A multiple-criteria framework for evaluation of decision support systems

Gloria E. Phillips-Wren\textsuperscript{a,*}, Eugene D. Hahn\textsuperscript{b}, Guiseppe A. Forgionne\textsuperscript{c}

\textsuperscript{a}Department of Information Systems and Operations Management, The Sellinger School of Business and Management, Loyola College in Maryland, 4501 N. Charles Street, Baltimore, MD 21210, USA
\textsuperscript{b}Department of Information and Decision Sciences, Perdue School of Business, Salisbury University, Salisbury, MD 21801, USA
\textsuperscript{c}Department of Information Systems, University of Maryland Baltimore County, 1000 Hilltop Circle, Catonsville, MD 21250, USA

Received 19 June 2003; accepted 8 January 2004

Abstract

In the literature, decision support systems (DSSs) have typically been evaluated on only a single criterion such as the outcome from decision making. However, it is clear that DSSs simultaneously have a critical impact on the process-oriented aspects of decision making, suggesting that a combination of both outcome and process criteria are highly relevant for DSS evaluation. Indeed, process characteristics are particularly crucial for web-based and real-time DSSs because of their ability to deliver timely, current information through features such as just-in-time information, real-time processing, on-line transaction processing, connectivity and globally up-to-date information. In this underexplored area, we propose a framework to evaluate DSSs that combines outcome- and process-oriented evaluation measures. The approach is demonstrated in the context of a real-time threat criticality detection DSS. Investigations are conducted using a multicriteria decision-making method called the Analytic Hierarchy Process (AHP) as well as a newly developed stochastic enhancement of AHP. We find that the real-time DSS offered a significant improvement in terms of process-related characteristics. However, it did not offer a statistically significant improvement in terms of outcome-related characteristics. The importance of simultaneously addressing both sets of considerations is discussed.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Decision support system; Evaluation; Analytic hierarchy process; Stochastic; Real-time

1. Introduction

In a study of 21 reported decision support systems (DSSs), Forgionne [1] found that most empirical studies of traditional DSSs focused on either process-oriented or outcome-oriented evaluation measures, but not both in the same study. Of studies that utilized multiple measures, outcome- and process-oriented measures were presented as individual values with no integrative assessment of the overall value of the DSS. To illustrate, we consider two DSSs in which the authors report several evaluation measures. Humphreys et al. [2] developed a DSS for strategic sourcing to assist companies in the make or buy decision. They claimed that firms do not have formal methods for evaluation of the make versus buy decision and often make the decision on the basis of a single criterion such as overhead costs. They used multi-attribute analysis to assess the trade-offs between make or buy. Rasmussen et al. [3] developed a DSS to characterize airport icing conditions to assist the decision of airplane deicing, and it was evaluated on the basis of user response. Both studies suggest domain-specific measures that are both process-oriented and outcome-oriented as categorized in Table 1 to evaluate the DSS, where the
Table 1
DSS studies using both outcome and process measures for evaluation

<table>
<thead>
<tr>
<th>Researchers</th>
<th>Date</th>
<th>Evaluation measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humphreys et al.</td>
<td>2002</td>
<td>Monitoring suppliers against performance benchmarks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Providing feedback to suppliers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improving cooperation between multifunctional purchasing team members</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reducing the time involved in conducting the make or buy evaluation</td>
</tr>
<tr>
<td>Rasmussen et al.</td>
<td>2001</td>
<td>Improving situational awareness and ability to anticipate storm conditions at the airport</td>
</tr>
<tr>
<td></td>
<td></td>
<td>More timely decision making</td>
</tr>
</tbody>
</table>

classification into process or outcome measures represents the opinion of the current authors and not those of the original researchers.

