challenge*. Research exists that compares *strategies* (for example, Avigad & Moshaiov, 2009), and significant and detailed research exists examining the efficiency and convergence properties automated algorithms. For example, (Marler & Arora, 2004) provide a comprehensive survey of *strategies* that use continuous nonlinear multi-objective optimization. They classify these *strategies* according to: ones that involve *a priori* articulation of preferences, ones that involve *a posteriori* articulation of preferences, and ones that involve no articulation of preferences. *Strategies* involving a progressive articulation of preferences are not included. They conclude the effectiveness of a *strategy* depends on the type of information provided in the *challenge*, user preferences, solution requirements, and software availability.

Research also exists which begins to evaluate the role of the designer on outcome in terms of the expertise or creativity as supported by various *strategies* for a given *challenge* (Dorst & Cross, 2001; Cross, 2004). In general, limited real-world data exists to evaluate *exploration* performance achieved by actual designers. This is due, primarily, to the fact that parallel or redundant *explorations* are not performed across *strategies* in the real-world due to limited project resources. Arguably, designers often execute similar strategies across *challenges* over time. In the absence of inter- or intra- project baselines, however, it is difficult to compare the effectiveness of various applications of the same *strategy*.

The methodology defined in this paper enables quantitative and objective assessment of the *guidance* afforded in human-performed or automated *explorations.* We use data from a laboratory experiment to assess the *guidance* that results from six design *strategies* across two design *challenges* to provide initial evidence for the method's power and generality. The domain of the *challenges* addressed is energy efficiency, although other *strategies* and *challenges* could be similarly tested across additional *explorations*.

Design Exploration Assessment Methodology (DEAM)

Using the defined terms and metrics we construct a Design Exploration Assessment Methodology (DEAM) to measure and compare the *guidance* afforded by distinct *strategies* applied to unique *challenges*. DEAM consists of the following six steps:

(1) Generate *value space*: Generate representative *alternatives* and assess multidisciplinary *impacts* and *values* for a given *challenge*.
(2) Assess *challenge*: apply *objective space size* (OSS), *alternative space interdependence* (ASI), *impact space complexity* (ISC), *value space dominance* (VSD) metrics to assess the level of *challenge* presented by the representative *alternatives*.
(3) Assess *strategies*: apply *objective space quality* (OSQ), *alternative space sampling*, (ASS), *alternative space flexibility* (ASF) metrics to the *objectives, options* and *alternatives* considered by candidate *strategies*.
(4) Conduct *explorations*: record the set and sequence of designer generated *alternatives* using distinct *strategies*. Assess *value* through analysis, results of which may or not be apparent to designer depending on the *strategy* implemented.
(5) Assess *explorations*: apply *value space average* (VSA), *value space range* (VSR), *value space* iterations (VSI), *value space maximum* (VSM) metrics to the *value space* generated in an *exploration*.
(6) Evaluate *guidance*: compare the *challenge*, *strategy* and *exploration* to deduce levels of *guidance* afforded by candidate *strategies*.

DEAM applied in synthetic experiment

We applied DEAM in a synthetic experiment using a charrette test to document the *explorations* performed by professional designers on two *challenges*, using six *strategies*. Two *challenges* were used to bring generality to the data. We acknowledge inherent differences exist between any two *challenges*, and we use *challenge* metrics to evaluate differences and allow for normalized comparison. We also used distinct *challenges* during the charrette test to discourage participant "learning" across *strategies* since different *strategies* are applied to the same *challenge*. While learning is a natural and intended consequence within any design *exploration,* in our experiment we changed *challenges* after the implementation of two *strategies* in an attempt to keep results representative of *strategy* rather than dependent (improved by) consecutive *explorations*. Figure 1 illustrates our application of DEAM.

