responsibility for promoting our field in the 21st century. She stressed the
criticality of being responsive to the needs of industry.

At the end of the day, all the speakers took part in a panel discussion
addressing professional development issues. Students asked questions on a
variety of topics from landing a job in a competitive market to integrating
work and family. Panelists candidly shared their views and experiences in
dealing with the issues. We’d like to express our appreciation to all the
panelists who made the consortium a success!

Doctoral students from over 20 graduate programs in industrial/organiza-
tional psychology and organizational behavior took part in the consortium.
Approximately 69% of the participants were from I/O psychology programs,
27% were from management departments, and the remaining 4% were from
other disciplines. Most participants (45%), indicated an interest in pursuing an
academic career path. Consulting careers were sought by 29%, applied
positions by 19%, and research positions by 7%.

According to our survey of participants, most students felt that the
consortium was an excellent career development opportunity. Comments
reflected an appreciation of the diversity of panelists’ backgrounds,
experiences, and interests. Student feedback also indicated that keeping the
consortium small was desirable in order to provide a better chance for
interaction among presenters and participants.

**Consortium 1996:** Plans for the 1996 I/O Psychology Doctoral
Consortium are already in the works. Demand for the consortium is typically
very high, thus each graduate program is only allowed to nominate one
participant. Space is filled on a first come first served basis, so send in your
registration materials early. Look for details in this issue of **TIP.**

---

**From Both Sides Now**

**Performance Appraisals:**

*Political Tools or Effective Measures?*

Allan H. Church

Nothing in organizational life these days is guaranteed, except, perhaps, a
performance review. Whether one is at the very bottom of the organization or
at the senior most level, someone somewhere is evaluating your performance.
Moreover, everyone seems to get in on the act—managers, stockholders,
financial advisors, the press, consultants, friends, and family. Of course, most
employees experience this process as the standard dyadic exchange between
supervisor and subordinate. In this time of massive staff reductions,
reengineering, and continuous improvement efforts, however, the evaluation
of employee work behavior has become an even more pressing and important
function in organizations (as if we haven’t spent the last 30 years of I/O
psychology studying these issues). A quick and dirty search on the
PSYCHINFO database for the term “performance appraisal” for example
yielded 403 article listings between 1985 and 1994 alone. Clearly the notion of
evaluating people’s performance is still a hot topic.

Unfortunately, it seems as though the practice of contemporary
performance reviews is akin to Forrest Gump’s box of chocolates—you never
know what you’re gonna get. I am always coming across practitioner pieces in
popular trade journals about taking the guess work out of the process,
increasing its validity and utility, or making it more effective. Practitioners are
not only concerned with improving existing performance measurement
systems, however. Some are calling for radical change. In a recent issue of
**Quality Progress** (1995), for example, George Eckes noted the following
conflicts between TQM and more traditional performance appraisals based on
Deming’s original criticisms:

- Appraisals nourish short-term performance (e.g., quotas, sales figures,
  forms processed, products shipped) and destroy long-term planning;
- Appraisals promote individualism and destroy teamwork;
- Appraisals assume that people are responsible for all results;
- Appraisals contain subjectivity and immeasurables.

He recommends instead either a customer-supplier appraisal or a process
(rather than output) oriented system that doesn’t pinpoint individuals but
rather suboptimal work processes. While the first option is similar to the 360°
notion of expanding perspectives to include other types of raters—e.g.,
customers, direct reports, peers, team members—the second is something very
different and probably difficult to implement. Other authors have expressed
concerns regarding the highly political nature of performance appraisals, particularly in executive populations. Based on their research with executives from a number of different companies, for example, Gioia and Longenecker (1994) assert that any attempt to increase the validity of a rating system:

"...is a future pursuit, mainly because the managers who use these instruments often actively pursue agendas that are incompatible with highly accurate ratings. Executives admit that, in appraising others, they often intentionally avoid meeting the goal of accuracy in favor of achieving goals that have more to do with exercising discretion and maintaining departmental effectiveness" (pp. 48-49).

Ironically, the researchers note, these same executives complain that the ratings they receive are often not accurate reflections of their ability and performance as well.

Despite such problems and concerns, given (a) the inevitability of some form of performance appraisal in contemporary organizations and (b) the important function that this type of measurement serves for performance and personnel management, I decided to pose the following questions:

1. Are contemporary performance appraisal systems doing what they were designed to do—i.e., measure/quantify performance—or are they shaped more by political, implementation, and measurement idiosyncrasies inherent in their use?

