Journal of the Operational Research Society (2006) 57, 1197–1215 0160-5682/06 $30.00

www.palgrave-journals.com/jors

A priority assessment multi-criteria decision model for human spaceflight mission planning at NASA

M Tavana*

La Salle University, Philadelphia, USA

Analog missions are real-life, Earth-based science missions whose purpose is to help understand the operations, techniques, and technologies required to perform similar tasks during future human spaceflight missions. The goal of performing an analog mission is to prepare crewmembers and support teams for future space missions in a low risk, low-cost environment. Vehicle, habitat, and surface terrain simulators are used to test hardware, operations, and tasks repeatedly for analog missions. This study presents a multi-criteria decision making model that was developed for the Integrated Human Exploration Mission Simulation Facility project at Johnson Space Center to assess the priority of a set of human spaceflight mission simulators. The proposed framework integrates subjective judgments derived from the analytic hierarchy process with entropy data into a preference model to prioritize five mission simulators for the human exploration of Mars.

Journal of the Operational Research Society (2006) **57,** 1197–1215. doi:10.1057/palgrave.jors.2602107
Published online 9 November 2005

Keywords: analytic hierarchy process; entropy; multi-criteria decision making; human spaceflight; NASA

1. Introduction

The primary goal in multi-criteria decision making (MCDM) is to provide a set of criteria aggregation methodologies that enable the development of models considering the decision makers' (DMs') preferential system and judgment policy (Doumpos and Zopounidis, 2002). Achieving this goal requires the implementation of complex processes. While intuition and simple rules are still favourite decision-making methods, they may be dangerously inaccurate for complex decision problems. The model presented here can help DMs improve their decision quality when they are confronted with complex decision problems. Our model ensures consistency and completeness of the required information and synthesizes a vast amount of information using a manageable and easy to understand structure, while incorporating the use of human intuition and subjective analysis skills.

The analytical processes in our model help a DM decompose complex MCDM problems into manageable steps, making this model accessible to a wide variety of DMs and situations. We use analytic hierarchy process (AHP), the entropy method, and preference modelling to help DMs crystallize their thoughts and reduce the inconsistencies associated with MCDM. Although technical details of the model may be beyond the reach of some DMs, the basic concepts are not difficult to understand or implement. As such, the DMs can use available analytical tools and techniques with some assistance from the experts (Schoemaker and Russo, 1993).

The structured framework presented in this study has some obvious attractive features:

(i) The generic nature of the model allows for the subjective evaluation of a finite number of decision alternatives on a finite number of performance criteria by a group of DMs. The mathematical and computational properties of the model are applicable to a wide range of real-world decision-making problems in MCDM.
(ii) The information requirements of the model are stratified into a hierarchy (criteria and their respective sub-criteria) that simplifies information input and allows the DMs to focus on a small area of the large problem. This process is also useful for seeking input from different experts or levels of management in the organization.
(iii) Inconsistencies are inevitable when dealing with subjective information from different DMs. The built-in inconsistency-checking mechanism of AHP helps to identify inconsistencies in judgments at very early stages of the computation process.

2. Problem description and background

The Integrated Human Exploration Mission Simulation Facility (*INTEGRITY*) project initiated at the Johnson Space Center (JSC) is expected to play an important role in

*Correspondence: M Tavana, Management Department, Management Information Systems, La Salle University, Philadelphia, PA 19141-1199, USA.
E-mail: tavana@lasalle.edu
URL: http://lasalle.edu/~tavana

increasing the success of analog missions. Analog missions are real-life, Earth-based science missions whose purpose is to help understand the operations, techniques, and technologies required to perform similar tasks during future human spaceflight missions. The goal of performing an analog mission is to prepare crewmembers and support teams and increase the productivity and scientific return during future space-based science missions in a low risk, low-cost environment. As nearly anyone with field experience can attest, even the most well-planned experiment usually does not go as expected—hardware does not work as desired, unforeseen logistical problems arise, planned procedures need to be modified to fit changing situations, etc. Theoretically, analog missions work many of these kinks out of the process before a similar mission is attempted in space. Although analog missions are able to perform these functions in a fashion that is considerably less costly than performing the missions in space, they still require budgets that are not insignificant.

Analog missions offer a unique opportunity to test hardware, operations, and tasks repeatability at a minimal cost and risk. Since these tests are performed in such a controllable environment, the opportunity arises to conduct numerous experiments that could help to answer several outstanding questions. For example, a time delay could be inserted easily into the communication systems that would help mission planners understand how to deal with this problem when actual crews are at a great distance from Earth. Another example of an item to be tested is the scheduling of a crew during an extended mission. Analog missions offer the opportunity to test various task schedules in order to help determine what reasonably can be expected of the crew over a given period of time. This information helps mission planners to set realistic goals and optimize the quality of the scientific return during actual missions. Analog missions also provide system designers with the opportunity to build, test, and evaluate their concepts. Once the system engineers and crews have been able to evaluate a given design, the feedback should flow back through the system engineer and their respective divisions. This information should then feedback into future design iterations made by the engineers, to create more robust systems and elements in subsequent designs.

The *INTEGRITY* Team (IT) is a 17-member team of experts and scientists from different divisions within the JSC. IT supports analog missions by providing human exploration mission vehicle, habitat, and planetary surface terrain simulation facilities and infrastructure. These simulation facilities are used to conduct integrated testing with human crews in support of advanced research and technology development efforts associated with future human exploration missions beyond low Earth orbit. Five simulators, capable of simulating the crew cabin architecture, integrated systems operations, and crew operations associated with a 1000-day class mission transit vehicle, will be housed within these simulation facilities. The IT is chartered with the task of assessing each of the following five simulators and developing an implementation plan that is in sync with JSC's mission, objectives and priorities.

Transit vehicle simulator (TVS): A transit vehicle for manned exploration missions would be used to transfer the mission crews between low Earth orbit, a gateway station such as the International Space Station, and the orbits for a mission destination such as an asteroid, the Moon or Mars. The TVS will be habitable and will attempt to accurately reflect appropriate mission scenarios and constraints by supporting a crew for 6–12 months.

Lander vehicle simulator (LVS): After the transit vehicle reaches orbit around a mission destination, the crew must descend to the surface by means of a landing vehicle. The lander vehicle must also provide ascent capability for rendezvous with and return to the transit vehicle in orbit. The lander vehicle simulator (LVS), a ground-based version of such a lander vehicle, will be habitable for short durations of several hours to several days.

Surface habitat simulator (SHS): A surface habitat on an extra-terrestrial surface must support and provide capabilities for a human crew to live and conduct an extra-terrestrial surface mission. It is likely that the duration of a surface mission will be in the range of 1 month to a year or more. The opportunities for Earth-based re-supply of a surface habitat will be limited, so the habitat must be largely self-supporting.

Roving vehicle simulator (RVS): Rovers offer several capabilities including surface translation for the crew, access to places considered too dangerous for manned exploration, maintenance of surface habitat, and scouting before surface operations by the crew. Exploration mission rovers consist of one or more surface-roving vehicles may be manned or unmanned and may have various levels of autonomy.

Surface terrain simulator (STS): Distinct from the other Integrity simulators, the STS does not involve designing and building a piece of technology or vehicle. Rather, it is a re-creation of an extra-terrestrial surface within the *INTEGRITY* facility. The surface simulation is expected to integrate and evaluate the technical abilities and operations of the lander (LVS), surface habitat (SHS), and rovers (RVS). It will consist of materials and features that are deliberately constructed to present mission simulations with realistic features of an extra-terrestrial surface.

Figure 1 shows how the analog mission vehicle, habitat, and surface terrain simulators will be utilized for a manned exploration mission to Mars. The diagram depicts outbound phases of the mission. Starting with mission phase 1, a crew would travel from Earth-orbit to a way station or Gateway in a Crew Transfer Vehicle. *INTEGRITY* would simulate this phase using the TVS. Continuing in a TVS, phase 2 would take the crew from the gateway station into orbit around Mars. From orbit, the crew would descend to the Martian surface in a lander vehicle (LVS) in phase 3. In

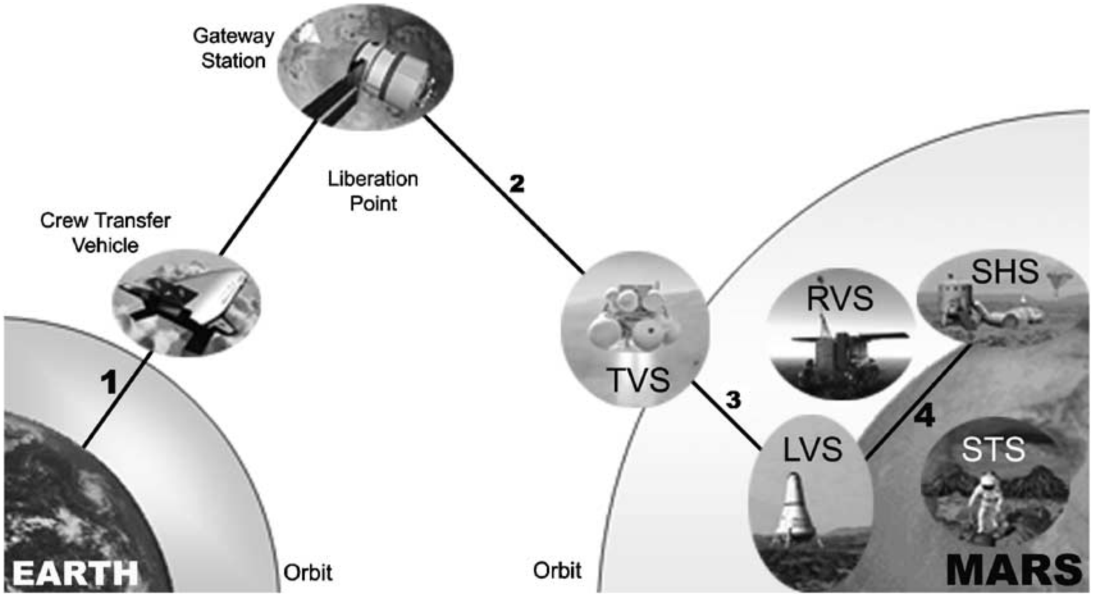

Figure 1 Mars mission elements and phases.

phase 4, the crew would egress the lander vehicle and traverse the Martian surface (STS) to man the surface habitat (SHS) using a roving vehicle (RVS).

