# Toward narrowing the research-practice gap in performance appraisal

2 authors:

Cristina G. Banks
University of California, Berkeley

13 PUBLICATIONS   182 CITATIONS

SEE PROFILE

Kevin Murphy
University of Limerick

153 PUBLICATIONS   10,410 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project    LLLightinEurope.com View project

# TOWARD NARROWING THE RESEARCH-PRACTICE GAP IN PERFORMANCE APPRAISAL

CRISTINA G. BANKS
Department of Management
The University of Texas at Austin

KEVIN R. MURPHY
Department of Psychology
Colorado State University

Recently, performance appraisal researchers have adopted a cognitive approach to analyzing judgment processes in performance appraisal. While this approach allows researchers to tap a wealth of knowledge applicable to the appraisal context, this line of research is likely to widen the already existing gap between research and practice. We argue that coordination of the talents of researchers and practitioners is essential for narrowing the gap. Specifically, we suggest researchers focus their research on the best *methods* of ensuring use of relevant and valid data in appraisal, given organizational constraints. We also suggest practitioners focus on determining observable and measurable aspects of performance, and thus, specify appropriate appraisal *content*. We also note that cognitive process research has promise for increasing raters' *ability* to judge accurately, but that this approach does not necessarily address the rater's *willingness* to provide accurate ratings.

Organizations continue to express disappointment in performance appraisal systems despite advances in appraisal technology. Appraisal reliability and validity still remain major problems in most appraisal systems, and new (and presumably improved) appraisal systems are often met with substantial resistance. In essence, effective performance appraisal in organizations continues to be a compelling but unrealized goal.

Problems associated with performance appraisal are documented in volumes of articles in scientific and trade journals. Both researchers

---

and practitioners have analyzed performance appraisal problems and
have suggested several and generally different remedies.[1] For exam-
ple, practitioners suggest that organizations need to increase manage-
ment's commitment to the appraisal system, increase communication
between supervisors and subordinates regarding tasks to be per-
formed, improve appraiser feedback skills, and clarify performance
objectives and criteria. Researchers, on the other hand, stress the need
to sharpen observational skills, reduce rating errors, use better for-
mats, and utilize information more effectively. The fact that research-
ers and practitioners focus on a different set of appraisal problems and
hence, advocate different solutions, suggests a lack of coordination in
solving appraisal problems. More important, this divergence in focus
indicates that researchers' solutions may not speak to practitioners'
problems.

In recent years, researchers have suggested analyzing the cognitive
processes underlying performance appraisal as a method of improving
appraisal judgments. This approach, which has taken performance ap-
praisal researchers by storm since Feldman's (1981) seminal article,
allows researchers to tap a wealth of research in social and cognitive
psychology that is applicable to performance appraisal. In particular,
this approach promises to generate a better understanding of *how* per-
formance judgments are formed and retained for use in appraisals.
Several studies using the cognitive processing approach have appeared
recently (Balzer, 1983; Banks, 1979, 1982; Barnes-Farrell & Couture,
1983; Cooper, 1981a; DeNisi, Meglino & Cafferty, 1984; Lord, 1985,
in press; Murphy & Balzer, 1983; Murphy, Balzer, Kellam, & Arm-
strong, 1984; Murphy, Balzer, Lockhart, & Eisenman, 1985; Murphy,
Garcia, Kerkar, Martin, & Balzer, 1982; Murphy, Martin, & Garcia,
1982). These studies have examined a variety of issues, ranging from
ways in which raters decide what to observe to distortions in memory
for behavior.