2. What can be done to improve their utility and effectiveness?

Jai Ghorpade, Professor of Management at San Diego State University and an active consultant on the side, provided the following comments on the most popular method of performance ratings used in the United States.

This is a highly relevant topic and one that needs to be confronted. Unfortunately, there currently do not exist any empirical studies that can provide definitive answers to all parts of the question. However, the literature does provide a guide to the origins and dynamics surrounding the Graphic Rating Scale (GRS), the tool most commonly used in performance appraisal in the United States. The GRS has two key elements: (1) focus on traits (e.g., dependable, cooperative, knowledgeable) in measuring performance, and (2) use of rating scales with subjective anchors (e.g., very low/very high; good/bad) for evaluating individual performance on each of the traits on the instrument.

The GRS was formally and ceremoniously introduced into American industry in the 1920s by prominent industrial-organizational psychologists of that time. This "new" method was designed to remedy problems with the established methods of the time—seniority, Frederick Taylor's appraisal through performance standards set by time and motion analysis, and man-to-man rating scales that required the preparation of elaborate master scales against which individual performance could be rated (Paterson, 1922-1923).

This new method was judged to be superior to the old methods for several reasons. Its major attraction, of course, was practicality—it was brief and could be completed within a short amount of time. The GRS was also claimed by its developers to have a couple of technical qualities that made it ideal as an appraisal tool. The first of these was that it freed the rater from having to link ratings to any actual, objective counts of performance (a pivotal feature of Taylor's Scientific Management). Second, the subjective scales enabled the rater to make "as fine a discrimination of merit as he chooses" (Paterson, p. 147). The original GRS had scales without numbers. The anchors were placed beneath blank lines, and stencils were provided for converting the check marks into numbers.

The introduction of the GRS paralleled the rejection of several of Taylor's ideas. It spread throughout industry like wildfire and remains the most common tool used today. It has great intuitive appeal and "makes sense" to managers, deans, and persons in positions of authority. "Improvements" over the original typically take the form of elaboration of anchors.

Any discussion of change thus needs to confront the preeminent position that the GRS has acquired in industry. Many "radical" and intelligent ideas for change have been offered by academics (Murphy & Cleveland, 1995). However, industry resists approaches that depart fundamentally from the GRS format. I am convinced, however, that it is possible to build on this format and to assemble systems that lessen the bias, inaccuracies, and psychological pain that accompany the undisciplined use of the GRS. In another publication, this issue was covered in a more leisurely way (Ghorpade & Chen, 1995).

However, I would like to highlight briefly the safeguards on the use of the GRS advocated by Donald Paterson himself, the original developer and promoter of the GRS. Paterson offered ten safeguards. In presenting this, I use his terminology but point out (in parentheses) links with current labels and trends wherever appropriate: (1) Instruct supervisors to follow directions (develop clear rules for administering appraisal systems); (2) ratings should be based on work that employees are actually doing and not on guesswork (assure job relatedness); (3) employees should only be compared with other employees doing the same work; (4) insist that ratings be based on defined characteristics (measure reliability through consistency of terminology); (5) spread the ratings over three-month periods; (6) secure independent ratings from two or more supervisors (include peers and co-workers); (7) use the independent ratings to isolate...
radical disagreements and hold conferences to detect and correct possible bias against the employee; (8) identify raters who rate all employees the same, confront them with distributions of ratings given by others, and encourage them to recognize real differences among employees; (9) study continuously the ratings statistically to detect and correct persistent leniency and strictness; and (10) hold conferences with raters to discuss the results of statistical analysis and to correct problems (Paterson, pp. 158-160).

Paterson’s safeguards anticipate the typical problems generated by the GRS format and provide a guide for corrective action. In particular, his insistence on continuous statistical analysis of appraisal ratings, and conferences to bring about corrective changes based on the analysis, would promote a healthy understanding of the consequences of established practices and inspire participants to develop methods that enhance utility and effectiveness of the appraisal process. Interestingly, this idea of continuous self-study is at the heart of the TQM movement.

Thus, for Jai, one of our biggest challenges to improving performance appraisal measures is to ensure a sound measurement system and more open and honest communication about the process and the actions that should result. I found it interesting that embedded in Paterson’s original recommendations, made in the early 1920s, mind you, was a call for multiple raters from different perspectives, behaviorally grounded items based on a job analysis, comparability in terms of equal standards of evaluation, and recognition of individual biases and ratings artifacts. Is it me or are these some of the same issues and recommendations still being made today?

Bob Cardy, a Professor of Management at Arizona State University, took a more direct approach in response to the question citing information processing models and trust as key components to the rating process.

