AN EXPERT SYSTEM PROTOTYPE FOR PERFORMANCE APPRAISAL OF MANAGERIAL COMMUNICATION

James S. Moore
Indiana University at Fort Wayne

A vital factor in the development of an organization's human resource is the performance evaluation, yet few management activities create the potential level of anxiety among all participants as does the process of performance appraisal of managers by their superiors. Common criticisms include favoritism, excessive subjectivity, and inconsistency among others. This paper examines the appropriateness of expert system technologies in evaluating one segment of managerial performance, that of communication skills. The expert system allows for the integration of the qualitative and quantitative aspects of performance assessment, while introducing the beneficial dimensions of increased objectivity, comprehensiveness, and consistency both across individuals and through time. A prototype expert system is offered that enables the non-expert end user manager to conduct a performance appraisal of a subordinate manager's communication proficiency. The system's early validation results hold promise of proving effective as an evaluation aid while potentially mitigating anxiety for all participants.

A critical factor in the proper use and development of an organization’s human resource is the performance appraisal. The search for objective tools that would enable performance evaluations to be made systematically and on a consistent basis can be traced to Bittner (1948). Despite the length of this noble quest, few contemporary management activities create the potential level of anxiety, among all participants, as does the process of performance evaluation of managers by their superiors. Major contributors to this anxiety creation are 1) the multiplicity of purpose inherent in any appraisal system that seeks to serve both the organization and its employees, and 2) concerns with the consistency of the assessment process, both through time and across participants.

McGregor (1960) summarized the diverse purposes of appraisals as:

1) Administrative: Appraisals permit an orderly and rational way of determining promotions, salary increases, transfers, terminations, and the like,

2) Informative: Appraisals supply data to management on the performance of subordinates, and to the individual regarding their superior’s perception of their strengths and weaknesses relative to their job expectations, and
3) Motivational: Appraisals create a learning experience for subordinates that motivates them to improve their job performance.

Such worthy, yet multiple dimensions of purpose may contribute to varying perspectives of the evaluation process by the participants, and thus be a source of anxiety concerning the appraisal that needs to be resolved through confidence in the objectivity and consistency of the process itself. Unfortunately, such confidence by the participants is not naturally inherent in the process. The importance of comprehensive, objective, and consistent performance assessments grows as the organization’s human resources become more costly and more concerned with professional development (DeNisi & Stevens, 1981).

Not only does the appraisal process mirror the individual’s performance, but the process can be a major influence on an individual’s future performance and satisfaction. When performed objectively, accurately, and consistently, performance assessments will create an aura of fairness about the entire evaluation process and will give the managerial subject the proper signals with respect to interpreting the performance expectations of the superior. Yet, if performed within an atmosphere of imprecise expectations and standards, excessive subjectivity and/or inconsistency, they risk the encouragement of only short-term performance at the expense of long-term planning, the development of bitterness, rivalry and politics at the expense of teamwork. Moreover, not only does such an appraisal process necessarily bear a high opportunity cost in time and energy for both the rater and ratee, but poorly designed or instituted systems may create considerable distrust by the subjects as well as discomfort for those forced to sit in judgement of others.

A comprehensive assessment of managerial performance would include evaluations for numerous behavioral dimensions including, but not limited to: Job Knowledge, Problem Solving Ability, Staff Management, Other Resource Management, Service to Others, Communication Skills, Effectiveness of Results, and Attitude and Enthusiasm. Each dimension could be evaluated separately, with an aggregate assessment assembled from such components. The intent of this study is to investigate the appropriateness of expert system technologies in evaluating one such segment of managerial performance, that of “Communication Skills”.

The next section reviews current evaluation techniques as well as efforts to incorporate traditional Computer Based Information Systems (CBIS) into the process. The paper then seeks to create an awareness of the potential contributing role of an expert system. The prototype model is then developed, and finally, initial deployment issues are discussed.