Despite the potential promise of cognitive research in appraisal, this
shift in the focus of performance appraisal research may have omi-
nous implications for performance appraisal practice. By investigating
appraisers' cognitive processes, researchers appear to be drifting even
farther from concerns which are most clearly voiced by practitioners.
Thus, although research on cognitive processes in performance ap-

---

[1]The terms "researcher" and "practitioner" describe roles rather than individuals,
since many psychologists working in the area of performance appraisal are involved in
both research and application. Thus, the same individual may employ different strate-
gies, or may be concerned with different issues in performance appraisal research and
practice.

praisal may advance our understanding of human judgment, it has not yet led to significant advances in the practice of performance appraisal. Indeed, very few applications of this approach have been suggested or applied: the cognitive processing approach will improve the practice of performance appraisal only to the extent it is applied in the field. A clear divergence between the concerns of performance appraisal researchers and the concerns of practitioners suggests that advances in cognitive research may have little impact on practice.

The purpose of this paper is to describe why the existing gap between research and practice is likely to become *wider* as a result of the widespread use of the cognitive processing approach, and to suggest what can be done to narrow the gap. In this paper, we address conceptual and methodological problems associated with cognitive processing research which make applications of this research difficult. To increase the usefulness of this research, we suggest how one aspect of the appraisal process, observation of reliable and valid ratee information, may lead to improved performance appraisals by designing studies that capitalize on the technical strengths of both researchers and practitioners. We believe that only through the joint effort of researchers and practitioners can useful products be generated and adopted in organizations.

## Products Generated From Performance Appraisal Research

Over the past 35 years, researchers have developed several products to assist performance appraisal in organizations. Contributions fall within three general categories: appraisal formats, rater training programs and strategies, and appraisal processes. Researchers developed numerous *formats* such as checklists, rating scales, narratives, and work samples that help to structure the appraisal (see Bernardin & Beatty, 1983, or Carroll & Schneier, 1982, for detailed descriptions of formats). Formats aid actual appraisals by determining the type and number of dimensions assessed, the types of judgments made (frequency of behaviors vs. evaluation of behaviors), appraisal length, and comprehensiveness. Some researchers argue that particular formats also guide appraisal judgments (e.g., Bernardin & Smith, 1981). *Rater training programs* were designed to promote proper utilization of appraisal systems and to improve rating skills. Some of these training programs incorporate learning principles such as practice, feedback, and active participation (Spool, 1978) and emphasize behavioral observation (Boice, 1983; Thornton & Zorich, 1980). Various *processes* were developed to assist the appraisal process. Examples of these processes are the critical incident method (Flanagan, 1954), di-

ary-keeping (Bernardin & Walter, 1977), participation in format development (Friedman & Cornelius, 1976), and goal setting (Latham & Locke, 1979). These processes, as well as others, consist of a set of techniques appraisers can use to help them generate valid ratee data.

These products are useful in an ideal sense because they promote (but do not guarantee) systematic, job-related, and relatively error-free evaluation. However, these products have not been adopted widely (cf., DeVries, Shullman, Morrison, & Gerlach, 1981); for the most part, the appraisal systems actually used in organizations have failed to draw upon this body of research. For example, despite the advances made in training programs, few training programs in industry employ learning principles or emphasize observations; most involve proper use of appraisal forms (Feldman, 1983; in press). Thus, while the products developed by performance appraisal researchers help form accurate appraisals in theory, they are not perceived as useful (and therefore are not applied) in practice.

We believe that the products generated from current cognitive processing studies will be perceived as even *less* useful than those developed in the past. There are two reasons for this belief. The first reason is that cognitive processes captured in laboratory studies are likely to be substantially different from cognitive processes in actual appraisals. Several factors contribute to the potential lack of generalizability of laboratory research on appraisal. First, cognitive process research designs routinely eliminate or control for contextual factors such as competing tasks, time pressures, and delay between ratee behavior and appraisal; these factors are thought to have substantial effects on actual appraisals. Second, laboratory studies eliminate or control for interpersonal and effective processes, which also have considerable impact on actual appraisals (Dipboye, 1985). Third, laboratory studies focus on the judgment process (evaluation of ratee performance), whereas appraisals in organizations include both the judgment process and a rendering process (the marking of the appraisal form). Because laboratory studies eliminate or control for motivational factors such as personal and political agendas, financial need, and avoidance of conflict in laboratory studies, the ratings recorded by subjects in these experiments are likely to correspond with their evaluation of the (often videotaped) "ratee." In contrast, raters in organizations are often strongly motivated to record ratings which differ significantly from their evaluations (Mohrman & Lawler, 1983). In particular, raters are often motivated to avoid giving low ratings regardless of how poorly the ratees perform. Cognitive processing research may tell us about evaluation but will not necessa-

rily tell us much about rendering behavior when motivational factors are present.