3. Theoretical justifications

Roy (1990) argues that solving MCDM problems is not searching for some kind of optimal solution, but rather helping DMs master the (often complex) data involved in their problems and advance towards an acceptable solution. Multi-criteria analysis is not a ready-made recipe that can be applied to every situation. Methods should be chosen according to the value and quantity of information available, the nature of the problem, and the expectations of the DMs. As often happens in applied mathematics, the development of MCDM models is dictated by real-life problems. Therefore, it is not surprising that methods have appeared in a rather diffuse way, without any clear general methodology or basic theory (Vincke, 1992). The model presented here has evolved in an attempt to solve a complex MCDM problem at the JSC. It is a unique model with several features:

(i) Traditionally, MCDM frameworks fall into three categories: the multi-objective value analysis (Keeney and Raiffa, 1976), the outranking method (Vincke, 1992), and the interactive methods (Vanderpooten and Vincke, 1989). The selection of a framework depends on the type of the problem, the type of the choices (continuous or discrete), the type of measurement scales, the type of importance weights, the type of dependency among the criteria, and the type of uncertainty (Vincke, 1992). We show how the integration of several mathematically sound techniques can reduce some of the difficulties in the selection of an appropriate framework. Rather than moulding the problem to fit into a framework, we integrate several techniques into a framework to address problem requirements.

(ii) Finding the 'best' MCDM framework is an elusive goal that may never be reached (Triantaphyllou, 2000). Pardalos and Hearn (2002) have argued that one of the major issues for future research in MCDM is to explore ways of combining criteria aggregation methodologies to enable the development of models that consider the DMs' preferential system in complex problems. Belton and Stewart (2002) also argue the need for integrating frameworks in MCDM. Our model has solved a complex and judgmental multi-criteria problem by suitably combining a set of well-known and proven techniques in MCDM. This integration allows for the objective data and subjective judgments to be used side-by-side in a mathematically sound decision model.

(iii) We have developed an alternative approach to traditional AHP problem structuring where hierarchies of decision criteria and alternatives are used to solve MCDM problems. Our framework integrates AHP preferences of decision criteria with the priority scores and entropy information. This structured framework aggregates the intuitive preferences of multiple DMs to assess the overall performance of alternative scenarios using a weighted sum model (Triantaphyllou, 2000).

(iv) The Entropy method is a commonly used method for calibrating the weights assigned to different decision criteria in MCDM (Hwang and Yoon, 1981; Zeleny, 1982). A criterion does not influence the final choice much when all the alternatives have a similar value for that criterion. The entropy concept suggests that if a criterion's values are the same, the criterion can be eliminated from further consideration. Alternately, the weight assigned to a criterion can be smaller if all the alternatives have similar values for a criterion. On the other hand, when the differences between a criterion's values across particular alternatives are greater, the criterion is viewed as more important. The entropy concept has been shown to be particularly useful to investigate contrasts between sets of data.

4. The procedure and the model

The MCDM model presented in this study was used by the IT to assess the importance of each *INTEGRITY* simulator. Schoemaker and Russo (1993) describe four general approaches to decision-making ranging from intuitive to highly analytical. These methods include intuitive judgments, rules and shortcuts, importance weighting, and value analysis. They argue that analytical methods such as importance weighting and value analysis are more complex but also more accurate than the intuitive approaches (Schoemaker and Russo, 1993). Our approach is a simple and yet sophisticated multi-objective value analysis model that attempts to uncover some of the complexities inherent in the evaluation. The proposed approach uses a series of intuitive and analytical methods to visually display an alternative's preference to the ideal and nadir alternatives as well as all other competing alternatives. The highest achievable scores with all currently considered criteria form a composite, an ideal alternative (Zeleny, 1982). Similarly, we form a nadir alternative with the lowest achievable scores. Using the preference scores and the theory of displaced ideal to grasp the extent of the emerging conflict between means and ends, the IT members explore the limits attainable with each criterion. An implementation plan was developed based on the group preference rankings provided by the model. A four-phase process as presented in Figure 2 was used by the IT for this prioritization:

Phase 1. The relevant criteria and sub-criteria and their importance weights: Each IT member was asked to develop a list of all possible criteria relevant to the simulator evaluation. All the individual responses were collated by the facilitator into a comprehensive list which was shared with the IT at subsequent meetings. After several meetings and lengthy discussions, the list was revised by the IT into the hierarchical set of criteria and sub-criteria presented in Table 1.

The discussions about criteria and sub-criteria raised some concerns about the natural sequencing that may or may not exist among the simulators. The facilitator developed a questionnaire for the IT to investigate this question further. The IT was asked in this questionnaire to offer their sequencing perception by providing a ranking between 1 and 5 to each simulator. They were instructed to give a ranking of 1 to the simulator that they believed had to be built first, 2 to the second simulator, etc. They were also instructed to use an average ranking for cases where they believed two or more simulators had to be built simultaneously. For example, a score of 2.5 was used for simulators B and C, if it was believed that both of them had to be built simultaneously after A. Table 2 presents the mean sequencing scores of the five simulators for the each IT member along with the median scores of the simulators. We performed Kruskal–Wallis test to see if there is a significant difference among the medians. The high P-value of 0.804 (relative to $\alpha = 0.1$) shows that there is no statistical difference between the medians of the five simulators. In addition, because all z-values are considerably less than 1.96 ($\alpha = 0.05$), there is no difference for each individual simulator from the overall median.

4.1. The analytic hierarchy process

AHP was used to develop a set of importance weights for the criteria and sub-criteria. The IT identified i criteria and j sub-criteria to be used as evaluation criteria in simulator evaluation. The importance weight of each criterion (w_i) and sub-criterion (w_{ij}) was captured and measured with AHP using the questionnaire presented in Appendix A. The IT members were asked to provide their subjective assessment of each pairwise comparison. Assuming that an IT member believes, $c_1, c_2, \ldots, c_i$ are the i criteria that contribute to the importance of a simulator, the team member's goal is to assess the relative importance of these factors.

Saaty's AHP (Saaty and Vargas, 1998; Forman and Gass, 2001) is a method of deriving a set of weights to be associated with each of the i criteria or ij sub-criteria. Initially, the team member is asked to compare each possible pair c_j, c_k of criteria and provide judgments about which criteria are more important and by how much. AHP quantifies these judgments and represents them in an $i \times i$ matrix:

$$A = (a_{jk}) \qquad (j, k = 1, 2, \ldots, i)$$

If c_j is judged to be of equal importance as c_k, then $a_{jk} = 1$
If c_j is judged to be more important than c_k, then $a_{jk} > 1$
If c_j is judged to be less important than c_k, then $a_{jk} < 1$

$$a_{jk} = 1/a_{kj}, \qquad a_{jk} \neq 0$$

Since the entry a_{jk} is the inverse of the entry a_{kj}, the matrix A is a reciprocal matrix. a_{jk} reflects the relative importance of c_j compared with criterion c_k. For example, $a_{12} = 1.25$ indicates that c_1 is 1.25 times as important as c_2.