Literature Review

Smith (1986) argues that most relatively successful performance appraisal systems incorporate detailed, accurate periodic job descriptions, specific performance criteria, periodic reviews, and rater training. Assessment of managerial performance is complicated by characteristics of the positions to be evaluated, often typified by non-routine, unprogrammable work, and the absence of a continual stream of identifiable outputs. In such managerial/administrative environments, where outputs are not directly measurable, performance appraisals can assess behaviors as surrogates for outputs. Such behaviors constitute the processes which contribute to output production and goal attainment. In principal, the rater observes behaviors that are job relevant, evaluates those behaviors, and weights the evaluations to arrive at a final rating.

Most performance appraisal methods adopt either a “traditional” approach based on analytic tools or a “collaborative” approach based on goals of personal growth and development. When the intent of the evaluation is to
serve as the basis for employee assignments, promotions, or merit increments, then the quantitative appraisal would better withstand formal challenges than would the more qualitative collaborative approach. Yet, when the emphasis is on communications, matching expectations between management levels, and goal setting, all of which contribute directly to the development of more effective managers, then the collaborative or joint goal setting approaches are recognized as being more effective. Schuler and Youngblood (1986) and Martin (1986) summarize major approaches to performance appraisal; the former focuses on the form of benchmark mechanism, while the latter categorizes appraisal systems, based on the process itself. Empirical examinations of case studies involving various participatory systems suggest that a formal combination of goal-setting and performance self-appraisal may raise productivity by as much as 16 percent (Waldrop, 1986; Gomez-Mejia, Page, & Tornow, 1985).

Historically, performance evaluation systems have not been highly regarded by managerial employees. Typical criticisms include favoritism, subjectivity, and inadequate training of the evaluators, (Fox, 1987-88; Friedman, 1987; & Altany, 1987). Much concern is often placed on the objectivity of the appraisals. Friedman (1987) identified two major roadblocks to a truly objective evaluation of employee performance to be measurement imprecision and managerial bias. Most ratees, and many raters, are often uncertain about what is being measured and how it is being measured. Consistency is another potential problem in performance appraisals. The rating of an individual’s performance can vary with different raters, with different units of the same organization, and even with the same rater over time. The case has been made by Nathan (1985), that the way raters process behavioral information has much more effect on the evaluation process than the behaviors themselves.

Additionally, there are legal pitfalls involved with the performance appraisal process, (Romberg, 1986; Martin, 1986). Performance appraisals are considered to be tests and are subject to the Equal Employment Opportunity Commission (EEOC) regulations. Improperly designed appraisal systems that are not strictly work related, do not adequately reflect the workforce, or can be perceived to discriminate can quickly precipitate litigation against the employer. Conversely, a sound, formally structured, and properly administered appraisal system can be a very valuable source of documentation to counter such challenges. Using survey responses from the Fortune 1000 population, Taylor and Zawacki (1984) find a significant shift during 1976 to 1981 toward the formal analytical methods. This shift was seen as a necessary response to the growing legal challenges to the prior qualitative systems; yet at the same time, they document adopters’ diminishing satisfaction with their evaluation system.

Behavioral research has long focused on methods for improving the quality of the performance evaluation process. Some studies critique the rating instrument, others examine the characteristics of the participants, while still others investigate the process itself. Jolly et al (1988) examine the evaluator’s own value system as it impacts the performance assessments of others. Potential judgement errors and biases of the rater will detract from the accuracy of the assessment. In addition to obvious biases resulting from overt discrimination, there are unintentional biases where individual differences among raters, such as age, race, personality type, and sex may dispose them to different perspectives on the observed behavior of the subject.

Combining with such dispositional factors to compromise the consistency of the appraisal are contextual factors associated with the assessment process itself. Recall of an individual’s prior rating may serve as a cognitive anchor for the evaluator and thus influence the current assessment. Often the current rating is insufficiently adjusted away from the anchor
Order bias exists when subjects evaluated early in the assessment process receive different ratings than those with similar objective performance that are simply rated later. This is most prevalent when evaluators initially lack specific benchmarks and must evaluate several subjects before acquiring a relative perspective of performance (Huber et al., 1987; Wagner & Hoover, 1974). Huber (1989) concludes that specific performance standards can reduce the rating bias of both contextual & dispositional types.