A less obvious problem with cognitive processes captured in the laboratory concerns the type of information available to appraisers. Laboratory studies are often constructed so that appraisers are not forced to differentiate between relevant and irrelevant ratee information. Typically, most of the information presented is relevant and, as a result, the appraiser's task is to attend to as much information as possible. Quite the opposite is true in organizational settings: appraisers need to suppress information that is irrelevant or unreliable. Accurate appraisal in organizations is likely to depend heavily on one's ability to *suppress* irrelevant information as well as *select* relevant information; accurate appraisal in the laboratory may depend primarily on accuracy in observing (Murphy, Garcia et al., 1982). In sum, we believe the cognitive processes studied in the laboratory do not simulate well enough the circumstances of actual appraisals to improve significantly the quality of appraisals in organizational settings. Although cognitive research has generated a good deal of information about processes involved in evaluating others, this knowledge has not yet proved useful in practice.

The second reason for our lack of faith in the applicability of the cognitive processing approach is the difficulty of accessing cognitive processes. Since these processes are not observable, we must assume that elicitation techniques capture representative portions of relevant mental processes. However, how do we measure error in the data we elicit? Should we treat all verbal statements and other measurable outcomes as valid indicators of appraisers' cognitive processes? Unfortunately, the cognitive processing approach deals with variables that are simply difficult to measure and for which research methodologies are primitive. If the relevant cognitive processes *are* measurable, they may be exceedingly difficult to manipulate and change permanently. Life-long biases, decision strategies, prejudices, world views, experiences, and feelings are likely to be highly resistant to change. Even if we learned a great deal about appraisers' cognitive processes, we may not be able to improve them significantly.

### Narrowing the Gap

At this point, the reader may be wondering what, if anything, can we gain from the cognitive processing approach? We argue that *given a particular focus*, cognitive processing studies can yield highly useful information for both researchers and practitioners. The focus we prefer examines appraisers' cognitive processes in settings that simulate

the essential features of appraisal in organizations, using methods that take organizational constraints into account. This focus and underlying rationale is resented below.

The goal for both researchers and practitioners is to develop measures of job performance that are reliable, valid, job-related, standardized, and based on criteria known to both appraiser and ratee. A central aspect of achieving this goal is the identification and use of valid ratee information in that appraisal. Toward this end, researchers are likely to be concerned with what information appraisers attend to in evaluating ratee performance, while practitioners are likely to be concerned with basing evaluations on job-related information. Fortunately, practitioners' concern for greater use of valid ratee data is addressed by cognitive processing research. One of the central components of cognitive process models of performance appraisal (e.g., Cooper, 1981b; Feldman, 1981; Landy & Farr, 1980) is *behavior observation*. Cognitive researchers have examined raters' ability to observe and select valid and relevant behavioral information in conducting appraisals (Banks, 1982; Lord, in press; Murphy, Garcia et al., 1982; Murphy et al., 1984; Murphy et al., 1985). To the extent this research can identify circumstances in which raters accurately observe and make sensible and reliable decisions about what to observe, the practitioners' goal of providing a base of valid behavioral information for appraisals will be met.