Then, the vector w representing the relative weights of each of the i criteria can be found by computing the

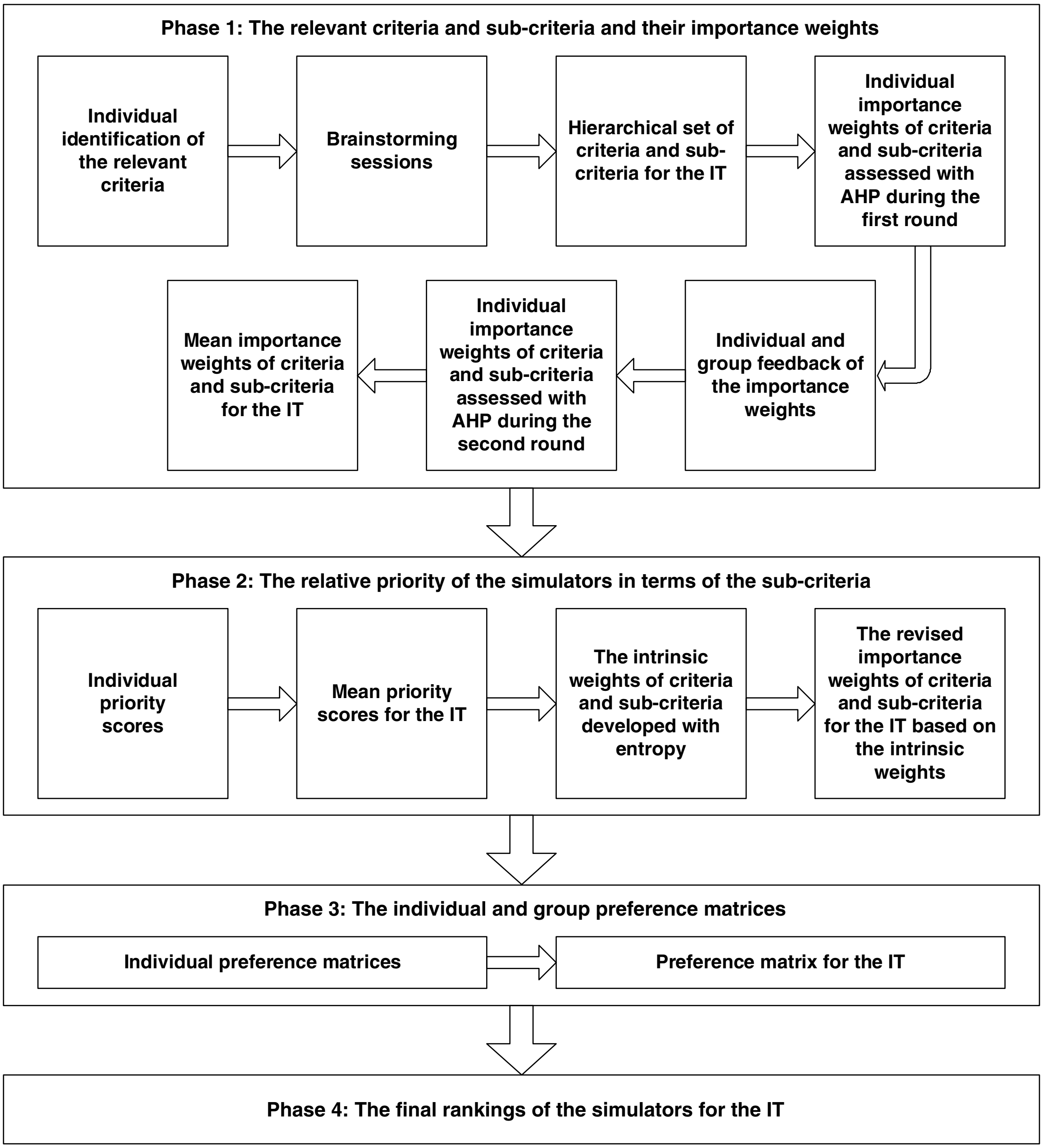

Figure 2 The overall procedure.

normalized eigenvector corresponding to the maximum eigenvalue of matrix A. An eigenvalue of A is defined as λ which satisfies the following matrix equation:

$$Aw = \lambda w$$

where λ is a constant, called the eigenvalue, associated with the given eigenvector w. Saaty (1977, 1980, 1983, 1989, 1990 and 1994) has shown that the best estimate of w is the one associated with the maximum eigenvalue ($\lambda_{\max}$) of the matrix A. Since the sum of the weights should be equal to 1.00, the normalized eigenvector is used. Saaty's algorithm for obtaining this w is incorporated in the software Expert Choice (Expert Choice, 2000).

One of the advantages of AHP is that it encourages team members to be consistent in their pairwise comparisons. Saaty suggests a measure of consistency for the pairwise comparisons. When the judgments are perfectly consistent, the maximum eigenvalue, $\lambda_{\max}$, should equal i, the number of criteria that are compared. In general, the responses are not perfectly consistent, and $\lambda_{\max}$ is greater than n. The larger the $\lambda_{\max}$, the greater is the degree of inconsistency.

Table 1 Simulator ranking criteria and sub-criteria

1. *Cost:* Cost-savings on future exploration should be an important benefit of the *INTEGRITY* Project. Each Integrity simulator should therefore provide high value to the project relative to its cost. This criterion is a measure of the resources required to accomplish a portion of the project.

1.1. *Cost to design (minimize):* The relative cost of designing the simulator, inclusive of all design efforts.
1.2. *Cost to build (minimize):* The relative cost of building the simulator including both material and labor costs.
1.3. *Cost to operate (minimize):* The relative cost of performing test operations in the simulator.
1.4. *Cost to reconfigure (minimize):* The relative cost of reconfiguring the simulator between tests.
1.5. *Cost to maintain (minimize):* The relative cost of maintaining the simulator over the 'long haul'.

2. *Strategic plan relevance:* The NASA Strategic Plan identifies a list of exploration, science, and discovery questions that will drive the mission of the Agency. 'Using our unique knowledge and expertise, we build the tools that enable revolutionary robotic and human missions. Through scientific research and strategic investments in transformational technologies, we open new pathways toward missions that were impossible only a few years ago'. This criterion examines how well a simulator under consideration meets the needs identified in the Agency Strategic Plan and enables exploration. It is a measure of the degree to which a simulator reflects, supports, or incorporates specific, identifiable features of the NASA Strategic Plan, the Code U/Bioastronautics Strategic Plan, the Integrated Space Plan, and/or the Bioastronautics Critical Questions List.

2.1. *Exploration development relevance (maximize):* The degree to which the element under consideration helps to resolve critical issues, answers critical science, operations, or technology questions, or clarifies exploration mission requirements.
2.2. *Mission architecture commonality (maximize):* All missions have certain operational methods and technology requirements that are common to all destinations (eg, sparing, approach to maintenance, repair, and redundancy, degree of automation, etc). Therefore, investigation of one of these issues will have benefits to all future missions, regardless of the destination. *INTEGRITY* should maximize the efforts to investigate the use of common operations and/or technology.
2.3. *Usefulness outside INTEGRITY (maximize):* A measure of the potential a simulator has to be useful for applications outside of the *INTEGRITY* project—including International Space Station, Space Shuttle Program, exploration, commercial, and other non-NASA.
2.4. *Publicity value (maximize):* Does this simulator have inspirational value from a visual and public/educational outreach perspective? What is its publicity value or 'cool factor''?
2.5. *Mission requirements understanding (maximize):* Many of the requirements for a future exploration mission are not yet understood or even identified. This sub-criterion measures the value of a simulator in uncovering hidden requirements and for defining requirements that are currently unspecified.
2.6. *Risk mitigation (maximize):* Consider all Programmatic risks, not only technical and operational risks, but also cost and schedule. Does this element have value to decrease the risk for a future exploration mission?

3. *Complexity:* This criterion is a measure of the overall intricacy of any simulator, relative to all of the other simulators or common NASA test facilities.

3.1. *Design and manufacturing difficulty (minimize):* A measure of the technical challenge of creating each simulator, as measured by things like number of drawings required, complex fabrication techniques, hard-to-work-with or rare materials, etc.
3.2. *Operational and safety difficulty (minimize):* The degree of complexity required for operating each simulator safely. The complexity may be revealed through number of failure modes, frequency/duration of required human interaction/operation, architecture-introduced safety risks (eg, confined spaces, elevated walkways, low passageways, etc).
3.3. *Sequencing with other stepping stones (maximize):* Integrity will use a stepping stone approach to gradually build up to a complete long-duration ground mission capability. How well do the simulators under consideration build upon previous stepping-stones and enable future stepping-stones? Is it a natural next extension of the project or facility that will seamlessly evolve into the next major objective, or is it a stand-alone evaluation of a component, operation, or science objective?
3.4. *Infrastructure impacts (minimize):* A measure of the degree to which the facility infrastructure must be modified to accommodate a simulator. Infrastructure includes buildings, utilities, data networks, etc.

4. *Science, operations, and technology: INTEGRITY* is part of a larger Agency effort aimed and developing our capabilities to fulfill our strategic vision for exploration. This criterion is a measure of the ability of a simulator to accommodate the range of exploration mission technologies, science and operations, and the degree to which a simulator allows their evaluation.

4.1. *Basic science opportunities (maximize):* This sub-criterion measures inherent or serendipitous opportunities for conducting basic science investigations within the facility and simulation whether it coincides with the test mission's science objectives or not.
4.2. *Exploration science diversity (maximize):* A measure of the degree to which a simulator reflects the spectrum of science activities and equipment needed to conduct exploration science (the science that will be conducted during exploration missions). These elements of exploration science may be simulations rather than actual experiments, scripted operational activities that are part of the overall simulation.
4.3. *Operations scenario diversity (maximize):* A reflection of the number and diversity of operations areas that can be evaluated in a particular simulator.
4.4. *Cross-cutting technology (maximize):* The number of cross-cutting or in-common technologies that are included or can be evaluated in a particular simulator. Cross-cutting technologies are those with broad applications for the entire Agency. This sub-criterion also reflects science and operations elements (eg medical contingency operations must be present at every phase of the mission and have applications for all manned space flight activities).
4.5. *Interfaces/integration (maximize):* The number of different interfaces and the level of systems integration that can be investigated with each simulator.
4.6. *Technology gaps/innovation (maximize):* The degree to which a simulator will require or can incorporate emerging and/or innovative technologies, science, or operations that address critical technology gaps.