This paper proposes that a combination of both outcome and process criteria are highly relevant for DSS evaluation such as those in Table 1, and that criteria can be simultaneously considered with a multicriteria decision-making method called the Analytic Hierarchy Process (AHP) [4] as well as with a newly developed stochastic enhancement of AHP. Our motivation for this research is to answer the following questions:

Can a framework be established to provide a single evaluation measure for the value of a DSS when multiple criteria are utilized? Can the criterion (or criteria) that most significantly contributes to the value of the DSS be identified? Can we determine if the DSS offers a statistically significant improvement over the alternative?

2. Background

The two major classes of performance measures for evaluating DSSs given by Turban and Aronson [5] are effectiveness and efficiency. Effectiveness is concerned with decision outputs, while efficiency measures the use of resources to achieve those outputs [5]. Forgione [1] suggested that evaluation of a DSS should simultaneously consider both of these types of metrics, and he grouped them under the two categories of process and outcome measures.

The process of decision making involves the procedures and steps that a decision maker utilizes in making a decision. The classical model of the decision-making process was provided by Simon [6] and has three phases of intelligence, design and choice as shown in Table 2, where the fourth phase of implementation is a more recent addition [5]. The phases generally proceed in order, although the decision maker may perform some phases concurrently or loop back to previous phases to refine the problem.

The phases can be expanded to provide the steps in the decision-making process as shown in Table 3 [1]. The steps are not necessarily sequential, and the decision maker may loop back, repeat steps and modify results from previous steps before proceeding. A DSS assists the decision maker with a particular decision problem and rarely supports all the
steps or phases of decision making. Therefore, the evaluation of the DSS would include only the relevant metrics.

The third component that could be considered as part of the process of decision making is changes in the organization or decision maker. For example, communication or morale could improve as a result of the process of decision making. The decision maker could become more proficient by reducing the time required for a decision or displaying more insight into the decision problem. These changes are a result of the decision making process rather than an outcome of the decision supported by the DSS.

The premise of our research is that DSSs affect both the process of, and outcomes from, decision making. The outcomes of decision making include organization performance measures such as reduced cost, increased profit, enhanced competitive position, or improved capital budgeting measures such as return on investment. Outcome measures are clearly problem-dependent and could include some of the metrics that are shown for organizational changes.

The consideration of multiple criteria is particularly relevant for web-based and real-time DSSs [7]. Such systems combine traditional decision-making support technologies with the benefits of advances such as the World Wide Web. They enhance and extend DSSs by providing features such as just-in-time information, real-time processing, on-line transaction processing, connectivity and globally up-to-date information [8–13]. Not only is the outcome of the decision-making process potentially affected by factors such as timely information available from the web or electronic sensor packages, but the process of decision making is affected as well. For example, the system described by Jensen et al. [8] provided advice in real-time that not only improved the outcomes from crop management, but also delivered personalized web pages with embedded graphics and links to additional information. The authors reported that the use patterns were different between subscriber types, indicating that the process of decision making was impacted by the web-based DSS.

We propose that multiple evaluation criteria based on both process and outcome metrics are needed to assess a DSS. We further propose a framework that combines these criteria to provide a single evaluation metric for a DSS and that is capable of discerning if the DSS offers a statistically significant improvement over the alternative. The AHP is utilized for the evaluation metric, and a newly developed methodology called stochastic AHP provides a framework for allowing us to statistically distinguish among the alternatives [14].

The paper is organized as follows. We first present the framework for DSS evaluation. A real-time DSS that utilizes wireless communication between an autonomous robotic system and a remote human decision maker is then presented. DSS effectiveness measures for the problem domain are examined in the next section. An evaluation is then discussed that considers both the outcomes from, and the process of, decision making by comparing decision making with and without the DSS from a multiple-criteria evaluation model implemented with the AHP. In the next section the results are analyzed with stochastic AHP to determine if the DSS is statistically different from the alternative. Finally, the work is summarized, and contribution to the literature is discussed.