Locke & Latham (1984) contend that formal scoring algorithms help raters unite information about performance and improve the accuracy of overall ratings. Efforts to incorporate computer based information systems into the evaluation process include software intended to make the reward process largely an automated task, (Cederblom, 1988). This system was successfully tested in a discrimination suit; where the court recognized the logic and objectivity of the procedures used. Edwards and Verdini (1986), develop a test around which a human performance measurement system can be designed with multiple raters. They then offer a computer program which not only combines the ratings but also detects biases among the raters. Davis and Mount (1984b) investigate the effectiveness of computer assisted instruction for training those who will conduct performance appraisals. A structural context for automating staff performance ratings was built using dBASE III by Tobin (1985). In addition to an overall rating, her system provides a profile of an employee’s strengths and weaknesses in a disciplined and rigorous fashion.

Such early efforts to employ some form of CBIS often permitted the pendulum to swing from the purely judgmental to the purely analytical. Schneier, Beatty, and Baird (1986) reiterate that any appraisal requires judgement and any attempt to make all measures purely objective is potentially unrealistic and often irrelevant. Yet, as long as performance appraisals of administrative personnel are exclusively dependent upon human evaluators, the potential for bias and inconsistency remains a serious concern. This study views these approaches as other than mutually exclusive. Moreover, through the development of an expert system, it is proposed that the formality of the analytical approach may be integrated with the qualitative perspectives to yield a more comprehensive, objective and consistent assessment of managerial performance.

The appraisal process must be part of a program beginning with performance planning in which expectations, goals, responsibilities, and the ultimate measurement criteria are explicitly laid out. Specific performance standards enable performance to be rated without being directly observed as outcomes can be evaluated against the standards. Between annual reviews, the subject should be given informal follow-up in the form of contemporaneous feedback on results obtained. When these communication steps are conscientiously carried out, the actual process should generate less stress, and the final assessment fewer surprises, for the participants.

While admittedly difficult to focus on one dimension of such a program to the exclusion of the others, this study will assume that a policy of performance planning and interim follow-ups is in place, and examine the potential role of computer assisted systems for evaluation of managerial communication performance.

**Potential Role of an Expert System**

Performance ratings of managers combine multivariate information. Managers are not necessarily good at integrating/synthesizing multidimensional performance data in order to construct a composite assessment, due to their vulnerability to numerous contaminants in the appraisal process (Nisbett & Ross, 1980; Northcraft, Neale & Huber, 1988; Feldman, 1981; Huber, Neale & Northcraft, 1987). As previously discussed, such extraneous factors may include inappropriate considerations (i.e.,
the “halo” effect), contextual and/or dispositional bias of the evaluator. At best, such contaminants may preclude the evaluation from yielding much useful and constructive information to the ratee; all too often, such compromises to objectivity and consistency in the process can create an atmosphere of anxiety, distrust, and even dread for not only the ratee but the evaluator as well. Any realizable improvement in the psychometric quality of such performance ratings would introduce greater consistency and equity into a very vital and judgmental evaluation. Recognition of such improved consistency would, in turn, introduce enhanced comparability of performance ratings, and thus, potentially enhance the employee’s perspective and attitude about both the appraisal process itself and its ultimate use(s). Moreover, the employee is more likely to perceive that the performance rating process is employing the same set of benchmarks across employees and is therefore more equitable.

Given the current difficulty in conducting accurate, timely, consistent and objective evaluations of managerial performance, such an appraisal process is an area where expert system technology has considerable potential. Most modestly successful evaluation systems involve a rater who observes behaviors that are job relevant, evaluates those behaviors, and weights the evaluations to arrive at a composite performance rating. Ilgen and Favero (1985) describe this as an interaction process followed by a judgement process. Given the concern for employee development, we add a third process of communication and explanation back to the ratee. It is precisely these processes that are the relevant domain of expert systems applications, and even more so when completeness, accuracy, timeliness, consistency and unbiasedness are recognized as worthy attributes. Moreover, the expert system can more accurately and consistently identify performance differences across people than can subjective non-automated systems. Thorough and fair appraisals require considerable time and effort. Yet managerial time is typically fragmented into numerous brief, and discrete units that make it difficult for the evaluator to accomplish the extended mental activity of synthesis, assimilation, and assessment of another’s performance with clarity of focus, objectivity and consistency across both time periods and individuals. Here again, an expert system can expedite the process by quickly integrating the many dimensions of the ratee’s performance data.