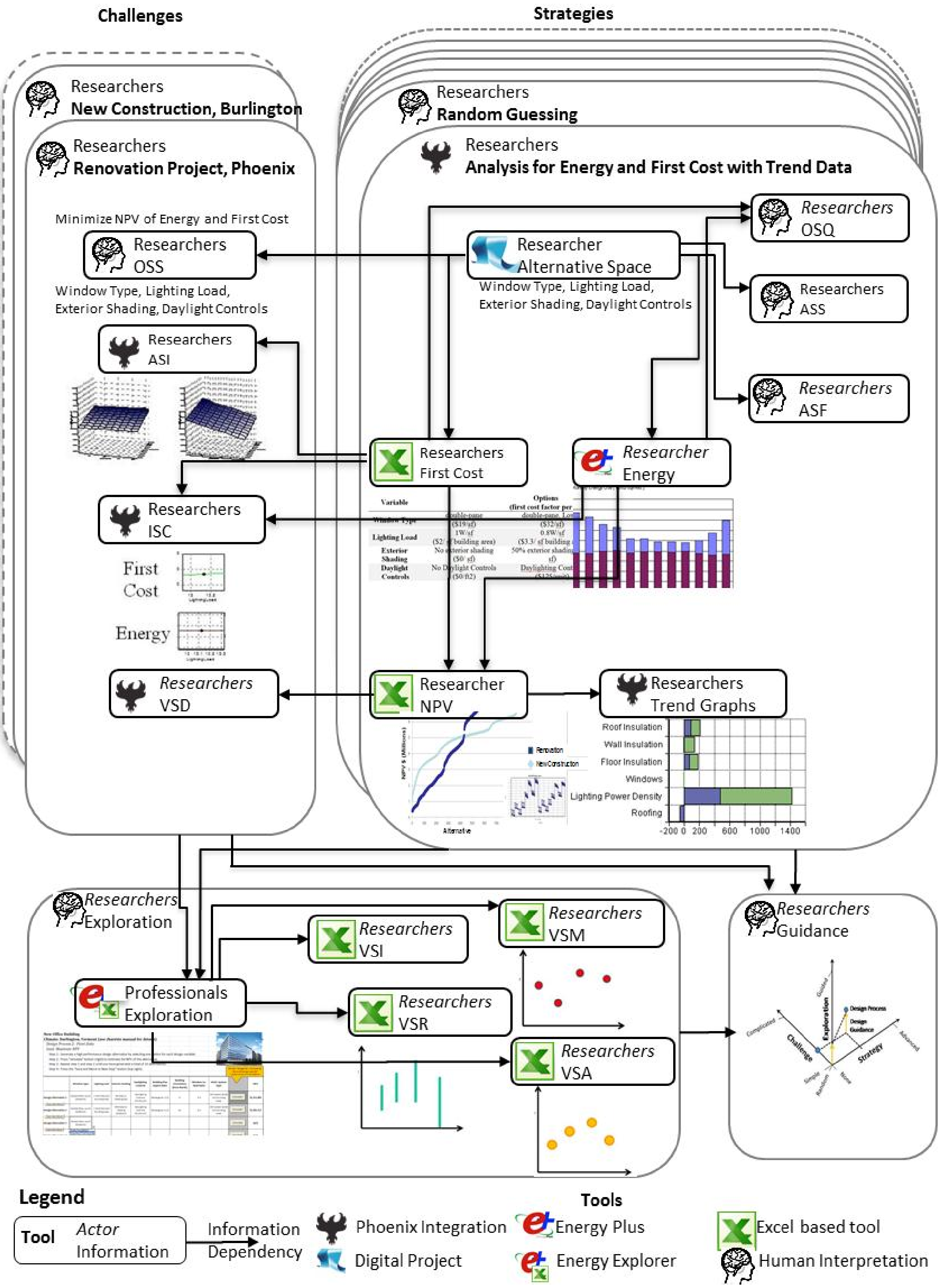

Figure 1. Process map showing how in this paper we use the Design Exploration Assessment Methodology (DEAM) to assess the *guidance* afforded from six *strategies* across two *challenges*.

In our synthetic experiment, we applied DEAM in the following manner.

(1) Generate *value space*: we used Phoenix Integration PIDO Software (Phoenix Integration, 2004) to automate input generation and analysis of EnergyPlus (LBNL, 2008) and to perform a full analysis of two simple building models, representing a new construction and a renovation project *challenge*. Full analysis provided first cost and annual energy cost estimates. We assessed *value* in units of Net Present Value ($) for all feasible *alternatives*, a total of 864 and 576 respectively. The *value spaces* for each *challenge* are shown in Figure 10.
(2) Assess *challenge*: we applied our *challenge* metrics to the results of our full analysis for both *challenges*. Results from applying metrics are presented in Table 3.
(3) Conduct *exploration*: we collected charrette test data using our custom software, EnergyExplorer™, to assist and document professional design *explorations*. EnergyExplorer™ is an interface that enables a user to easily generate *alternatives* using *strategies* different types of access to the previously simulated *value space*.
(4) Assess *strategy*: we applied *strategy* metrics to six *strategies* implemented in the experiment. The six *strategies* evaluated include: random guessing, tacit knowledge, point-based analysis, trend-based analysis, trend and point-based analysis, and full analysis. Results from applying metrics are presented in Table 4. Process diagrams of the *strategies* are provided in Figures 4-9.
(5) Assess *exploration*: we applied *exploration* metrics to charrette test results for the six *strategies* applied to two *challenges*. Results from applying metrics are presented in Table 5.
(6) Evaluate *guidance*: we compared *explorations* and *guidance* afforded by various combinations of *strategies* and *challenges*. Our comparison is shown in Figure 13.

While the order of steps 3 and 4 appears counter-intuitive, our evaluation method tests *strategies* that incorporate human decision (e.g.; tacit knowledge) rather than prescriptive algorithms alone (see Marler & Arora, 2004). In order to assess non-prescriptive *strategies,* it is necessary to first observe decisions made. While prescriptive or fully automated *strategies* such as optimization or full analysis are *strategies* that can be assessed independently of resultant *explorations*, we assess all *strategies* post *exploration* for consistency.

Next we outline further detail of the individual *strategies* tested in our synthetic experiment and their process maps.

Design Strategy 1: Random Guessing - Random algorithm generates *alternatives*.

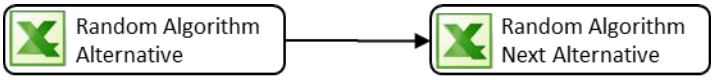

Figure 2. Random Guessing *strategy* process.

Design Strategy 2: No NPV Data - No NPV information is provided. Participants generate *alternatives* using intuition and tacit knowledge.

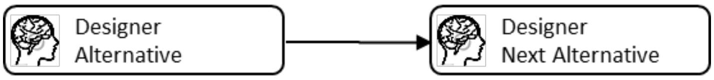

Figure 3. Tacit Knowledge *strategy* process.