3. Model framework for evaluation of DSSs

DSS-effectiveness measures can be developed to include both the outcomes from, and the process of, decision making [1]. Grabowski and Sanborn [15] offered a similar viewpoint during their evaluation of software embedded in larger host systems by concluding that evaluation criteria should include both technology and human-oriented dimensions. Parikh et al. [16] used four evaluation criteria—decision quality, decision maker satisfaction, decision maker learning, and decision maker efficiency—to evaluate decisional guidance effectiveness, and these include both process and outcome variables. A model framework for the evaluation is shown in Fig. 1, where the specific measures under process and outcome are individualized to the decision problem domain.

It is possible to associate these criteria using a multiple-criteria evaluation model of DSS effectiveness implemented with the AHP [1,17]. The AHP provides a logical and scientific basis to decision making in which pairwise comparisons of components are made with respect to a common goal or criteria [18]. The decision problem is structured as a hierarchy of criteria, subcriteria and alternatives, with the number of levels being determined by the problem. Once the hierarchy is established, the alternatives are evaluated in pairs with respect to the criteria on the next level. The criteria can be weighted, if desired, according to the priority of each criterion [18]. The AHP is widely used in individual and group decision making scenarios [19].

Applying the AHP to DSS-effectiveness evaluation, there are two criteria: the process of decision making and the outcomes from decision making. Outcome improvements could come from gains in organizational performance such as increased sales or decreased costs. Process improvements could involve gains in the user’s ability to perform the phases and steps of decision making, or increased efficiency and productivity. The process and outcomes criteria in the DSS-effectiveness evaluation can have further subcriteria in the AHP model as dictated by the decision problem.

Measures and subcriteria to evaluate DSS effectiveness under the AHP structure have been suggested. For example, Mirani and Lederer [20] comprehensively studied the benefits of IS projects (as opposed to purely DSSs) from the literature and suggested three outcome categories: strategic, informational, and transactional. Strategic benefits were further refined to competitive advantage, alignment and customer relations. Informational benefits were comprised of information access, information quality and information flexibility. Transactional benefits were communications efficiency,
systems development efficiency and business efficiency. This taxonomy could be applied to AHP subcriteria for a DSS as well.

4. Real-time DSS

Autonomous, mobile robotic systems are beginning to emerge from the pure research phase into applied technology. A recent example is a robotic nurse called Pearl under development at Carnegie Mellon University [21]. Pearl is designed to escort nursing home residents to appointments, remind them to take pills, and check on them. Pearl is reported to have passed the first test by escorting residents at a local nursing home to their physical therapy appointments. As another example, using a sophisticated sensor suite and software architecture, a robotic system is able to serve as a scout by navigating on and off roadways to obtain information about its environment and relay the information to a human using wireless communication channels [22–24]. Such systems operate in real-time with current information, and their capabilities could be extended with a DSS placed on-board the robotic system or with a human user who can communicate with the system. One example is a DSS for threat assessment developed for an autonomous robotic system [13]. Threat assessment currently requires direct intervention and judgment by personnel who are remotely located from the robotic system and who must rely on information such as visual images provided by the sensor systems and transferred via wireless communication to a human who acts as the decision maker concerning the appropriate response.

One approach to mitigating the need for bandwidth and time to transfer information between the robotic system and the human is to embed a DSS within the robotic system to determine the threat criticality [13]. Such a system utilizes the sensor data in real-time to determine econometrically estimated models and forecast the threat criticality from an external source. Using intelligent agent technology, the DSS is able to update data and perform functions autonomously, reducing the need for communication with the remote human. Such real-time processing is similar to the timeliness of web-based DSSs. In both cases, the outcomes from decision making are expected to be affected by the DSS since data are more current and relevant.