The expert system can blend the computational scoring of the traditional approach with the judgmental assessment of the qualitative approach. Use of such a system draws upon the predeveloped knowledge base which serves as the repository of “collected wisdom” of the domain experts. Granted, the knowledge acquisition by the system involves considerable subjectivity, human discretion, and judgement, as it should, but it is introduced quite apart from the evaluation of any given employee and thereby introduces an element of consistency rather than capriciousness. Distinct from the human evaluator, the expert system can be purged of contextual biases such as rating order and prior rating anchors.

Unlike conventional programming languages, a rule-based expert system depicts the problem domain as a set of rules rather than as a sequence of processing steps. Such rules, or heuristics, are sought out and executed by the system’s inference engine as they are needed to make a deduction, rather than in the sequential order in which they exist in the system’s repository. Once both the factual data and heuristics involving criteria and performance indicators are carefully assimilated in the system’s knowledge base, the system will ask the user for the inputs relevant to a given subject, thus enabling the inference engine to simulate the collective expertise and judgement of the contributing experts and effectively complete an appraisal, while systematically avoiding the pitfalls often encumbering the human evaluator’s process. Should the user/rater elect to challenge the relevance of any inquiry from the system, he may simply ask “why
is this query important?”; the expert system will then explain the intermediate level conclusion which it is attempting to draw and how the subject question relates to its ability to do so. Another form of self-explanation occurs when the expert system has concluded its evaluation. The user may seek clarification of just “how did it arrive at this conclusion?” The expert system will possess the ability to explain the foundations for its final assessment so as to better identify areas of strength or weakness to the ratee. Such a reconstruction of the logic steps is not necessarily possible for all human evaluators, regardless of their level of expertise.

One of the most useful features of the expert system is its ability to make a single change in any of the input parameters and re-evaluate the case. Such sensitivity analysis permits both the rater and the ratee to explore “what if” situations concerning the impact of a selected issue on the ultimate assessment.

The “Model”

In order to investigate the appropriateness of expert system technologies to the evaluation of managerial performance, a rapid prototype ES was developed on an IBM compatible microcomputer using the commercial shell, EXSYS1. The intent was to model the evaluation process of only the communication skills component of the performance of the non-teaching administrative and professional staff at a state supported university. The model’s knowledge base was developed from numerous interactions with middle and senior level administrators serving as the domain experts. In the early stages of the knowledge acquisition process, a frame of reference was provided to the knowledge engineer through a standard performance appraisal document used by the institution. The knowledge acquisition process sought to describe factual data and rules employed in appraising the interpersonal skills of the first line administrators in such a manner that an ultimate rating of “Excellent”, “Good”, “Improvement Needed”, and “Unsatisfactory” could be assigned, each with an accompanying confidence factor.

The contributions of the domain experts in this application are many. Initially, they must identify characteristics that are indicative of good communication skills, given the job descriptions of the managers whom they must evaluate. They must profile conditions that adequately describe performance options and allow the non-expert user of the system to accurately discern among such conditions. The most critical role of the domain expert in this model is in the sensitive assignment of confidence factors for each ultimate rating, given its link to the various behavioral conditions. Both the absolute values of such factors and their relative values between various qualifier conditions enable the expert to fine tune the system. For example, areas deemed more critical to proper performance, such as customer service, can be assigned a wider range of possible confidence levels from top to bottom than those deemed of lesser importance, e.g., distracting mannerisms.