Design Strategy 3: Point NPV Data - participants generate an *alternative* and then "simulate" NPV performance. Instant feedback regarding NPV of the generated *alternative* is provided and, presumably, assists in selection of the next *alternative*.

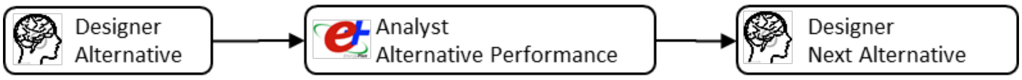

Figure 4. Point NPV Data *strategy* process.

Design Strategy 4: Trend NPV Data - Prior to *exploration*, participants are given first cost and lifecycle energy cost *impact* and NPV trend data illustrating dominance and interactive effects among *variables* and trade-offs among *impacts*. Participants were not given instruction on how to interpret this information.

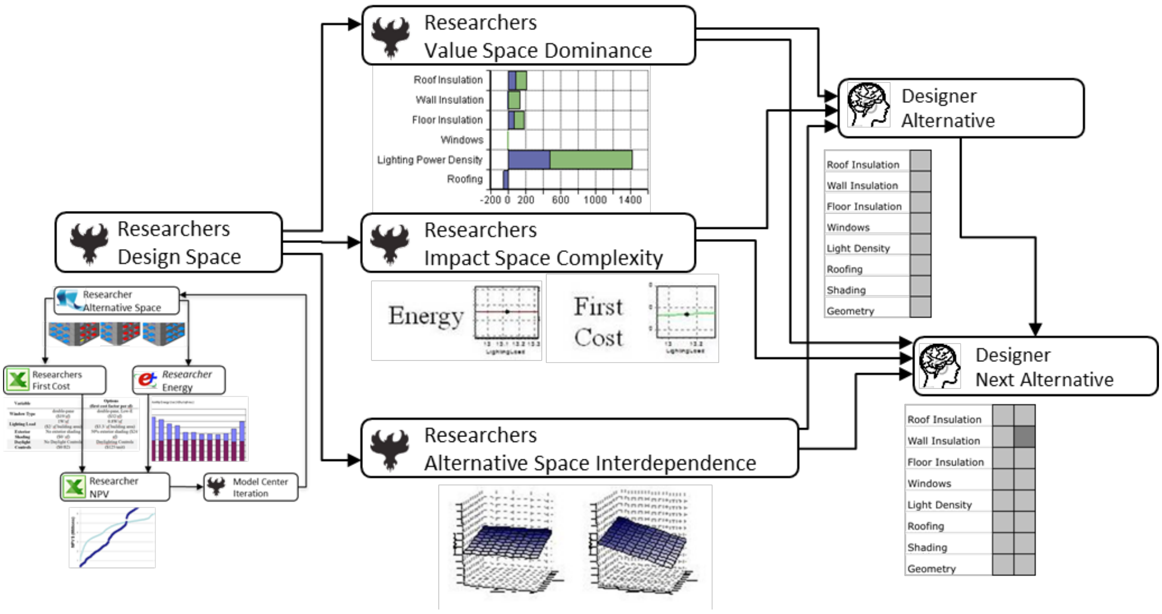

Figure 6. NPV Trend Data *strategy* process.

Figure 7 shows examples of the trend data representations provided to participants.

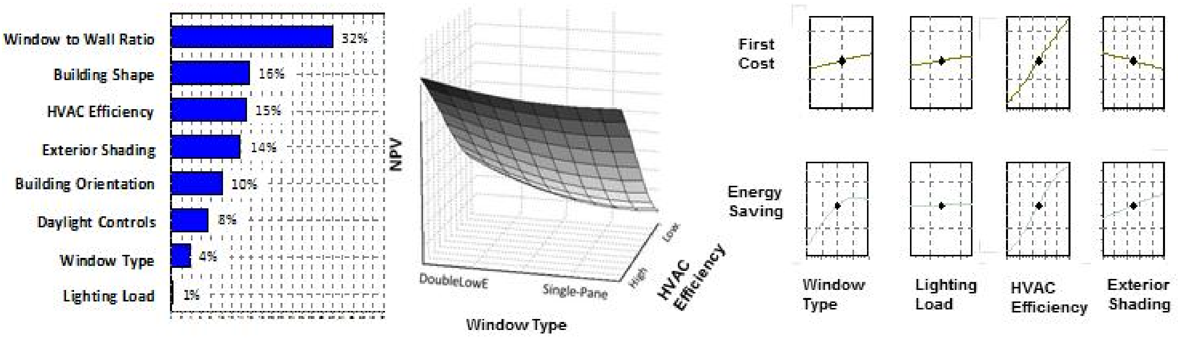

Figure 7. Sample trend data for the new construction *challenge* provided to participants implementing the NPV Trend Data *strategy*. The bar chart on the left represents the extent to which a *decision* affects an *alternative's value* (either positively or negatively). The graph in the center shows the relationship of one *decision* to another. The diagrams on the right show the *impact* of individual *decisions* on individual *goals* (ie.; first cost and energy savings).