The process of decision making using the real-time DSS is affected as well. One description of the process of decision making utilizes the generally accepted phases of decision making: intelligence, design, choice and implementation [5,6,25]. The process is continuous, the phases may overlap, and the decision maker may loop back to a previous phase during decision making [6,26]. Decisions may be further influenced during the decision-making process by the anticipation of a review that provides some control and evaluation of the decision [6]. In web-based and real-time DSSs, the process of decision-making is affected by the methods used within the DSS itself. For example, the intelligence phase may be modified by the methods of extracting the data from the WWW. An alternate description of the process of decision making is to break the phases down into the steps of decision making such as identifying objectives, recognizing the problem, gathering data, generating alternatives, establishing criteria, evaluating alternatives, making a final choice, and implementing the final choice [1]. In the real-time DSS in our application, for example, the data-gathering step is affected by the implementation of the DSS.

5. DSS-effectiveness evaluation

5.1. Experimental design

A DSS developed for an autonomous robotic system to assist in the threat criticality decision was encoded in SAS
Version 8 and evaluated with a multiple-criteria model implemented in the AHP [13]. The two alternatives of decision making with and without the DSS, i.e. DSS and No-DSS, were evaluated using the threat assessment decision hierarchy shown in Fig. 2, where the criteria used were

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>REC</td>
<td>time to recognize problem</td>
</tr>
<tr>
<td>QUA</td>
<td>amount of data, i.e. number of variables, considered in TC decision</td>
</tr>
<tr>
<td>GEN</td>
<td>number of TC alternatives generated for one scenario</td>
</tr>
<tr>
<td>CRT</td>
<td>time to establish criteria, i.e. develop a model</td>
</tr>
<tr>
<td>EVA</td>
<td>time to evaluate alternatives</td>
</tr>
<tr>
<td>FIN</td>
<td>time to choose final TC</td>
</tr>
<tr>
<td>DEC</td>
<td>time to make one TC decision</td>
</tr>
<tr>
<td>ALT</td>
<td>number of TC decisions made in a given time period (1 min)</td>
</tr>
<tr>
<td>ATL</td>
<td>closeness to actual TC</td>
</tr>
</tbody>
</table>

where TC is the threat criticality criterion.

The experiment tested whether the DSS could improve the outcomes from, and the process of, threat assessment decision making compared to No-DSS. After an initial training session with the training scenarios, the decision maker was presented with data for a threat assessment scenario. The decision maker used the available information together with his experience, judgment and knowledge to determine the threat criticality. Subsequently, the decision maker utilized the DSS to determine the threat criticality. Since the DSS has been constructed to determine the threat criticality autonomously, no bias is introduced by using the same decision maker for the No-DSS cases and DSS cases.

5.2. Performance results

The results from the No-DSS and DSS cases for the criteria in the decision hierarchy are shown in Table 4. The data in Table 4 were obtained from a user who was experienced with the application area, but not an expert in the application domain. The user viewed a 21” screen on a 750 MHz Pentium® III computer. Oral and written definitions of the data variables were provided. A data training set was used to familiarize the user with the data format,
and the data were interpreted relative to the variables. The measures for the DSS-effectiveness model in Table 4 were explained. For the No-DSS case, the user’s time was recorded for REC, CRT, EVA, FIN, DEC and ALT. The user provided the values for QUA and GEN immediately after the session.

A rough estimate of the timing for the variables with the DSS was developed by using the clock cycles for the Pentium® chip, estimating the number of operations needed to complete the task, and using the computer speed of 750 MHz. The latency due to the DSS software operation is included in ALT. QUA was obtained from the number of variables considered in the model, and GEN is the number of models in the model base. Although the times for the DSS case are gross estimates, they provide adequate information for the AHP analysis. A study of imprecision in pairwise comparisons in the AHP showed that even under highly imprecise cases, the relative ranking of the alternatives was unchanged [27]. The imprecision in the estimates in Table 4 is not expected to change the ranking of the No-DSS and DSS alternatives.

The numerical values for the criteria shown in Table 4, with equal weights, were evaluated with the AHP model shown in Fig. 2 to determine the AHP priority. The values in Table 4 were utilized to determine a numerical rating of the factors for the two alternatives. The values of the criteria were compared with a numerical preference scale, and the specific pairwise comparisons for the No-DSS and DSS cases for the criteria used in the AHP analysis are also shown in Fig. 2.