In an attempt to give a sense of structured development to a fluid and evolutionary process, a Dependency Diagram is offered as Figure 1. The Dependency Diagram continually evolves during the development process of the early prototype. Its value, as a tree-like form of visual documentation, lies in depicting the knowledge segments, their inter-relationships, and their contributing roles to the final decision/recommendation. Specifically, the Dependency Diagram depicts the case specific information that is requested of the user during a consultation (acknowledged via “?”), intermediate level milestones (depicted as ovals), and the actual decision points in the process where heuristic reasoning occurs (acknowledged as triangles). Moreover, the Dependency Diagram focuses on examining the decision process so as to lay out the “reason-
Figure 1: Dependency Diagram
ing paths’ that enable a decision.

In this prototype, a composite profile of “communication skills” was initially established reflecting the various divergent perspectives of the domain experts. The most immediate components of an individual’s overall communication were established to be: (1) dealing with external entities (“customers”/vendors); (2) internal, spoken, one-on-one interactions; (3) internal, spoken, group communications; and (4) written communications. The relative role of each of these would be individually tailored to the subject manager’s job responsibilities, and would be established in advance of the performance evaluation period. Performance in each of these areas was influenced by several “layers” of intermediary factors, which, in turn, were determined by numerous behavioral proficiencies of the subject employee. Moreover, the “communication”, of any given employee, was ultimately decomposed into an applicable subset of the following issues:

1) the employee’s response to “customer” requests,
2) the employee’s resolution of “customer” complaints,
3) the employee’s response to vendor inquiries,
4) the employee’s resolution of problems with vendors,
5) the employee’s ability to influence the thinking of others,
6) the employee’s ability to elicit action by others,
7) the employee’s ability to listen to others and incorporate that input into his own communication output,
8) the employee’s dialogue style and/or mannerisms,
9) the frequency of the employee’s appeals for intervention by, or advice, or information from superiors,
10) the frequency of the appeals from subordinates for clarification for additional information,
11) the employee’s ability to address groups, and participate effectively in group meetings, and
12) the employee’s ability to prepare effective reports, letters, directives and memoranda.

Each of these primary determinants was assessed through a series of up to 29 interrogatories; the specific subset of questions relevant to any particular subject was determined during run time by the system based upon responses to early questions. The user’s responses then feed a collection of decision rules developed in the conditional “if-then-else” format. An initial screening by the expert system distinguished those managers without routine contact with “customers” and/or vendors from those with such regular contact. All first line managers and above were expected to be evaluated on, at least, the remaining issues. In this early prototype system, a model consisting of 168 if-then-else rules was built around these issues. As each was critiqued, and given influence within its relevant major component, a confidence factor was assigned for each of the ultimate communication performance ratings possible. The confidence levels for the final composite assessment were established as weighted averages of the confidence factors contributed by the four major dimensions acknowledged above. These weights should reflect the subject’s specific job description and responsibilities, and consequently should be an integral part of the annual goal setting process performed jointly by the first line manager and his superior.

Deployment of the Model

As discussed earlier, instead of pursuing the broader objective of developing an expert system to aid in the evaluation of total managerial performance, this model has been scaled back to address merely one dimension of such overall performance, namely communication skills. This sub-system approach is consistent with the incremental design strategy of prototyping. This
development strategy assumes that the requirements of the system, and/or the formal articulation of behavioral rules can be known only partially at the beginning of the system development. It permits the clarification of users’ needs and the domain experts’ knowledge by actively involving them in a rapidly iterating development process. A small-scale system, that does not purport to be comprehensive, is quickly built in order to “get something in the hands of users as quickly as possible”. The developer anticipates errors and shortcomings in the system, but attempts to learn as much as possible from such errors, as the users have something to react to. Such user feedback will hopefully result in a system that is refined iteratively, with rapid turnaround, over a long trial period.

The mini-system discussed here has gone through alpha testing, or verification, which seeks to detect logic errors by executing the system in a simulated environment. Test cases were fabricated with the express intent of determining whether the system could process them correctly. These tests resulted in considerable revision of selected confidence assignments to better balance their relative values between various qualifiers.