One of the clearest manifestations of the present gap between appraisal research and appraisal in organizations is the role of observation in research versus practice. Researchers have believed for many years, probably since the development of the critical incident technique (Flanagan, 1954), that observation of job behavior is critical for appraisal reliability and validity. Observation of actual ratee behavior is assumed to be a determinant of appraisal accuracy, based on the assumption that such observations are objective, verifiable, and can be tied closely to job analysis information (Dunnette, 1966). Recently, preliminary links have been established between observation and appraisal accuracy (Bernardin, 1981; Bernardin & Walter, 1977; Murphy, Garcia et al., 1982; Thornton & Zorich, 1980). Despite the evidence supporting the importance of observation, relatively few organizations have incorporated observation-driven methods and techniques (e.g., behavior diaries, behavior observation scales) to improve the accuracy of their ratings. Why is there a gap when researchers have argued so strongly for its importance?

The gap becomes obvious when we consider in what *manner* observation will lead to greater accuracy. In order to improve appraisal

accuracy, observations must be frequent, unbiased, and representative. Researchers often design laboratory studies in which appraisers have access to all information, most of the information presented is relevant, and appraisers' time is devoted exclusively to the appraisal task. Under these circumstances, observation is almost assured to be frequent, and because ratee behavior is likely to be internally consistent, almost any sampling strategy will result in unbiased and representative sampling. In contrast, observation in organizations is fraught with distortions and inadequacies: observation is infrequent and noisy, and appraisers lack focus due to competing pressures, motivations, and demands. Perhaps it is not surprising, then, to find that adoption of techniques aimed at increasing observational accuracy is rare. Researchers' recommendations may work well under laboratory conditions but become impractical or inapplicable in practice because of inherent shortcomings in the quality of observations in organizational settings.

*Specific recommendations.* Instead of applying laboratory-generated and, to some extent, laboratory-bound methods, researchers need to generate methods of observation and cognitive aids that ensure valid input data (e.g., ratee job behaviors) *given organizational constraints.* This strategy suggests more attention needs to be focused on the problem of separating relevant from irrelevant information and on obtaining valid information when observation is infrequent. Specifically, appraisers need to learn to discriminate between relevant and irrelevant information and then select relevant information for use in appraisals. Recent empirical studies underscore the importance of separating the two (Banks, 1982; Murphy et al., 1985). In addition, appraisers need to learn how to employ work sampling to obtain valid data when observation is necessarily infrequent.

Given these needs, training programs should not train appraisers merely to observe; rather they should train them *how to decide what to observe* (Banks & Roberson, 1985). Researchers could contribute most by determining what are the best *methods* of encouraging appraisers to use relevant as opposed to irrelevant information in appraisals and to systematically collect valid ratee information through work sampling. For example, it would be useful to compare different memory (or "forgetting") aids to determine which ones allow appraisers to retain relevant information and suppress irrelevant information. Another example is a systematic analysis of work sampling approaches for jobs that vary in task complexity and concreteness.

If the role of the researcher now is to determine the best *method* of achieving utilization of relevant information, then the practitioner's role becomes one of specifying the *content* of appraisals. First, practi-

tioners could determine what aspects of performance are *measurable.* That is, they can identify which aspects of performance can be observed validly. In a way, we are suggesting an anthropological approach to specifying the content of appraisals. If managers are viewed as a culture, they may be studied in such a way to develop characterizations of effective and ineffective performance. Practitioners who are most familiar with job activities would be the best candidates for developing these characterizations.

In a related vein, a second contribution practitioners can make is development of effective prototypes of performance in organizations. We use the term "prototype" here to refer to a specific, coherent behavioral definition of satisfactory or superior performance. Thus, a prototype is a description of what the typical satisfactory or superior worker *does.* Observable and measurable aspects of performance gleaned from the work described above can be organized into prototypes of performance for specific jobs. These can be used subsequently to help appraisers develop stable and consistent performance schemas. Practitioners can help shape appraisers' prototypes by comparing individual schemas with a preferred prototype. In addition, practitioners could also identify instances in which stereotypes and biases are most intrusive or most strongly held. Revision of firmly entrenched stereotypes may be extremely difficult; practitioners may be able to identify which stereotypes can or cannot be changed.