Fig. 2 also shows the values of the nodes obtained with the AHP analysis. Personal productivity and personal efficiency are significantly enhanced with the DSS compared to No-DSS. Step proficiency is greatly improved with the DSS, and organizational performance is improved as well. The DSS resulted in improvement in both the process of, and outcomes from, decision making, with the process showing the greatest improvement. Overall, a priority of 0.242 was obtained for the No-DSS case compared to a value of 0.758 for the DSS case, indicating that the DSS improved decision making. The results demonstrate that a consideration of both outcome and process simultaneously may provide an improved evaluation of real-time decision support systems.

6. Empirical analyses

The AHP priorities appearing in Fig. 2 indicated that the DSS was associated with enhancements to both the process of and the outcomes from decision making. However, the improvement associated with the DSS relative to No-DSS varied considerably as a function of the factors in the model. Indeed, if the focus were to be placed exclusively on the GEN criterion, DSS was actually less preferable than No-DSS. Thus, a question naturally arises: does the DSS offer a statistically significant improvement over the alternative? In previous years, this type of question was unanswerable. However, this type of question can now be answered with a newly developed methodology called stochastic AHP [14].

We review the essentials of stochastic AHP briefly as follows. Priorities, such as those displayed in Fig. 2, are the result of a series of pairwise judgments in which two alternatives are compared on a particular criterion. Each pairwise judgment can further be decomposed into two weights, one for each of the pair of alternatives under consideration. For example, we can see that the priorities of No-DSS versus DSS with respect to the ATL criterion are 0.333 and 0.667, respectively. This corresponds to a relatively mild 2:1 preference for DSS over No-DSS.

It can be shown that the two weights of the rth series of pairwise comparisons follow a binomial distribution. That is,

\[(w_{1r}, w_{2r}) \sim \text{Binomial} \left( \sum_{k=1}^{2} w_{rk}, p_{rk} \right).\]

Here, \(p_{rk}\) is the vector of priorities such that \(\sum_{k=1}^{2} p_{rk} = 1\). We take advantage of the fact that the \(p_{rk}\) quantities are binomially distributed in order to determine whether or not the priorities are different from one another. In the case where the number of alternatives \(K\) is greater than two, the method can be extended via a multinomial generalization (see [14]).

In AHP, the pairwise comparisons are arranged into a judgment matrix. The decision maker’s priorities (or underlying weights for the alternatives) are obtained via eigenvalue decompositions on the collection of judgment matrices. The sub-diagonal entries in the judgment matrices are the reciprocals of the entries above the diagonal. Thus, the sub-diagonal entries are redundant. However, it can be shown that the redundancy can be eliminated by an appropriate system of weighting the judgments. Then a weighted binomial (or multinomial) logit model can be used to allow one to examine inferential questions regarding the priorities. In [14] it is shown that an appropriate
weight is
\[ Q = \frac{K^2 + K - 2}{2K^2}. \]
Thus, the weighted formulation is
\[
(w_1, w_2) \sim \text{Binomial} \left( Q \sum_{k=1}^{2} w_k, p_k \right),
\]
where \( Q = 1/2 \). Either marginal or hierarchical formulations of the weighted logit model can be used when \( K > 2 \), but in the present case where there are two alternatives, the marginal formulation is appropriate.

Since the sampling distribution of the priorities is specified by a probability distribution (the binomial), we have a parametric model. Here interest centers on the priority parameters, \( p_k \). In particular, we would like to determine whether the overall priority of one of the alternatives (e.g., DSS) is significantly different from the overall priority of the other alternatives (e.g., No-DSS). We examine this issue by calculating a difference score for the priorities, and then determining whether the probability interval contains zero. If this interval does not contain zero, then we may reject the null hypothesis that the two alternatives have an equal priority with respect to a particular criterion.