Design Strategy 5: Trend + Point NPV Data - Prior to *exploration*, participants are given the same information provided in the Trend NPV Data *strategy*. Armed with this trend information, participants generate an *alternative* and, in addition, "simulate" NPV performance. Instant feedback regarding NPV is provided to participants after each *alternative* is generated to supplement the trend data alone.

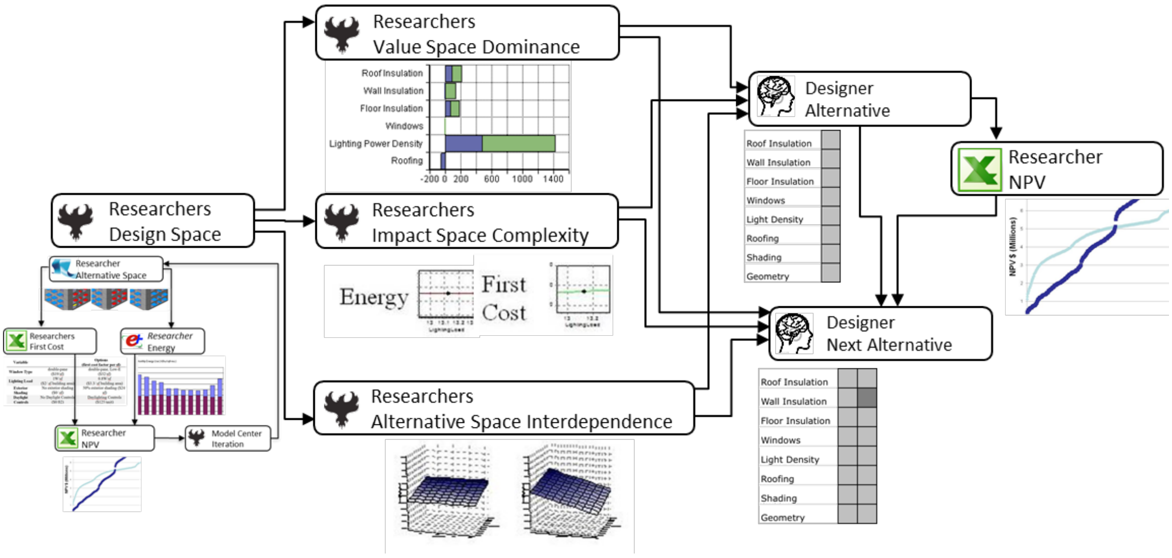

Figure 8. Trend NPV Data and Point NPV Data *strategy* process.

Design Strategy 6: Full Analysis - a full Design of Experiment (DoE) (Box et al, 2005) analysis of the *value* (i.e.; NPV) of all *alternatives* in the *alternative space* are generated. *Exploration* based on this *strategy* simply consists of selecting an *alternative* with maximum NPV since the space is fully explored.

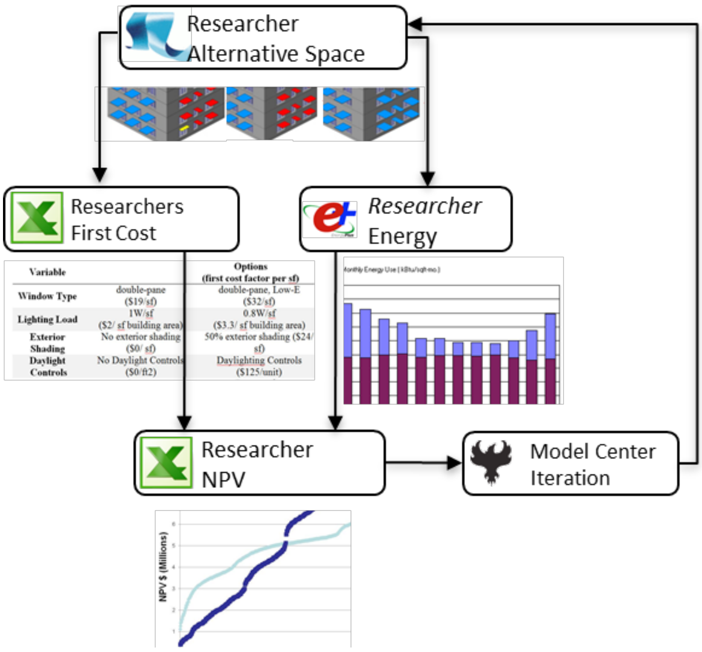

Figure 9. Full Analysis *strategy* process.

Next we discuss the application of the six steps of DEAM in our synthetic experiment:

Value Space Generated

The first step of DEAM is to generate a *value space*. We built a DoE using Phoenix Integration PIDO software and EnergyPlus energy modeling software (Welle & Haymaker, 2011, Flager et al., 2009)). We applied this full analysis to two distinct *challenges*. The *objective* for both *challenges* is to maximize the Net Present Value (NPV) of decisions regarding energy efficiency. The first *challenge* simulated the renovation of a 3 story, 100,000 sf, rectilinear office building located in Phoenix, Arizona. We modeled eight *variables* representing design *decisions* typical to an energy efficiency upgrade. Table 1 lists the *options* for the *variables* modeled in the renovation *challenge* with associated first cost implications. The second *challenge* simulated the design of a new 3 story, 100,000 sf, rectilinear office building located in Burlington, Vermont. We modeled eight *variables* representing typical design *decisions* that have effects on energy performance in new construction. Table 2 lists the *options* for the *variables* modeled in the new construction *challenge* with associated first cost estimates. The main difference between the renovation and new construction *challenges* is the inclusion of geometric *variables* in the new construction *challenge*. We model all *variables* as discrete options. However, the *variables* modeled and the number of *options* for these *variables* differed between the two *challenges*. After presenting the *variables* and *options* considered, we will use *challenge* metrics to characterize how the two *challenges* differ to assess the impact of these differences on the *guidance* provided.