Yes, I believe that performance ratings, for the most part, do a reasonable job of measuring/quantifying performance. The dominant appraisal systems focus on concrete specifics and emphasize documentation. Appraisal systems that force raters to make distinctions among ratee performance levels are commonly used. In short, appraisal systems have largely been designed to be measurement tools and seem to have been based on the idea that the more specific the better, and that raters may not be trustworthy.

Despite these characteristics, it seems that the prevalent opinion in organizations is that appraisals, due to their subject nature, are errorful and biased. Belief that biases occur can be seen in countless grievances and lawsuits. It appears that the perception is that performance ratings don’t do a good job of measuring performance and that they are guided by political, personal, and other factors. On net, however, I believe that performance ratings do a reasonable job of capturing performance.

While there is mistrust of appraisal, I still believe that performance ratings largely reflect performance. Numerous laboratory and field studies have found performance ratings to be valid and substantially accurate. However, because of the dominance of the measurement concern in the design of appraisal systems, raters are routinely asked to focus on specific and particular outcomes and to finely distinguish among performance levels. Perhaps we have developed appraisal systems too much from the perspective of documentation and lack of trust in raters. The more concrete and specific the appraisal format, and the less flexibility allowed raters, the better the system seems assumed to be. The emphasis on the measurements of specifics can backfire, since fine distinctions may be difficult and not needed. Specifics may not be how we process information. There are studies that indicate that we tend to process information in a global fashion and that global judgments can be pretty accurate. Raters may also be fully aware that these specifics give a woefully deficient picture of the performance level of the workers. Rater prototypes of raters may be quite accurate and not as deficient at capturing the overall performance construct as some of the specific measures can be. Besides, while we may be reluctant to admit it, a lot of the specific judgments are probably based on prototypes anyway. Unfortunately, to the extent that specifics are deficient, raters are forced to go beyond this simple detailed level if they are to capture performance of the ratees adequately. This can then be interpreted as bias and evidence of politics or the role of personal feeling.

Rather than making assumptions about what appraisal system characteristics will lead to good measurement, we have an opportunity to take what we know about modern organizational life and what we've learned about information processing and use it to construct an improved rating system. While perhaps contrary to the measurement approach, we need to allow flexibility in the system. An appraisal system that is highly specific and concrete may be legally defensive, but will be outmoded and useless in no time in today’s dynamic organizational environments. There is a need to broaden and make the system more flexible. At the same time, this will open up the system to more political influence. However, I believe that you have to trust your raters or get rid of appraisal altogether. It may be a common paradox that the less the system is controlled and contrived in an effort to get better ratings, the better the ratings! What's critical isn't the rating format, it's the users—both raters and ratees.
The rating system needs to be responsive to the needs of the rates. What kind of information do they need to do a better job? That is, what should be the focus of the rating system? Raters are capable of underestimating the most sophisticated appraisal system or of providing effective performance management with the crudest of appraisal systems. Raters need to be trained in making good performance judgments and in coaching and feedback skills. These efforts should reduce the perception of bias in appraisal. However, it needs to be recognized that, for the most part, the influence of raters' feelings, such as liking, probably are a function of ratee performance characteristics. Besides, being likable and engaging in impression management tactics can help to assure that the job gets done efficiently and effectively. A broadened appraisal system may allow for the inclusion of these factors to the extent that they contribute to the system. Separating out performance caused by person from that caused by system is another critical factor for which training should be designed.

The utility and effectiveness of appraisal systems would improve if the emphasis would be on performance improvement and if the system would be future oriented and flexible enough to recognize the variety of ways people may contribute to the mission of the organization. Certainly, measurement of performance would be important. However, the focus would be on what’s needed to provide adequate feedback and improve rather than making fine-grade distinctions among performance levels.

From Bob’s perspective, performance appraisals are effective most of the time, despite what people think. Although he makes a number of interesting points, I was intrigued by the idea of insuring the flexibility of the appraisal system to adapt to future needs. There is a key distinction between redesigning a performance rating system every time it’s needed and designing a single system from the start that allows for these types of changes on an on-going basis. Speaking as one with experience designing automated feedback systems for leadership and management development purposes, building a truly flexible reporting process is not an easy task, particularly when data from multiple perspectives and different versions of existing instruments are involved. Of course, as Bob notes, without the ability to adapt performance systems over time they will, in all certainty, become obsolete.

The third contributor for this topic is David Bracken, Director of Organization Assessment and Research at Personnel Decisions Inc. David’s comments focus on the concept of performance management (versus appraisal) and, once again, incorporating 360° feedback into the process.