Beta testing, or validation, is the use of the system in a live environment, with the awareness that the system can fail. A pseudo beta test for this system was created using a hold-out sample of real cases for end of year evaluation. During the earlier knowledge acquisition phase, the thoughts of the domain experts were extracted by observing their behaviors as they performed the evaluation of junior level managers’ communication skills. A random sample of such assessments was recorded but not permitted to enter into the knowledge engineer’s development of the knowledge base. Such recording of hold-out cases took place under ideal circumstances when the appraiser was aware of no compromising influences on his objectivity; thus, these seven cases would serve as benchmarks which the expert system would seek to replicate. Upon completion of the current iteration of the system, a comparison of the performance of the expert system to that of the pre-recorded evaluation revealed that, in each hold-out case, the profile of the assessment distribution generated by the system tightly paralleled that of the human evaluator. Whenever the human was confident of a “most appropriate” classification for a particular subject, the system rendered the same assessment; whenever the human evaluator was “torn” between two classifications, the system established similar confidence levels for those same assessment categories. Representative comparisons of the assessment profiles of a human evaluator and the expert system are provided in Table 1 for two cases.

A “suggestive” observation from these results is that in five of the seven hold-out cases, including the two above, the expert system was more “willing” to go to the polar categories than were the human evaluators. Should human performance appraisers have a propensity to hesitate to make assessments at the extreme classifications, this subtle skewness of the expert system evaluation may not be degrading. Next a modified Turing test was performed in which a case situation was analyzed independently by the expert system and by a human personnel evaluator. Other managers were given the conclusions and rationale of both appraisals, without knowing which is which, and charged with comparing them. With the output media form and format neutralized, no substantive differences were detected. Granted that these preliminary validation tests, with extremely small samples, are not suf-

### Table 1: ES/Domain Expert Appraisal Comparisons for Two Hold-Out Cases

<table>
<thead>
<tr>
<th>CASE 1</th>
<th>CASE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BY</strong></td>
<td><strong>BY</strong></td>
</tr>
<tr>
<td><strong>RATING</strong></td>
<td><strong>HUMAN</strong></td>
</tr>
<tr>
<td>Excellent</td>
<td>5%</td>
</tr>
<tr>
<td>Good</td>
<td>35%</td>
</tr>
<tr>
<td>Needs Improvmt</td>
<td>50%</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>10%</td>
</tr>
</tbody>
</table>
icient to make the communication module operational, they do confirm the appropriateness of the expert system application and encourage further refinements and the extension to other modules.

Conclusions

The intent of the expert system approach is to make available to the non-expert user the expertise of the scarce human expert while enhancing the dimensions of objectivity, comprehensiveness, and consistency. Many of the attributes of expert system technology address the historical shortcomings of traditional managerial performance appraisal systems. The model developed here allows for a consistent, and objective assessment of an employee’s performance in the area of communication skills and is offered as evidence that an expert system approach to administering managerial performance appraisals can prove effective. Further iterations of this prototype, coupled with similar assessments of the other areas of managerial performance, can be used to develop an overall performance evaluation aid that would mitigate anxiety for rater and ratee alike. Performance of human resources should not merely be appraised, it should also be managed. By assisting with the difficulties of the assessment process, the expert system lays the foundation for improved management of administrative performance and human resource development.

References


---

*James S. Moore has been on the Indiana University faculty since 1980. His teaching and research interests are in microcomputer-based decision support systems, management science models, expert systems, and their intersection.*
Related Content

Effectiveness of Systems Analysis and Design Education: An Exploratory Study
www.irma-international.org/article/effectiveness-systems-analysis-design-education/3754/

A Theoretical Investigation and Extension of a Model of Information Technology Architecture Maturity
Randy V. Bradley and Terry A. Byrd (2011). Organizational and End-User Interactions: New Explorations (pp. 352-375).
www.irma-international.org/chapter/theoretical-investigation-extension-model-information/53099/

The Design Implementation Framework: Iterative Design From the Lab to the Classroom
www.irma-international.org/chapter/the-design-implementation-framework/183013/

End User Computing Ergonomics: Facts or Fads?
www.irma-international.org/article/end-user-computing-ergonomics/3815/

Designing End-User Geographic Information Systems
www.irma-international.org/article/designing-end-user-geographic-information/3725/