Note that a number of non-branching nodes appear in Fig. 2 for clarity of exposition. Nodes that do not branch can effectively be collapsed over in both AHP and stochastic AHP. This was done here to facilitate the empirical analyses.

In the current study, inference was conducted via Bayesian Markov chain Monte Carlo methods. Relatively flat but proper priors were used for all parameters. Specifically, normal priors with means of zero and variances of 100,000 were used for all parameters. These priors were effectively uniform across all plausible parameter values. As the binomial logit model is well-behaved, the convergence of the Markov chain was rapid. Evidence for this contention appears in Fig. 3. Fig. 3 contains trace plots of the Markov chain for the four significant parameters (the non-significant parameter essentially conveyed the same information). At the left of each trace it is possible to discern a very small tail that begins at the chain’s initial value of zero. The Markov chain drifts downward for a few hundred iterations and then very quickly converges to its final stationary distribution before the 500th iteration. In spite of this quick convergence, the chain was allowed to burn in for a total of 5000 iterations. These iterations were discarded and the chain was run for an additional 30,000 iterations. Inference was based on these latter 30,000 samples.

6.1. Step proficiency

We first examined DSS and No-DSS with respect to the criterion of step proficiency. The stochastic AHP model estimated the priorities of No-DSS and DSS as 0.249 and 0.751, respectively. These were the same priorities as those provided by AHP. In general, the priorities that arise from stochastic AHP will have the same value as those from AHP, excluding possible trivial differences at the third or fourth significant digit due to Monte Carlo error. Thus, we do not report the estimates of the priorities further, as these values are provided in Fig. 2.

We may determine whether these two priorities are different from one another using the approach specified as follows. First we construct a difference score, \( \delta \), which is the difference of the two priorities. We then obtain the 95% posterior probability interval for \( \delta \). Note that this 95% probability interval has a broadly similar interpretation to the 95% confidence interval. Specifically, if the 95% probability interval does not contain the value zero, then we can conclude that the two priorities are different from one another at the 95% level.

With respect to step proficiency, the estimate of the difference score, \( \delta_p \), was \(-0.502\) (see Table 5). The standard deviation of \( \delta_p \) was 0.115 and the 95% probability interval of \( \delta_p \) was \(-0.719\) to \(-0.270\). Thus, we find empirical support for the notion that step proficiency is significantly improved with the DSS.

6.2. Personal productivity and personal efficiency

In analyzing the personal productivity judgments, we find the DSS strongly associated with enhanced personal productivity. Here, the estimate of the difference score, \( \delta_{pp} \), was \(-0.703\). The standard deviation of \( \delta_{pp} \) was 0.175 and the 95% probability interval was \(-0.994\) to \(-0.351\). Thus, the difference score was significantly different from zero at well beyond the 95% probability level. The judgments for personal efficiency were the same as personal productivity. As such, the resulting estimates of the difference score for personal efficiency were the same as those for personal productivity.

6.3. Process and outcome

The analyses showed that the DSS was also favored when one aggregates over all of the process-related considerations. The estimate of the overall difference score for process considerations, \( \delta_p \), was \(-0.703\). The standard deviation of \( \delta_p \) was 0.085 and the 95% probability interval was \(-0.841\) to \(-0.507\). By contrast, the DSS was favored by

\[ 1 \text{ A careful reading of the above figures will seem to suggest that the standard deviation and the 95% probability interval are slightly discordant. This result occurs because here the distribution of the } \delta_p \text{ is not perfectly symmetric, but is somewhat skewed toward zero.} \]
Table 5  
Stochastic AHP parameter estimates of DSS effectiveness