Table 1. *Variables* and *options* modeled for Renovation Project, Phoenix Arizona.

Variable	Existing Condition (baseline cost)	Options (cost delta)	
Window Type	single-pane ($0)	double-pane, Low-E ($32/sf)	Argon filled, Low-E ($39/ sf)
Lighting Load	1.2 W/sf ($0/sf of building area)	1 W/sf ($2/sf of building area)	0.8W/sf ($3.3/sf)
Exterior Shading	No exterior shading ($0/sf)	50% exterior shading ($24/sf)	
Daylight Controls	No Daylight Controls ($0/sf)	Daylight Controls ($125/unit)	
Roof Type	Uninsulated Concrete Roof ($0/sf)	2" Rigid insulation added to Concrete Roof ($1.2/sf)	
Interior Office Equipment	5W/sf ($0/sf of building area)	2W/sf ($2/sf of building area)	
Wall Insulation	R-11 Insulation ($0/sf)	R-19 Insulation ($.5/sf)	
HVAC Efficiency	Existing VAV System ($0/sf of building area)	High Efficiency VAV ($/sf)[1]	

1. If the HVAC system is upgraded, the size (and resulting cost) of the system depends on other *options*. For the existing system the size is fixed and independent of other *options*.

Table 2. *Variables* and *options* modeled for New Construction Project, Burlington Vermont.

Variable	Options (first cost factor per sf)		
Window Type	double-pane ($19/sf)	double-pane, Low-E ($32/sf)	double-pane, low-e, argon filled ($39/ sf)
Lighting Load	1W/sf ($2/sf of building area)	0.8W/sf ($3.3/sf of building area)	
Exterior Shading	No exterior shading ($0/sf)	50% exterior shading ($24/sf)	
Daylight Controls	No Daylight Controls ($0/sf)	Daylight Controls ($500/floor)	
Building Shape	Square [1:1 aspect ratio] ($0/sf)	Rectangular [1:2 aspect ratio] ($0/sf)	Long-Skinny [1:5 aspect ratio] ($0/sf)
Building Orientation	0 (rotation from N)	45 (rotation from N)	90 (rotation from N)
Window to Wall Ratio	40% ($/sf)[1]	90% ($/sf)[1]	
HVAC Efficiency	Low Efficiency[2] ($8/sf of building area)	High Efficiency (~$9.3/sf of building area)	

1. Cost dependent on window type and aspect ratio.
2. The size (and resulting cost) of the HVAC system depends on other *options*.

We used the following equations to calculate *value* in units of NPV for the two *challenges*, and assumed $.10/kWh energy cost with 3% inflation:

Equation 1, Renovation:

NPV = Pre-renovation Energy Budget - 30 year Discounted Annual Energy Cost ($) – First Cost ($)

Equation 2, New Construction:

NPV = Baseline Project Budget - 30 year Discounted Annual Energy Costs ($) - First Cost($)

We developed NPV estimates primarily to be internally consistent, rather than accurate representations of true or full costs of actual building projects. For example, we excluded plumbing costs, and potential impact of site conditions relative to building orientation, and assumed utility rates to be fixed over the 30 years. We felt the abstractions were necessary and appropriate to reduce calculation time, manage the number of *alternatives* analyzed, and avoid introducing the concept of uncertainty into the experiment.

Figure 10 graphs the results of the full analysis of the two *challenges* in ascending order. It shows that although the renovation *challenge* includes fewer *alternatives*, a larger range of performance and a higher maximum *value* exists relative to the new construction *challenge*. The inset is an illustration of the clustering of results that occurs for various *options*. Each cluster in the inset represents the impact of all *options* relative to a single *option*. The *impact* of each single *option* can be seen between clusters. The inset serves as a visual demonstration that certain *decisions* have more *impact* than others.

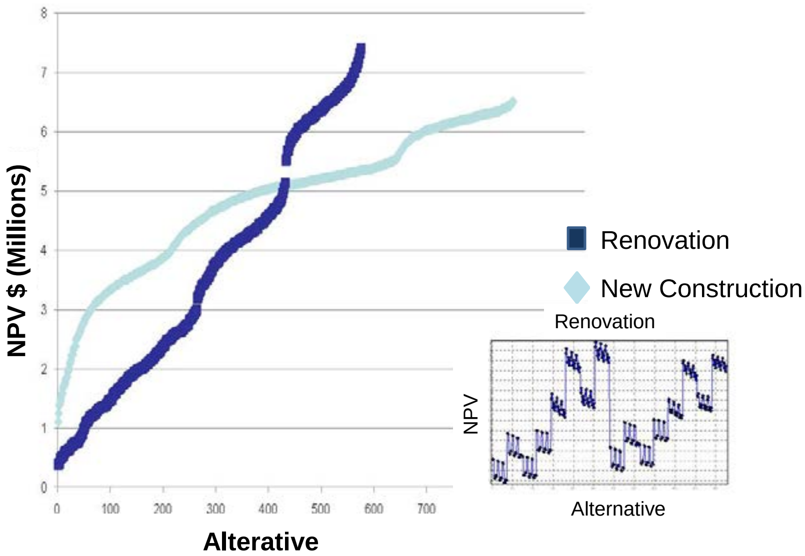

Figure 10. Full analysis of *alternatives* in renovation and new construction *challenges* ordered by NPV ($). Inset shows clusters of *value* generated by changing *options* in a given *challenge*.