Table 2 Simulator sequencing scores

IT member	*TVS*	*LVS*	*SHS*	*RVS*	*STS*
A	4.0	5.0	1.0	3.0	2.0
B	1.5	3.0	1.5	4.0	5.0
C	1.0	2.5	2.5	5.0	4.0
D	5.0	2.0	1.0	4.0	3.0
E	5.0	3.0	2.0	1.0	4.0
F	3.5	3.5	5.0	2.0	1.0
G	3.0	1.0	2.0	5.0	4.0
H	5.0	2.0	3.0	4.0	1.0
I	2.0	4.0	1.0	5.0	3.0
J	5.0	4.0	2.5	1.0	2.5
K	3.5	5.0	3.5	1.0	2.0
L	1.0	5.0	2.0	3.0	4.0
M	4.0	1.5	5.0	1.5	3.0
N	1.0	4.0	2.5	5.0	2.5
O	4.0	1.0	5.0	3.0	2.0
P	2.0	4.0	3.0	5.0	1.0
Q	2.0	4.5	4.5	1.0	3.0
Median	3.5	3.5	2.5	3.0	3.0
z-value	0.27	0.76	−0.77	0.48	−0.75

Saaty defines the consistency index (*CI*) as $(\lambda_{\max}-i)/(i-1)$, and provides the following random index (*RI*) table for matrices of order 3–10:

n	3	4	5	6	7	8	9	10
RI	0.58	0.90	1.12	1.32	1.41	1.45	1.49	1.51

This *RI* is based on a simulation of a large number of randomly generated weights. Saaty recommends the calculation of a consistency ratio (*CR*), which is the ratio of *CI* to the *RI* for the same order matrix. A *CR* of 0.10 or less is considered acceptable. When the *CR* is unacceptable, the team member is made aware that his or her pairwise comparisons are logically inconsistent, and he or she is encouraged to revise them.

The responses were processed with Expert Choice (Expert Choice, 2000) and those with inconsistency ratios greater than 0.10 were asked to reconsider their judgments as suggested by Saaty. The mean importance weights were calculated for the IT after the necessary adjustments were made to inconsistent responses. Each IT member was presented with his/her individual score along with the group mean weights. IT members were given the opportunity to revisit their judgments and make revisions to their pairwise comparison scores based on this feedback. Some IT members took advantage of this opportunity and revised their judgments in the second round. The mean importance weights for the first and second round are presented in Table 3. While the second round results differ slightly from the first round results, the inconsistency was improved significantly. A similar approach was used to determine the relative importance of each sub-criterion.

Table 3 Rounds 1 and 2 importance weights of the criteria and sub-criteria

Criterion	*Round-1*	*Round-2*
1. Cost	0.167	0.170
2. Strategic plan relevance	0.439	0.429
3. Complexity	0.113	0.124
4. Science, operations, and technology	0.281	0.278
Inconsistency ratio	0.070	0.049
1. Cost sub-criteria		
1.1. Cost to design	0.234	0.231
1.2. Cost to build	0.166	0.163
1.3. Cost to operate	0.286	0.286
1.4. Cost to reconfigure	0.125	0.125
1.5. Cost to maintain	0.189	0.195
Inconsistency ratio	0.069	0.058
2. Strategic plan relevance sub-criteria		
2.1. Exploration development relevance	0.274	0.273
2.2. Mission architecture commonality	0.167	0.168
2.3. Usefulness outside *INTEGRITY*	0.086	0.086
2.4. Publicity value	0.084	0.084
2.5. Mission requirements understanding	0.175	0.179
2.6. Risk mitigation	0.215	0.211
Inconsistency ratio	0.071	0.061
3. Complexity sub-criteria		
3.1. Design and manufacturing difficulty	0.202	0.197
3.2. Operational and safety difficulty	0.436	0.431
3.3. Sequencing with other stepping stones	0.167	0.170
3.4. Infrastructure impacts	0.195	0.202
Inconsistency ratio	0.064	0.049
4. Science, operations, and technology sub-criteria		
4.1. Basic science opportunities	0.163	0.157
4.2. Exploration science diversity	0.170	0.174
4.3. Operations scenarios	0.204	0.206
4.4. Cross cutting technology	0.126	0.126
4.5. Interfaces/integration	0.203	0.201
4.6. Technology gap/innovation	0.135	0.136

There has been some criticism of AHP in the operations research literature. Harker and Vargas (1987) show that AHP does have an axiomatic foundation, the cardinal measurement of preferences is fully represented by the eigenvector method, and the principles of hierarchical composition and rank reversal are valid. On the other hand, Dyer (1990a) has questioned the theoretical basis underlying AHP and argues that it can lead to preference reversals based on the alternative set being analysed. In response, Saaty (1990) explains how rank reversal is a positive feature when new reference points are introduced. We use the geometric aggregation rule to avoid the controversies associated with rank reversal (Dyer, 1990a,b; Saaty, 1990; Harker and Vargas, 1990).

Phase 2. The relative priority of the simulators in terms of the sub-criteria: The priority scores of the five simulators for all sub-criteria were needed to develop the preference scores in our model. The Simulator Assessment Questionnaire presented in Appendix B was given to the IT to capture their perception of the relative importance of each simulator for each sub-criteria using a 10-point Lickert scale. The individual priority scores were averaged into the group priority scores given in Table 4.

Next, we used the group priority scores presented in Table 4 to revise the initial weights of the sub-criteria (w'_{ij}—second-round weights in Table 3) using the entropy concept. The essential idea in entropy method is that the relative importance of a criterion is directly related to the information conveyed by the criterion relative to the set of alternatives under consideration. This means that the greater the dispersion in the evaluations of the alternatives for a given criterion, the more important the criterion. In other words, the most important criteria are those which have the greatest discriminating power between alternatives. In this method the importance weight of criteria could be determined without the direct involvement of the DM, in terms of the values derived from the evaluation of the alternatives. However, this is a complete contradiction to the notion that weights should represent the relative importance the DM attaches to the criteria. Therefore, we multiply the values of weights obtained by the entropy method (intrinsic weights) by the subjective weights representing the judgments of DMs obtained by the AHP. The final result, once normalized, is used in our model to represent the importance weight of the criteria and sub-criteria.

Each sub-criterion is an information source; therefore, the more information a sub-criterion reveals, the more relevant it is. Consequently, the more information the *j*th sub-criterion of the *i*th criterion reveals, the more relevant the sub-criterion is to the decision analysis. Zeleny (1982) argues that this intrinsic information must be used in parallel with the initial weights the DMs assigned to various sub-criteria. In other words, the overall importance weight of a criterion, w_{ij}, is directly related to the intrinsic weight, w''_{ij}, reflecting the average intrinsic information developed by the priority scores of the simulators, and the subjective weight, w'_{ij}, reflecting the IT members' subjective assessment of its importance.

The more different the scores of a sub-criterion are for a set of simulators, the larger is the contrast intensity of the sub-criterion, and the greater is the amount of information transmitted by that sub-criterion. Assuming that the vector $z_{ij} = (z^1_{ij}, \ldots, z^q_{ij})$ characterizes the *i*th criterion, the *j*th sub-criterion, and the *q*th simulator; the entropy measure of the

Table 4 The IT mean priority scores

Criterion	*LVS*	*RVS*	*SHS*	*STS*	*TVS*
1. Cost sub-criteria					
1.1. Cost to design	3.88	4.76	3.88	8.00	4.00
1.2. Cost to build	3.94	4.53	3.12	7.94	3.76
1.3. Cost to operate	4.71	5.29	3.41	8.65	4.35
1.4. Cost to reconfigure	4.53	5.24	3.94	7.88	4.06
1.5. Cost to maintain	4.53	4.12	3.76	7.65	4.35
2. Strategic plan relevance sub-criteria					
2.1. Exploration development relevance	6.29	6.53	9.06	5.53	8.53
2.2. Mission architecture commonality	7.12	6.53	8.94	4.29	8.59
2.3. Usefulness outside *INTEGRITY*	5.53	4.82	8.00	3.88	7.65
2.4. Publicity value	6.76	8.35	8.18	6.76	7.24
2.5. Mission requirements understanding	6.88	7.12	8.88	4.76	8.35
2.6. Risk mitigation	7.24	6.88	9.00	5.29	8.82
3. Complexity sub-criteria					
3.1. Design and manufacturing difficulty	4.76	5.18	3.82	7.24	4.35
3.2. Operational and safety difficulty	4.47	5.53	4.18	7.18	4.94
3.3. Sequencing with other stepping stones	6.76	5.12	7.82	5.29	7.71
3.4. Infrastructure impacts	5.00	5.59	4.53	5.76	4.47
4. Science, operations, and technology sub-criteria					
4.1. Basic science opportunities	5.18	5.24	9.00	4.88	7.18
4.2. Exploration science diversity	4.53	5.88	8.12	5.82	5.94
4.3. Operations scenarios	6.76	6.41	8.59	6.00	7.41
4.4. Cross cutting technology	6.53	5.94	8.94	3.82	7.82
4.5. Interfaces/integration	7.76	6.47	9.06	3.94	8.35
4.6. Technology gap/innovation	6.71	7.06	9.29	3.76	7.94