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard deviation</th>
<th>2.5% percentile</th>
<th>97.5% percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>δp</td>
<td>−0.502</td>
<td>0.115</td>
<td>−0.719</td>
<td>−0.270</td>
</tr>
<tr>
<td>δpp</td>
<td>−0.802</td>
<td>0.175</td>
<td>−0.994</td>
<td>−0.351</td>
</tr>
<tr>
<td>δp</td>
<td>−0.703</td>
<td>0.085</td>
<td>−0.841</td>
<td>−0.507</td>
</tr>
<tr>
<td>δo</td>
<td>−0.330</td>
<td>0.469</td>
<td>−0.974</td>
<td>0.687</td>
</tr>
<tr>
<td>δmodel</td>
<td>−0.517</td>
<td>0.239</td>
<td>−0.859</td>
<td>−0.006</td>
</tr>
</tbody>
</table>

a non-significant margin with respect to outcome-related considerations. The estimate of the overall difference score for outcome considerations, δo, was −0.330. However, the standard deviation of δo was 0.469. This resulted in a broad 95% probability interval, one which ranged from −0.974 to +0.687.

6.4. Overall model

In general, the DSS was favored overall. However, this overall judgment of favorability resulted from a combination of outcome-related considerations, where the DSS did not provide a significant improvement, and process-related considerations, where the DSS provided a highly significant improvement. This result is unexpected, and one would expect a concomitant improvement in outcome. As such, it is of interest to estimate the difference score for the overall model, δmodel. The estimate of this quantity was −0.517, indicating the DSS was significantly preferred at the overall level. The standard deviation of δmodel was 0.239 and the 95% probability interval was −0.859 to −0.006. Thus, we can be 95% confident that the DSS offered an improvement over the alternative. For reference purposes, Fig. 2 displays the model, specifies the location of the parameters with respect to the model, and indicates which model parameters are statistically significant.

7. Summary and contributions to the literature

This research provides a fundamental framework for evaluation of decision support systems and a methodology for multicriteria evaluation. Secondly, we have developed a model to identify the criterion (or criteria) that most significantly contributes to the benefit (or lack thereof) of the DSS. Thirdly, we have demonstrated a method of determining if the DSS offers a statistically significant improvement over the alternative.

The use of multiple criteria to evaluate DSSs is essential since both the process of, and the outcomes from, decision making are affected. In particular, the increasing use of web-based and real-time DSSs suggests that both types of metrics are relevant to represent the value of a DSS. Web-based and real-time DSSs deliver timely, current information by providing features such as just-in-time information, real-time processing, on-line transaction processing, connectivity and globally up-to-date information. We
have proposed a framework to evaluate DSSs that combines outcome- and process-oriented evaluation measures using the Analytic Hierarchy Process, and the statistical difference between a DSS and the alternative can be statistically determined using a newly developed analytic technique called stochastic AHP. The framework was illustrated with the evaluation of a real-time DSS, and we suggest that the evaluation framework offers an improvement over traditional single measures.

The AHP was chosen as the multicriteria model due to its capability to quantify and rank the alternatives using simple pairwise comparisons of criteria [18]. The model allows the user to exercise judgment in both the selection of criteria and the relative ranking of the alternatives within each criterion, and the AHP has demonstrated robustness across a range of application domains [28]. Its application to evaluation of DSSs provides an accessible methodology for the practitioner.

With regard to the DSS application described as an example in this paper, a certain limitation may be discerned. Our application is a single experiment employing a single user taken from a population of users. However, it is possible that a different user would evaluate the DSS and the No-DSS cases differently; thus, different conclusions could be reached. Further evaluation studies of threat-criticality DSSs need to be conducted to confirm or refute the specifics of the empirical findings pertaining to the DSS application described in this paper. Nonetheless, we are proposing that the AHP can evaluate DSS effectiveness from sample applications, with our application offering an illustration. Most importantly, the framework and the specific evaluation criteria can be applied to a variety of DSSs to determine the usefulness of the single metric in comparing and evaluating DSSs.

Acknowledgements

The authors would like to express their sincere appreciation to the anonymous reviewers whose thoughtful comments greatly contributed to the quality of the manuscript.

References