Stating that the objective of contemporary performance appraisal systems is to “measure/quantify performance” vastly understates the purpose of an ideal performance appraisal. While valid, reliable measurement is certainly the keystone, the system must also set the course for action (e.g., alignment), establish an incentive for performance (e.g., motivation) and set a course for development. Our clients view appraisal more in the context of a system, usually under the banner of performance management.

With those added purposes in mind, contemporary performance management systems seem to be making strides toward creating alignment, specifically tying together organization, group, and individual goals. The “state of the art” performance management system will be designed with a downward communication of the strategic vision, integration into goals and objectives at each level, and a closing of the loop with an indication of the extent to which the goals are being achieved.

The advent of the 360° feedback systems has also served to improve performance appraisals and performance management systems. For example, a well-designed 360° system will
- Clearly communicate the purpose, method and desired outcomes of the evaluation process
- Create alignment throughout the organization around competencies and/or values
- Standardize the evaluation process, creating acceptance and perceived fairness;
- Increase participation throughout the organization for helping develop individuals
- Train participants and users of the feedback to increase the accuracy of the ratings and set expectations
- Establish accountability for change with the ratee, acknowledging the role of the organization in providing developmental resources.

There is no question that our well-designed systems sometimes fail at the point of delivery. Our role as I/O psychologists is to increase the probability that the tools will be used effectively and to prevent and/or address abuses of the system. Despite all the critical incidents and horror stories, we have reason to believe that some of the barriers to effective performance appraisals are being overcome.

David’s points are well taken regarding the larger purpose of performance measurement systems and their contribution to shaping future behaviors, as well as assessing them in the present. I was also intrigued by his comments regarding the role of I/O psychologists in the process. More often than not, we are so focused on the quality of the rating system itself—making it more effective, valid, efficient, and meaningful—that the implementation and issues of practice, particularly regarding ethical uses of this type of data, are left to chance. Having seen firsthand the impact that individual feedback can have
on managers’ feelings and cognitions, it is paramount that we, as practitioners, make a concerted effort to manage the integrity of such systems with our clients and in our own organizations.

I would like to thank Jai, Bob and David for their thoughtful contributions to the performance appraisal discussion above. If you have a topic you would like to see discussed in an upcoming issue or you would like to write about one yourself, please give me a call or drop me a note at W. Warner Burke Associates Inc., 201 Wolfs Lane, Pelham, NY 10803, (914)738-0080 tel. / (914)738-1059 fax.

References

Biographies
Jai Ghepardo is a Professor of Management in the College of Business Administration at San Diego State University. He has published articles in such journals as the Academy of Management Journal, Multivariate Behavioral Research, and Personnel Psychology. His teaching, research and consulting interests include general management, human resource management, organization development and comparative industrial relations. He received his Ph.D. from the Graduate School of Management at UCLA in 1968.

Robert L. Cardy is a Professor of Management in the College of Business at Arizona State University. He has also served on the faculty of the State University of New York at Buffalo. Dr. Cardy has authored a number of articles in such publications as Basic and Applied Social Psychology, the Journal of Applied Psychology, and Organizational Behavior and Human Decision Processes. His interests include personnel appraisal and training, selection, information processing, TQM, and performance management. Dr. Cardy received his B.S. and M.A. from Central Michigan University and his Ph.D. from Virginia Polytechnic Institute.

David W. Bracken is Director of Organization Assessment and Research at Personnel Decisions Inc. Previously he was Director, Consulting Services for National Computer Systems. Dr. Bracken has over 15 years of experience in organization assessment as a researcher, practitioner and consultant. More recently he has focused on the development and implementation of 360° rating systems and is a co-founder of the Upward Feedback Forum. Dr. Bracken received his B.A. from Dartmouth College, and his M.S. and Ph.D. degrees in I/O Psychology from the Georgia Institute of Technology. He is a licensed psychologist in the State of Georgia.

Complete the Circle
Round out your 360° feedback or upward appraisal assessments with the
Management Development Report
From Hogan Personality Inventory (HPI)

The HPI Management Development Report is the only available management assessment procedure with these features:
• Keyed into the success factors identified by major management consulting firms
• Provides specific behavior tips for development
• Normed by managers
• Empirically derived feedback from upward appraisals

The HPI-based Management Development Report is available through:
Hogan Assessment Systems
P.O. Box 521176
Tulsa, OK 74152
Phone: 1-800-756-0632
Fax: 918-749-0635