jth sub-criterion for the ith criterion is

$$e(z_{ij}) = -K \sum_{k=1}^{q} \frac{z_{ij}^k}{Z_{ij}} \ln \frac{z_{ij}^k}{Z_{ij}} \quad (1)$$

where

$$Z_{ij} = \sum_{k=1}^{q} z_{ij}^k \qquad (i = 1, 2, \ldots, m \text{ and } j = 1, 2, \ldots, n)$$

and $K>0$, ln is the natural logarithm, $0 \leqslant z_{ij}^k \leqslant 1$, and $e(z_{ij}) \geqslant 0$. When all z_{ij}^k are equal for a given i and j, then $z_{ij}^k/Z_{ij} = 1/q$, and $e(z_{ij})$ assumes its maximum value, which is $e_{\max} = \ln q$. By setting $K = 1/e_{\max}$, we achieve $0 \leqslant e(z_{ij}) \leqslant 1$. This normalization is necessary for meaningful comparisons. In addition, the total entropy is defined as

$$E = \sum_{j=1}^{n} e(z_{ij})$$

The smaller $e(z_{ij})$ is, the more information the jth sub-criterion transmits for the ith criterion, and the larger $e(z_{ij})$ is, the less information it transmits. When $e(z_{ij}) = e_{\max} = \ln q$, the jth sub-criterion for the ith criterion transmits no useful information. Next, the intrinsic weight is calculated as

$$w''_{ij} = \frac{1}{I - E}[1 - e(z_{ij})] \quad (2)$$

where I is the total number of sub-criteria for a particular criteria under consideration.

Since w''_{ij} is inversely related to $e(z_{ij})$, $1-e(z_{ij})$ is used instead and normalized to make sure $0 \leqslant w''_{ij} \leqslant 1$ and $\sum_{j=1}^{n} w''_{ij} = 1$. The higher $e(z_{ij})$, the less information content is provided by the jth sub-criterion for the ith criterion. When the information content of the jth sub-criterion for the ith criterion is low, the corresponding intrinsic weight (w''_{ij}) should be low. Thus, the intrinsic weight is assumed to be inversely related to the entropy and therefore, we use $1-e(z_{ij})$ in the definition of the intrinsic weight.

The more different the priority scores z_{ij}^k are, the larger w''_{ij} is and the more important the jth sub-criterion for the ith criterion is. When all the priority scores z_{ij}^k are equal for the jth sub-criterion for the ith criterion, then $w''_{ij} = 0$ for that sub-criterion. However, this is not true if the priority scores z_{ij}^k are equal for all the sub-criteria j. In that case, the weights are assumed to be equal or $w''_{ij} = 1/n$, where n is the number of sub-criteria for a given criterion. Entropy multiplies the intrinsic weight w''_{ij} by the subjective weight w'_{ij} and normalizes the product to calculate the overall importance weight of the jth sub-criterion for the ith criterion w_{ij},

$$w_{ij} = \frac{w''_{ij}\, w'_{ij}}{\sum_{j=1}^{n} w''_{ij} w'_{ij}} \quad (3)$$

When there is more than one priority score (n-ary sub-criteria), these priority scores are used to calculate the entropy within each simulator. These within-simulator intrinsic weights can influence the overall weight of the sub-criteria. In other words, the overall importance weight for an n-ary sub-criterion w_{ij} is related to its between-simulator intrinsic weight w''_{ij} the subjective weight $w_{ij}{}'$ and the within-simulator intrinsic weight. Table 5 presents the initial weights, the intrinsic weights and the revised weights of each sub-criterion.

There are two other methods for calculating the intrinsic weights of criteria. Diakoulaki *et al* (2000) proposes a method based on the correlation between the columns of the decision matrix. Another method consists in measuring the importance of each criterion as a member of a coalition by means of the Shapley value (Grabisch and Roubens, 1999). We use the entropy method suggested by Zeleny (1982, Chapter 7) in this study because it is readily available in MCDM, provides consistent results, and easily accepted by DMs (Pomerol and Barba-Romero, 2000, Chapter 4).

Phase 3. The individual and group preference matrices: The next model is used to identify a preference matrix for each IT member and an overall preference matrix for the IT as a group. To formulate the model algebraically, let us assume, w_i is the weight of the ith criterion ($i = 1, 2, \ldots, m$), w_{ij} the weight of the jth sub-criterion of the ith criterion ($i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$) and z_{ij}^k the priority score of the kth alternative for the jth sub-criterion of the ith criterion ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$; and $k = 1, 2, \ldots, q$).

Assuming that a^k is the kth alternative and z_{ij}^k represents the priority score of the kth alternative for the jth sub-criterion of the ith criterion:

$$\begin{array}{cccccc}
w_i & w_{ij} & a^1 & a^2 & \cdots & a^q \\
w_1 & w_{11} & z_{11}^1 & z_{11}^2 & \cdots & z_{11}^q \\
w_1 & w_{12} & z_{12}^1 & z_{12}^2 & \cdots & z_{12}^q \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_1 & w_{1n} & z_{1n}^1 & z_{1n}^2 & \cdots & z_{1n}^q \\
w_2 & w_{21} & z_{21}^1 & z_{21}^2 & \cdots & z_{21}^q \\
w_2 & w_{22} & z_{22}^1 & z_{22}^2 & \cdots & z_{22}^q \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_2 & w_{2n} & z_{2n}^1 & z_{2n}^2 & \cdots & z_{2n}^q \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_m & w_{m1} & z_{m1}^1 & z_{m1}^2 & \cdots & z_{m1}^q \\
w_m & w_{m2} & z_{m2}^1 & z_{m2}^2 & \cdots & z_{m2}^q \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_m & w_{mn} & z_{mn}^1 & z_{mn}^2 & \cdots & z_{mn}^q
\end{array}$$

We identify the Ideal Alternative (a^*) whose priority scores are ($z_{11}^*, \ldots, z_{mn}^*$), where $z_{ij}^* = \max z_{ij}^k$ ($k = 1, 2, \ldots, q$).

Similarly, we identify the Nadir Alternative ($\underline{a}$) whose priority scores are ($\underline{z}_{11}, \ldots, \underline{z}_{mn}$), where $\underline{z}_{ij} = \min z_{ij}^k$ ($k = 1, 2, \ldots, q$).

Table 5 The initial, intrinsic, and revised importance weights of sub-criteria

Criterion	*Initial weight* (w_{ij}')	*Intrinsic weight* (w_{ij}'')	*Revised weight* (w_{ij})
1. Cost sub-criteria			
1.1. Cost to design	0.231	0.1992	0.226
1.2. Cost to build	0.163	0.2566	0.205
1.3. Cost to operate	0.286	0.2267	0.318
1.4. Cost to reconfigure	0.125	0.1574	0.097
1.5. Cost to maintain	0.195	0.1602	0.153
2. Strategic plan relevance sub-criteria			
2.1. Exploration development relevance	0.273	0.1421	0.236
2.2. Mission architecture commonality	0.168	0.2332	0.239
2.3. Usefulness outside *INTEGRITY*	0.086	0.2882	0.151
2.4. Publicity value	0.084	0.0332	0.017
2.5. Mission requirements understanding	0.179	0.1654	0.181
2.6. Risk mitigation	0.211	0.1378	0.177
3. Complexity sub-criteria			
3.1. Design and manufacturing difficulty	0.197	0.3826	0.290
3.2. Operational and safety difficulty	0.431	0.2957	0.490
3.3. Sequencing with other stepping stones	0.170	0.2389	0.156
3.4. Infrastructure impacts	0.202	0.0829	0.064
4. Science, operations, and technology sub-criteria			
4.1. Basic science opportunities	0.157	0.1818	0.180
4.2. Exploration science diversity	0.174	0.1073	0.117
4.3. Operations scenarios	0.206	0.0489	0.063
4.4. Cross cutting technology	0.126	0.2208	0.175
4.5. Interfaces/integration	0.201	0.2131	0.270
4.6. Technology gap/innovation	0.136	0.2280	0.195