Challenge Assessed

After *value space* is generated, the next step in DEAM is to assess the *challenge*. Table 3 shows the *challenge* metrics evaluated for the renovation and new construction *challenges.*

Table 3. *Challenge* metrics evaluated for renovation and new construction *challenges*.

	Challenge	
Metric	**Renovation**	**New Construction**
Objective Space Size (OSS)	2	2
Alternative Space Interdependence (ASI)	.58	.70
Impact Space Complexity (ISC)	.25	.25
Value Space Dominance (VSD)	.31	.63
Total:	3.14	3.58

Evaluation of these metrics supports the following observations:

(1) *Objective space size* (OSS) is 2 for both *challenges* since first cost and annual energy cost are considered in both.
(2) *Alternative space interdependence* (ASI) is higher in the new construction *challenge*, meaning a higher number of interactions exist among the 26 pairings of *variables* in the new construction *alternative space* than the renovation *alternative space*. This is not surprising since changing building geometry can affect numerous *variables*.
(3) *Impact space complexity* (ISC) is equal in both *challenges* meaning the renovation and new construction *impact spaces* include a similar number of design trade-offs. Both *challenges* have the same number of *goals* (2) and *variables* (8). Analysis reveals that for both *challenges* two *variables*, window type and exterior shading, have competing *impacts*.
(4) *Value space dominance* (VSD) is significantly lower for the renovation *challenge*, meaning select *variables* in the renovation *challenge* play a more dominant role in the renovation *value space*, than select *variables* in the new construction *challenge*. Evaluation of ranked *impacts* of individual *variables* on *value* (i.e., NPV) demonstrates that one *decision*, HVAC efficiency, is highly dominant during renovation. This dominance is the result of the fact that, unlike in the new construction *challenge*, in the renovation *challenge*, the size of an existing (low efficiency) system is fixed, but a new system is “right-sized” doubly increasing the efficiency *impact*.

Assessment of the *challenge* metrics indicates that the new construction *challenge* is more complex than the renovation *challenge*: 3.58 versus 3.14 according to crude evaluation by simple summation. In general, although the number of *objectives* and complexity of *impacts* is comparable, more interactions occur between similarly influential *variables*, in the new construction *challenge*. Specifically, a greater number of *alternatives* exist in the new construction *alternative space*, but with less variation in *value space* (see Figure 10). Intuition might tell us that new construction is generally a simpler *challenge* with more opportunity for gain in energy efficiency. However, some of the *variables* in the renovation *challenge* include *options* that do not meet energy code requirements or include an improperly sized HVAC system. Such *options*, excluded from the new construction *challenge* assuming code-compliant design for new construction, result in an artificially low baseline in building renovation by providing “low-hanging fruit.” In addition, as previously noted, the renovation *challenge* is highly dominated by the HVAC selection *variable*. By comparison the new construction *challenge* is not strongly dominated, but has more interdependences resulting in more trade-offs and less obvious *option* selections.

Exploration Conducted

The next step in this application of DEAM is to observe designers conducting *explorations*. To gather this data, we used the Charrette Test Method (Clayton et al, 1998), an established research technique that “employs a short but intensive design problem and compares the performance of several designers in undertaking the problem using various carefully defined design processes.” For this purpose, we developed a custom interface we called the EnergyExplorer^{TM}, shown in Figure 11. Using this tool, designers are able to quickly and easily generate and record design *alternatives*, and, employing certain strategies, assess their energy performance. Supporting the interactive interface are hidden libraries in excel that contain the results from pre-simulated full NPV analyses for both the new construction and renovation *value spaces.* Participants have access to different levels of data analysis (none, point, or trend data) depending on which *strategy* they implement.