Next, Z is revised by adding the Ideal and Nadir alternatives:

$$\begin{array}{cccccccc}
w_i & w_{ij} & a^1 & a^2 & \cdots & a^q & a^* & \underline{a} \\
w_1 & w_{11} & z_{11}^1 & z_{11}^2 & \cdots & z_{11}^q & z_{11}^* & \underline{z}_{11} \\
w_1 & w_{12} & z_{12}^1 & z_{12}^2 & \cdots & z_{12}^q & z_{12}^* & \underline{z}_{12} \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
w_1 & w_{1n} & z_{1n}^1 & z_{1n}^2 & \cdots & z_{1n}^q & z_{1n}^* & \underline{z}_{1n} \\
w_2 & w_{21} & z_{21}^1 & z_{21}^2 & \cdots & z_{21}^q & z_{21}^* & \underline{z}_{21} \\
w_2 & w_{22} & z_{22}^1 & z_{22}^2 & \cdots & z_{22}^q & z_{22}^* & \underline{z}_{22} \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
w_2 & w_{2n} & z_{2n}^1 & z_{2n}^2 & \cdots & z_{2n}^q & z_{2n}^* & \underline{z}_{2n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_m & w_{m1} & z_{m1}^1 & z_{m1}^2 & \cdots & z_{m1}^q & z_{m1}^* & \underline{z}_{m1} \\
w_m & w_{m2} & z_{m2}^1 & z_{m2}^2 & \cdots & z_{m2}^q & z_{m2}^* & \underline{z}_{m2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_m & w_{mn} & z_{mn}^1 & z_{mn}^2 & \cdots & z_{mn}^q & z_{mn}^* & \underline{z}_{mn}
\end{array}$$

Then, the *Preference Matrix* showing the pairwise preference strength of one alternative compared with another $P(a^c,a^d)$ is constructed using the following equation:

$$P(a^c, a^d) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_i \cdot w_{ij} \cdot (z_{ij}^c - z_{ij}^d) \Big/ \sum_{i=1}^{m}\sum_{j=1}^{n} w_i \cdot w_{ij} \cdot (z_{ij}^* - \underline{z}_{ij}) \qquad (4)$$

Note that $P(a^*,\underline{a}) = +1.000$ is the maximum possible strength and the preference of a_c compared with a_d is a percentage of this maximum possible strength. In addition, the number of pairwise comparisons needed to complete the Preference Matrix in our model is $(n+2)(n+1)/2$ because $P(a^c,a^d) = -P(a^d,a^c)$ and when comparing one element with itself, the comparison must give no preference (0).

$$\begin{array}{ccccccc}
 & a^1 & a^2 & \cdots & a^q & a^* & \underline{a} \\
a^1 & 0 & P(a^1,a^2) & \cdots & P(a^1,a^q) & P(a^1,a^*) & P(a^1,\underline{a}) \\
a^2 & -P(a^1,a^2) & 0 & \cdots & P(a^2,a^q) & P(a^2,a^*) & P(a^2,\underline{a}) \\
\vdots & \vdots & \vdots & 0 & \vdots & \vdots & \vdots \\
a^q & -P(a^1,a^q) & -P(a^2,a^q) & \cdots & 0 & P(a^q,a^*) & P(a^q,\underline{a}) \\
a^* & -P(a^1,a^*) & -P(a^2,a^*) & \cdots & -P(a^q,a^*) & 0 & P(a^*,\underline{a}) \\
\underline{a} & -P(a^1,\underline{a}) & -P(a^2,\underline{a}) & \cdots & -P(a^q,\underline{a}) & -P(a^*,\underline{a}) & 0
\end{array}$$

We used pairwise comparisons to measure the relative importance of one alternative to another. Using the importance weights and the priority scores, a preference matrix similar to the one presented in Table 6 was developed for each IT member. The preference matrices of the individual IT members were averaged to obtain the final group preference matrix presented in Table 7. The ideal and nadir alternatives are added to the list of the alternatives to show the relative relationship of each alternative to the ideal and nadir. The formula for the preference between two alternatives is normalized so that the preference of the ideal alternative compared to the nadir is the greatest and equals 1.

Table 6 Individual A's preference matrix

	LVS	*RVS*	*SHS*	*STS*	*TVS*	*Ideal*	*Nadir*
LVS	—	0.114	−0.025	0.280	0.221	−0.370	0.630
RVS	0.886	—	−0.139	0.165	0.106	−0.485	0.515
SHS	1.025	1.139	—	0.304	0.246	−0.346	0.654
STS	0.720	0.835	0.696	—	−0.059	−0.650	0.350
TVS	0.779	0.894	0.754	1.059	—	−0.591	0.409
Ideal	1.370	1.485	1.346	1.650	1.591	—	1.000
Nadir	0.370	0.485	0.346	0.650	0.591	0.000	—

Table 7 Group preference matrix

	LVS	*RVS*	*SHS*	*STS*	*TVS*	*Ideal*	*Nadir*
LVS	—	0.174	−0.090	0.184	0.263	−0.378	0.622
RVS	−0.174	—	−0.264	0.010	0.089	−0.552	0.448
SHS	0.090	0.264	—	0.274	0.353	−0.288	0.712
STS	−0.184	−0.010	−0.274	—	0.079	−0.562	0.438
TVS	−0.263	−0.089	−0.353	−0.079	—	−0.641	0.359
Ideal	0.378	0.552	0.288	0.562	0.641	—	1.000
Nadir	−0.622	−0.448	−0.712	−0.438	−0.359	−1.000	—

Phase 4. The final rankings of the simulators for the IT: The preference scores show the relative distance of one simulator from the ideal, nadir, and all other simulators. For example, in Table 7, LVS is more preferred to nadir with a relative distance of 0.622 while it is less preferred to ideal with a relative distance of 0.378 (given the 0–1 distance of nadir from ideal). Similarly, we can compare LVS with any other simulator such as RVS. LVS is preferred to RVS with a relative distance of 0.174. The pairwise preferences provided in Table 7 can be translated into the following rankings and relative distances:

Ideal > SHS > LVS > RVS > STS > TVS > Nadir
|← 0.288 →|← 0.090 →|← 0.174 →|← 0.010 →|← 0.079 →|← 0.359 →|

Next, by reconstructing the above rankings and distances for all individual IT members, we performed a one-sample *t*-test to see if the differences of the means are equal to zero (see Table 8). The *t*-test, with $\alpha = 0.05$, showed that Ideal–SHS, SHS–LVS, LVS–RVS, and TVS–Nadir were significantly different while there was no difference between RVS–STS and STS–TVS. We concluded that, with $\alpha = 0.05$, the relative distance of 0.010 between RVS and STS and the relative distance of 0.079 between STS and TVS are not different. We shared this result with the IT. While the difference between STS and TVS was not statistically significant at the $\alpha = 0.05$ confidence level, the team felt that there was a noticeable difference that was significant at $\alpha = 0.10$. The IT finalized its recommendation and agreed on the following final consensus rankings of the simulators: SHS > LVS > RVS = STS > TVS. The decision was made to implement SHS first, LVS second, both RVS and STS together next, and TVS last.

5. Evaluation

Prior to implementation of our model, the IT used simple voting to solve group decision problems. As a result of the lack of a systematic process and inadequate data gathering and analysis, the team was concerned about the potential for poor decision-making. Their desire to replace the voting method with a more complete, structured, and analytical process resulted in the creation of the model.

The proposed model was evaluated by the IT based on seven factors: decision quality, decision process quality, analytical capabilities, completeness, structured, flexibility, and simplicity. The questionnaire presented in Appendix C, was used to compare the *POST-INTEGRITY* framework with the *PRE-INTEGRITY* voting method. The results presented in Figure 3 shows that *POST-INTEGRITY* scores are higher on all of the assessment dimensions with the exception of simplicity. The *PRE-INTEGRITY* method was slightly preferred over the *POST-INTEGRITY* system in terms of simplicity.

Next, we performed Wilcoxon signed ranks test of medians. While computation of a median is easily justified for ordinal data, some statisticians have reservations about computing a mean for ordinal data. Therefore, we used median as the measure of central tendency. All the medians were statistically different from 0 for $\alpha = 0.05$. In summary, the medians were statistically different from 0 for all evaluation criteria. In addition, the medians showed that *POST-INTEGRITY* was favoured over *PRE-INTEGRITY* for all evaluation criteria with the exception of simplicity. For simplicity, *PRE-INTEGRITY* was marginally favoured over *POST-INTEGRITY* (see Table 9).

Table 8 Individual and group relative distances

IT member	*Ideal-SHS*	*SHS-LVS*	*LVS-RVS*	*RVS-STS*	*STS-TVS*	*TVS-Nadir*
A	0.346	0.025	0.114	0.165	−0.059	0.409
B	0.360	0.054	0.126	−0.132	0.060	0.533
C	0.265	0.000	0.247	−0.001	0.120	0.369
D	0.223	0.112	0.046	0.096	0.291	0.232
E	0.284	0.035	−0.052	0.266	0.168	0.299
F	0.299	0.037	0.337	−0.201	0.126	0.401
G	0.286	0.166	0.315	−0.047	−0.042	0.322
H	0.310	−0.039	0.283	0.226	−0.075	0.295
I	0.242	0.198	0.119	−0.276	0.447	0.271
J	0.101	0.234	0.218	0.226	0.040	0.182
K	0.251	0.000	0.436	−0.074	0.095	0.292
L	0.172	0.010	0.332	0.060	0.264	0.163
M	0.217	0.416	−0.019	0.022	0.000	0.364
N	0.347	0.016	0.390	−0.066	−0.200	0.514
O	0.553	−0.003	−0.023	−0.028	−0.157	0.658
P	0.288	0.036	0.073	−0.010	0.325	0.288
Q	0.356	0.231	0.024	−0.058	−0.072	0.519
Mean	0.288	0.090	0.174	0.010	0.079	0.359
t-value	13.05	3.26	4.88	0.29	1.92	11.76
Significant at 95%	Yes	Yes	Yes	No	No	Yes
Significant at 90%	Yes	Yes	Yes	No	Yes	Yes

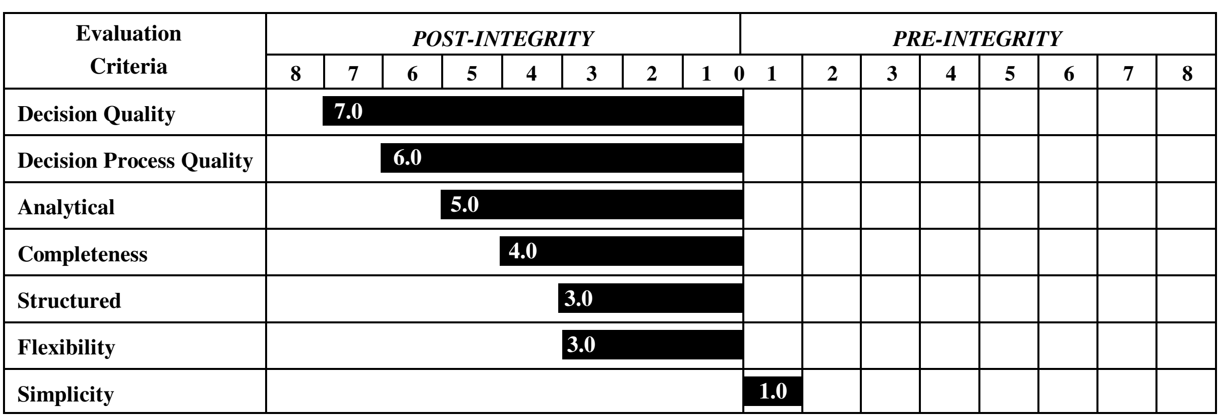

Figure 3 POST-INTEGRITY and PRE-INTEGRITY median scores.

Table 9 *POST-INTEGRITY* test of medians

Evaluation factor	*Median*	*Significant ($\alpha = 0.05$)*
Decision quality	7	Yes
Decision process quality	6	Yes
Analytical	5	Yes
Completeness	4	Yes
Structured	3	Yes
Flexibility	3	Yes
Simplicity	−1	Yes

6. Conclusion

Advances in computer technology and the availability of data have made MCDM more complex than ever. Schoemaker and Russo (1993) argue that as the significance and the complexity of a decision problem increases, so does the importance of the solution quality. While practicing mangers may favour simple approaches, they can be dangerously inaccurate for complex decision problems. Our model helps DMs (i) decompose a complex problem into manageable steps, (ii) ensure the consistency and completeness of the information, and (iii) synthesize the results through a series of logically sound techniques. This decomposition and synthesis is not intended to replace DMs; rather, it provides a systematic approach to support and supplement their judgments. Our model has several attractive features that address some of the limitations inherent in the previous voting method used by the IT:

(i) *Structured:* The model stratifies the information requirements into a hierarchy that simplifies information input and help DMs focus on a small area of the large problem. This process can also help by dividing the problem into different levels of detail for purposes of seeking input from different experts in the organization.
(ii) *Complete:* The model processes a wide range of importance weights and preferences concerning multiple criteria, multiple alternatives, and multiple DMs.
(iii) *Analytical:* The value-analysis model utilized here is considered a MCDM approach.
(iv) *Simple:* The model has a simple to understand and yet theoretically sound scoring system.
(v) *Flexible:* Our model does not limit the number of criteria, alternatives, or DMs.
(vi) *Inconsistency:* Inconsistencies are inevitable while dealing with subjective information from different DMs. The built-in inconsistency-checking mechanism of AHP helps in identifying inconsistencies in judgments at very early stages of the computation process.
(vii) *Sensitivity analysis:* The most important potential benefit of our model is its usefulness in examining how sensitive the results are to changes in the selected parameters.

It should be noted that using the step-by-step and structured approach presented here does not imply a deterministic approach to problem solving. While our model has enabled the IT members crystallize their thoughts and organize their judgments, it should be used very carefully. Team member judgment is an integral component of the model. While these judgments often mirror a team member's belief, they should be used with caution. As with any MCDM, the practicing managers must be aware of the limitations of subjective estimates. When empirical analysis is feasible and makes economic sense, it should be utilized to improve these estimates (Lodish, 1982).

Acknowledgements—This research was supported by NASA Grant number NAG 9-1526. I am grateful to Jack Knight, Steven Gonzales, Anthony Bruins, Terry Tri, Gretchen Thomas, John Park, Kevin Wells, Lac Nguyen, and the *INTEGRITY* Team, all from the Johnson Space Center, and my colleagues Dennis Kennedy and Jack Rappaport for their assistance with this research project.

References

Belton V and Stewart TJ (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers: Boston.
Diakoulaki D, Mavrotas G and Papayannakis L (2000). Objective weights of criteria for interfirm comparisons. *Journées du groupe européen Aide Multicritére à la Décision 36*[e], Luxembourg.
Doumpos M and Zopounidis C (2002). *Multicriteria Decision Aid Classification Methods*. Kluwer Academic Publishers: Boston.
Dyer JS (1990a). Remarks on the analytic hierarchy process. *Mngt Sci* **36**: 249–258.
Dyer JS (1990b). A clarification of 'Remarks on the analytic hierarchy process'. *Mngt Sci* **36**: 274–275.
Expert Choice (2000). *Computer software*. Decision Support Software: McLean.
Forman EH and Gass SI (2001). The analytic hierarchy process—an exposition. *Opns Res* **49**: 469–486.
Grabisch M and Roubens M (1999). An axiomatic approach to the concept of interaction among players in cooperative games. *Int J Game Theory* **28**: 547–565.
Harker PT and Vargas LG (1987). The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Mngt Sci* **33**: 1383–1403.
Harker PT and Vargas LG (1990). Reply to 'Remarks on the analytic hierarchy process' by JS Dyer. *Mngt Sci* **36**: 269–273.
Hwang LC and Yoon K (1981). *Multi Attribute Decision-Making: A Methods and Applications*. Springer-Verlag: Berlin.
Keeney RL and Raiffa H (1976). *Decisions with Multiple Objectives: Preference and Value Tradeoffs*. Wiley: New York.
Lodish LM (1982). Experience with decision calculus models and decision support systems. In: Schulz R and Zoltners A (eds). *Marketing Decision Models*. North-Holland: New York.
Pardalos PM and Hearn D (2002). *Multicriteria Decision Aid Classification Methods*. Kluwer Academic Publishers: Boston.
Pomerol J-C and Barba-Romero S (2000). *Multicriterion Decision in Management: Principles and Practice*. Kluwer Academic Publishers: Boston.
Roy B (1990). Decision-aid and decision making. *Eur J Opl Res* **45**: 324–331.
Saaty TL (1977). A scaling method for priorities in hierarchical structures. *J Math Psychol* **15**: 234–281.
Saaty TL (1980). *The Analytic Hierarchy Process*. McGraw-Hill: New York.
Saaty TL (1983). Axiomatic foundations of the analytic hierarchy process. *Mngt Sci* **32**: 841–855.
Saaty TL (1989). Decision making, scaling, and number crunching. *Decision Sci* **20**: 404–409.
Saaty TL (1990). An exposition of the AHP in reply to the paper 'Remarks on the analytic hierarchy process'. *Mngt Sci* **36**: 259–268.
Saaty TL (1994). How to make a decision: The analytic hierarchy process. *Interfaces* **24**: 19–43.
Saaty TL and Vargas LG (1998). Diagnosis with dependent symptoms: Bayes theorem and the analytic hierarchy process. *Opns Res* **46**: 491–502.
Schoemaker PJH and Russo JE (1993). A pyramid of decision approaches. *California Mngt Rev* **36**: 9–31.
Triantaphyllou E (2000). *Multi-criteria Decision Making Methods: A Comparative Study*. Kluwer Academic Publishers: Boston.
Vanderpooten D and Vincke P (1989). Description and analysis of some representative interactive multicriteria procedures. *Math Computer Modeling* **12**: 1221–1238.
Vincke P (1992). *Multicriteria Decision Aid*. Wiley: New York.
Zeleny M (1982). *Multiple Criteria Decision Making*. McGraw-Hill: New York.

Received November 2004
accepted September 2005 after one revision

Appendix A. Simulator Criteria Weight Assessment Questionnaire

INSTRUCTIONS: Please respond by placing an 'X' on the appropriate location to indicate your perception of the relative importance of the ***INTEGRITY* criteria** for each of the following pairwise comparisons.

Placing an 'X' at equal indicates that criteria A and B are equally important

Placing an 'X' to the left of equal indicates that criterion A is more important than criterion B

Placing an 'X' to the right of equal indicates that criterion B is more important than criterion A

INTEGRITY CRITERION A	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	*INTEGRITY* CRITERION B
	Extreme		Very Strong		Strong		Moderate		Equal		Moderate		Strong		Very Strong		Extreme	
Cost	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Strategic Plan Relevance**
Cost	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Complexity**
Cost	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Science, Operations, and Technology**
Strategic Plan Relevance	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Complexity**
Strategic Plan Relevance	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Science, Operations, and Technology**
Complexity	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Science, Operations, and Technology**

INSTRUCTIONS: Please respond by placing an 'X' on the appropriate location to indicate your perception of the relative importance of the **COST Sub-criteria** for each of the following pairwise comparisons.