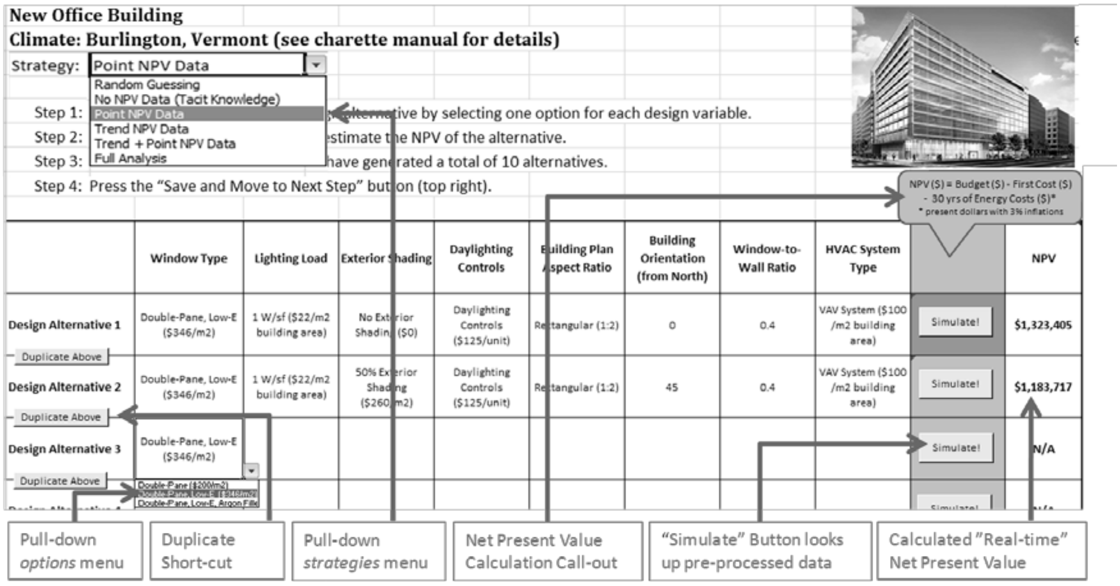

Figure 11. Custom interface for EnergyExplorer™, an interactive software tool used by charrette participants to document *explorations* supported by various *strategies*.

Figure 11 shows sample new construction *alternatives* generated by participants using the EnergyExplorer™ interface. If the tacit knowledge *strategy* had been shown, by way of example, participants would not have access to NPV data (far right).

Charrette tests were conducted with 15 building industry professional participants. These participants were asked to answer several questions regarding their professional background. Information regarding professional background was not used in the statistical analysis of results presented in this paper; nevertheless, we briefly summarize participant profiles to highlight the diversity and experience among the professionals who participated in the two charrettes. Professional roles of participants included: 4 Energy Analyst/Consultants, 2 Construction Managers, 3 Mechanical Engineers, 4 Program Managers, 1 Designer/Architect, 1 Owner/operator; years of experience in industry ranged from 0-5 to over 20; and, level of self-reported energy expertise ranged from low to high (with no individuals claiming to be an expert although several worked as energy consultants). Not surprisingly, given the variety of industry roles and experience represented, participants had significantly different exposure to energy modeling in practice with a few individuals reporting that energy modeling was used on 0-5% of their projects and others reporting it was used on >75% of their projects. Finally, on the real-world projects where energy modeling was performed, nearly all professionals report that typically 2-5 energy simulations were run. This estimate is consistent with the findings of other researchers (Flager & Haymaker, 2007), (Gane & Haymaker, 2010), although in another similar case study (Clevenger & Haymaker, 2011) 13 energy simulations were performed. Regardless of whether 2 or 13 energy model simulations are performed, observations confirm that traditional industry energy modeling practice most closely aligns with Design Strategy 3: Point NPV Data.

Strategies Assessed

Each participant completed a total of four *explorations* across two *challenges*, each using a different *strategy* (No NPV Data, Point NPV Data, Trend NPV Data and Trend + Point NPV Data). Each participant generated up to 10 *alternatives* using each *strategy* to create approximately 40 *alternatives* per participant during the charrette test. PIDO was used to generate the *alternatives* produced by random guessing and full analysis *strategies* for both *challenges*. We used EnergyExplorer^{TM} to record the total of up to 150 *alternatives* per *strategy,* or approximately 75 *alternatives* generated for each of the two *challenges*.

Table 4 shows the *strategy* metrics evaluated for the six *strategies* tested averaged over both renovation and new construction *challenges*. The *objective space quality* (OSQ) calculation reflects the degree (0 to 1) to which a given *strategy* provides designers information directly representative of a stated *goal(s)* (e.g.; maximize NPV). For the *strategies* that provide *value* (i.e.; NPV) results to designers, OSQ = 1. For the *strategies* where no data is provided about NPV, (i.e.; random guessing), OSQ = 0. Finally, where *impact* patterns are identified, but individual *alternative's impact* data are not provided, we assigned OSQ = .5, and where presumably designers had some tacit knowledge but no definitive analysis, we assigned OSQ = .33. Although these evaluations of OSQ are internally consistent, they remain somewhat subjective. Future research is necessary to more fully define the evaluation of OSQ, particularly for complex *challenges* with wide-ranging and disparate *goals*.

Alternative space sampling (ASS) indicates the percentage (although not the distribution) of the *alternatives* generated relative to the total number of feasible *alternatives* in *value space*. We generated the trend data provided during the charrette from a DoE using a PIDO script to automate EnergyPlus. Nevertheless, ASS for trend (NPV) data requires only a statistical sampling to determine trend data, which can be added in future research.

Alternative space flexibility (ASF) calculations communicate the number of differences between *options* among the *alternatives* generated by a given *strategy* regardless of sequence. For automated *strategies* (e.g.; full analysis) this metric is evaluated as a very small number of changes, but is deterministic and can be calculated prior to *exploration*. For *strategies* that incorporated human decision (e.g.; tacit knowledge, point analysis, trend analysis etc.) the metric is assessed after human *exploration*. As observed, the *strategy* that generated *alternatives* with more flexibility among *options* was trend data, followed by tacit knowledge and trend + point data. Full analysis has an extremely low ASF since each alternative only differs by one *option* of one *variable*, while random guessing is likely to have different *options* for up to 50% of the *variables*. ASF signifies that while varying one *option* in isolation provides significant information in the comparison of two *alternatives*, it provides almost no information about the overall *design space*.