Placing an 'X' at equal indicates that sub-criteria A and B are equally important

Placing an 'X' to the left of equal indicates that sub-criterion A is more important than sub-criterion B

Placing an 'X' to the right of equal indicates that sub-criterion B is more important than sub-criterion A

COST SUB-CRITERION A	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	COST SUB-CRITERION B
	Extreme		Very Strong		Strong		Moderate		Equal		Moderate		Strong		Very Strong		Extreme	
Cost to Design	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Build**
Cost to Design	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Operate**
Cost to Design	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Reconfigure**
Cost to Design	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Maintain**
Cost to Build	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Operate**
Cost to Build	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Reconfigure**
Cost to Build	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Maintain**
Cost to Operate	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Reconfigure**
Cost to Operate	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Maintain**
Cost to Reconfigure	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cost to Maintain**

INSTRUCTIONS: Please respond by placing an ‘X’ on the appropriate location to indicate your perception of the relative importance of the **STRATEGIC PLAN RELEVANCE Sub-criteria** for each of the following pairwise comparisons.

Placing an 'X' at equal indicates that sub-criteria A and B are equally important

Placing an 'X' to the left of equal indicates that sub-criterion A is more important than sub-criterion B

Placing an 'X' to the right of equal indicates that sub-criterion B is more important than sub-criterion A

STRATEGIC PLAN RELEVANCE	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	STRATEGIC PLAN RELEVANCE
SUB-CRITERION A	Extreme		Very Strong		Strong		Moderate		Equal		Moderate		Strong		Very Strong		Extreme	**SUB-CRITERION B**
Exploration Development Relevance	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Mission Architecture Commonality
Exploration Development Relevance	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Usefulness Outside *INTEGRITY*
Exploration Development Relevance	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Publicity Value
Exploration Development Relevance	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Mission Requirements Understanding
Exploration Development Relevance	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Risk Mitigation
Mission Architecture Commonality	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Usefulness Outside *INTEGRITY*
Mission Architecture Commonality	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Publicity Value
Mission Architecture Commonality	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Mission Requirements Understanding
Mission Architecture Commonality	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Risk Mitigation
Usefulness Outside *INTEGRITY*	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Publicity Value
Usefulness Outside *INTEGRITY*	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Mission Requirements Understanding
Usefulness Outside *INTEGRITY*	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Risk Mitigation
Publicity Value	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Mission Requirements Understanding
Publicity Value	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Risk Mitigation
Mission Requirements Understanding	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	Risk Mitigation

INSTRUCTIONS: Please respond by placing an 'X' on the appropriate location to indicate your perception of the relative importance of the **COMPLEXITY Sub-criteria** for each of the following pairwise comparisons.

Placing an 'X' at equal indicates that sub-criteria A and B are equally important

Placing an 'X' to the left of equal indicates that sub-criterion A is more important than sub-criterion B

Placing an 'X' to the right of equal indicates that sub-criterion B is more important than sub-criterion A

COMPLEXITY SUB-CRITERION A	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	COMPLEXITY SUB-CRITERION B
	Extreme		Very Strong		Strong		Moderate		Equal		Moderate		Strong		Very Strong		Extreme	
Design & Manufacturing Difficulty	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Operational & Safety Difficulty**
Design & Manufacturing Difficulty	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Sequencing with other Stepping Stones**
Design & Manufacturing Difficulty	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Infrastructure Impacts**
Operational & Safety Difficulty	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Sequencing with other Stepping Stones**
Operational & Safety Difficulty	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Infrastructure Impacts**
Sequencing with other Stepping Stones	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Infrastructure Impacts**

INSTRUCTIONS: Please respond by placing an 'X' on the appropriate location to indicate your perception of the relative importance of the **SCIENCE, OPERATIONS, AND TECHNOLOGY Sub-criteria** for each of the following pairwise comparisons.

Placing an 'X' at equal indicates that sub-criteria A and B are equally important

Placing an 'X' to the left of equal indicates that sub-criterion A is more important than sub-criterion B

Placing an 'X' to the right of equal indicates that sub-criterion B is more important than sub-criterion A

SCIENCE, OPERATIONS, & TECHNOLOGY SUB-CRITERION A	Extreme		Very Strong		Strong		Moderate		Equal		Moderate		Strong		Very Strong		Extreme	SCIENCE, OPERATIONS, & TECHNOLOGY SUB-CRITERION B
	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	
Basic Science Opportunities	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Exploration Science Diversity**
Basic Science Opportunities	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Operations Scenarios**
Basic Science Opportunities	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cross Cutting Technology**
Basic Science Opportunities	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Interfaces/Integration**
Basic Science Opportunities	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Technology Gap/Innovation**
Exploration Science Diversity	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Operations Scenarios**
Exploration Science Diversity	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cross Cutting Technology**
Exploration Science Diversity	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Interfaces/Integration**
Exploration Science Diversity	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Technology Gap/Innovation**
Operations Scenarios	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Cross Cutting Technology**
Operations Scenarios	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Interfaces/Integration**
Operations Scenarios	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Technology Gap/Innovation**
Cross Cutting Technology	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Interfaces/Integration**
Cross Cutting Technology	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Technology Gap/Innovation**
Interfaces/Integration	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	**Technology Gap/Innovation**

Appendix B. Simulator Assessment Questionnaire (Abbreviated Version)

INTEGRITY is designed to provide the building facilities and infrastructure required to house and operate all analog mission vehicle, habitat, and surface terrain simulators and to manage and coordinate all test mission operations associated with these simulators. *INTEGRITY* planners have identified five simulators, all capable of simulating the crew cabin architecture, integrated systems operations, and crew operations associated with a 1000-day class mission transit vehicle:

Transit Vehicle Simulator (TVS)
Lander Vehicle Simulator (LVS)
Surface Habitat Simulator (SHS)
Roving Vehicle Simulator (RVS)
Surface Terrain Simulator (STS)

The purpose of this questionnaire is to capture your perception of the sub-criteria scores for the simulators using a 10-point Lickert Scale (1 = Very Bad…10 = Very Good). For example, when evaluating 'Cost to Design', note that lower cost is preferred to higher cost. If you think a simulator is very costly (bad), provide a low (bad) score and if you think the simulator is not too costly (good), provide a high (good) score. However, when evaluating 'Publicity Value', it should be noted that higher publicity value is preferred to lower value. Therefore, if you think a simulator has a great publicity value (good), provide a high (good) score and if you think it provides little value (bad), provide a low (bad) score.

	TVS	LVS	SHS	RVS	STS
		(1 = Very Bad..........10 = Very Good)			
1.1. Cost to Design (Minimize)					
1.2. Cost to Build (Minimize)					
1.3. Cost to Operate (Minimize)					
1.4. Cost to Reconfigure (Minimize)					
1.5. Cost to Maintain (Minimize)					
2.1. Exploration Development Relevance (Maximize)					
2.2. Mission Architecture Commonality (Maximize)					
2.3. Usefulness Outside *INTEGRITY* (Maximize)					
2.4. Publicity Value (Maximize)					
2.5. Mission Requirements Understanding (Maximize)					
2.6. Risk Mitigation (Maximize)					
3.1. Design and Manufacturing Difficulty (Minimize)					
3.2. Operational and Safety Difficulty (Minimize)					
3.3. Sequencing with other Stepping Stones (Maximize)					
3.4. Infrastructure Impacts (Minimize)					
4.1. Basic Science Opportunities (Maximize)					
4.2. Exploration Science Diversity (Maximize)					
4.3. Operations Scenario Diversity (Maximize)					
4.4. Cross Cutting Technology (Maximize)					
4.5. Interfaces/Integration (Maximize)					
4.6. Technology Gaps/Innovation (Maximize)					

Appendix C. INTEGRITY System Evaluation Questionnaire

INSTRUCTIONS: The following criteria are adopted by the INEGRITY Team to evaluate the simulator assessment system. Please respond by placing an 'X' on the appropriate location to indicate your preference for the *POST*-System or *PRE*-System based on Flexibility, Simplicity, Analytical Capabilities, Completeness, Structured Framework, Decision Process Quality, and Decision Quality.

Placing an 'X' at equal indicates
NO PREFERENCE between Post-Integrity and Pre-Integrity

Placing an 'X' to the left of equal indicates that POST-Integrity is MORE PREFERRED than the PRE-Integrity ⇩ *Placing an 'X' to the right of equal indicates that PRE-Integrity is MORE PREFERRED than POST-Integrity*

	O	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**
	Extreme		Very Strong		Strong		Moderate		Equal		Moderate		Strong		Very Strong		Extreme
	<<*POST-INTEGRITY* System Preferred>>								<< *PRE-INTEGRITY* System Preferred>>								
Flexibility	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**
Simplicity	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**
Analytical.	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**
Completeness	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**
Structured	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**
Decision Process Quality	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**
Decision Quality	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**	O	**O**