Table 4. *Strategy* metrics evaluated for six *strategies* tested. Results support characterization and comparison of *strategies*.

Metric	Random Guessing	No NPV Data (Tacit Knowledge)	Point NPV Data	Trend NPV Data	Trend + Point NPV Data	Full Analysis
Objective Space Quality (OSQ)	0	.33	1	~.5	1	1
Alternative Space Sampling (ASS)	0	0	~.015 (10/576; 10/864)	~.40 (231/576; 266/864)	~.415	1 (576/576; 864/864)
Alternative Space Flexibility (ASF)	~.5	.25	.19	.31	.23	~.001
Total:	.5	.58	1.205	1.21	1.645	2.001

Debate exists in design theory regarding the importance of individual metrics. For example, is depth versus breadth a more effective *strategy?* (Goldschmidt, 2006) argues in favor of the depth *strategy* where novices to expert designers work in a limited *alternative space* to conduct a more complete *exploration*. In contrast, (Akın, 2001) argues that expert designers tend to start with breadth before depth *strategies*. Many researchers argue that design improvement and innovation are generally supported by breadth (Sutton, 2002). Multiple assessments comparing *strategies* are possible using the *strategy* metrics presented in Table 4. For example, depth versus breadth can be assessed and compared as a function of a combination of variously weighted ASS, the number of *alternatives* generated, and ASF, the level of similarity among *alternatives.*

For this research, we simply sum the three metrics without weights to evaluate the six *strategies* from least to most advanced (Table 5). Future research could be performed that varies the assessment algorithm to assess and compare *strategies* differently. Although certain *strategies* score closely in our research, the contribution is that application of these metrics supports a range of comparisons of *strategies*.

Table 5. Ranking of *strategy* from least to most advanced, based on *strategy* metrics.

Level of Advancement (1= least, 6= most)	**Strategy**	**Sum of Strategy Metrics**
1	Random Guessing	.5
2	No NPV Data, Tacit Knowledge	.58
3	Point NPV Data	1.205
4	Trend NPV Data	1.21
5	Trend + Point NPV Data	1.645
6	Full Analysis	2.001

Initial assessment, while crude, appears intuitive since rank is ordered according to the level of data provided to the designer by a given *strategy*. Future research may be performed to determine if past performance on a given *challenge* with a given *strategy* is a reliable predictor of *exploration* performance on similar *challenges* in the future. If so, this will enable consideration of the effectiveness of *strategies* relative to *challenge* prior to *exploration*.

Explorations Assessed

The next step in DEAM is to analyze the *exploration* achieved by the *strategy* implemented. Table 6 summarizes the *exploration* metrics assessed for the six *strategies* across the two *challenges*. Results, originally NPVs, are normalized to percentages of *value space* maximum. For example, VSA for an *exploration* is the average *value* achieved among *alternatives* generated using a given *strategy* over the maximum *value* of the full *value space*. In the renovation *challenge* a significant number of *alternatives* have a low *value*, which results in a low VSA. Conversely, in the new construction *challenge* a significant number of *alternatives* have relatively high *value*, which results in a high VSA.

Table 6. *Exploration* metrics evaluated for the six *strategies* and two *challenges* tested. Data is shown as a percentage of the maximum *value* of the *value space*. VSI is the exception; assessments represent the number of iterations generated prior to achieving maximum *value* in a given *exploration.* Data sample size generated by charrette participants using the various *strategies* is provided under each assessment in parentheses.

Metric	Challenge	Random Guessing	No NPV Data (Tacit Knowledge)	Point NPV Data	Trend NPV Data	Trend + Point NPV Data	Full Analysis
Value Space Average (VSA)	Renovation	~30% (100)	51% (52)	58% (53)	73% (36)	53% (41)	30% (576)
	New Construction	~74% (100)	79% (37)	86% (38)	87% (41)	84% (41)	74% (864)
Value Space Range (VSR)	Renovation	<37% (100)	41% (52)	38% (53)	28% (36)	43% (41)	70% (576)
	New Construction	<18% (100)	15% (37)	13% (38)	17% (41)	24% (41)	74% (864)
Value Space Iterations (VSI)	Renovation	~5	4.2	5.7	4.4	3.8	576
	New Construction	~5	5.0	5.8	2.4	4.0	864
Value Space Maximum (VSM)	Renovation	~67% (10)	92% (7)	92% (7)	93% (7)	76% (7)	100% (1)
	New Construction	~83% (10)	97% (6)	95% (6)	99% (6)	98% (6)	100% (1)

The following notes discuss the findings presented in Table 6. Statistical significance was determined using the threshold of a T-distribution test < .10. While sample size (shown in Table 6 in parentheses) was sufficient in certain cases to make claims about statistical significance, in many cases it was not. The real value of the data is less in the findings themselves and more in the fact that this methodology support systematic comparison of *strategies*. Furthermore, such comparison may be customizable depending on designer or researcher preference. Observable findings from this initial data set include:

(1) Tacit knowledge guides the generation of a set of *alternatives* that have a higher average *value* than those generated using random guessing.
Supporting evidence: The VSA of *alternatives* generated using tacit knowledge was, on average, superior to the average *value* of *alternatives* guided by random guessing. The improvement maintains statistical significance by a margin of 44% for the renovation *challenge*, and 3% in the new construction *challenge*.