A final contribution practitioners can make is separating the multiple functions of performance appraisal data. There are statistical as well as practical reasons for minimizing multiple use of appraisal data. When the same appraisal data is used for multiple personnel decisions (e.g., pay increases, training interventions, suitability for promotion), error associated with these judgments will be greater than the error expected if judgments were made on independent sets of data. This is because errors associated with each judgment combine rather than cancel each other out, thus dramatically underestimating the probability of error with each assessment. Basically, the error associated with the set of unrelated judgments taken together is greater than the error indicated by the alpha level (Hays, 1981). As a result, this strategy might present a highly unreliable picture of the ratee. Practitioners should heed these measurement guidelines whenever possible. To avoid this problem, practitioners could help by developing different performance prototypes for each purpose. For example, prototypes involved in assessing candidates for promotion might concentrate on behaviors which are required for more advanced jobs, whereas prototypes used in salary administration should focus on be-

havior required in the ratee's own job.

*Capability versus willingness.* Most performance appraisal inter-
ventions and technologies, including those described herein, are
designed to increase the rater's ability to accurately evaluate ratees.
However, as we have noted earlier, raters are rarely motivated to pro-
vide accurate appraisals and may, in some cases, be strongly moti-
vated to provide inaccurate appraisals (e.g., refusing to give low rat-
ings). Research on performance appraisal has consistently failed to
distinguish between evaluation and rating behavior. Cognitive process
research is clearly relevant to evaluation but may be only tangentially
relevant to rating behavior. The interventions we describe could in-
crease raters' ability to make accurate discriminations but will not
have a significant impact on rating behavior unless raters are willing,
as well as able, to rate accurately.

We believe that the problem of increasing the rater's willingness to
provide accurate ratings falls largely outside of the domain of cogni-
tive process research. The same, however, can be said of almost all of
the intervention and rating technologies developed to date. The dis-
tinction between capability and willingness is especially useful, since it
helps clarify what sort of problems can or cannot be addressed with
specific interventions, and also suggests appropriate criteria for evalu-
ating these interventions. We believe that the problem of making
someone a better rater and the problem of obtaining better ratings
from that person are conceptually and practically distinct. Although it
makes sense to attack the problem of capability first, organizations
must also consider factors which affect each rater's willingness to rec-
ord faithfully the judgments he or she has made.

## Conclusion

Performance appraisal researchers are now at the crossroads of
making significant strides in achieving a better understanding of the
appraisal process; in theory, this information should contribute to in-
creasing appraisal effectiveness in organizations. By adopting the cog-
nitive processing approach, researchers have the opportunity to tap a
wealth of information relevant to the appraisal context. Regretably,
this current line of research is likely to widen the already existing gap
between research and practice in performance appraisal unless re-
searchers and practitioners coordinate their efforts. We suggest that
one way to achieve coordination is for both to undertake the task of
determining how to increase appraisers' use of valid input data in ap-
praisal decisions. Researchers could identify methods and other aids
that will best facilitate appraisers' selection of valid ratee information.

Practitioners could determine the content of appraisals and specify how job behaviors could be assessed most accurately in organizational settings.

An analogy might illustrate how we envision coordination between researchers and practitioners. If the goal is to hit a nail on the head, then we need to find a hammer and locate the nail. Researchers can provide the hammer by determining what *methods* would be most effective for hitting the target. Practitioners can locate the nail by determining what *nails* can be observed best and which ones to hit. Appraisers simply hit the nail. We argue that collaboration between researchers and practitioners is essential to build a stronger and more appealing structure for performance appraisal.

## REFERENCES

Balzer WK. (1983). *Biased attention to and encoding of behaviors during performance appraisal: The effects of initial impression and centrality of the appraisal task.* Unpublished dissertation, Department of Psychology, New York University.

Banks CG. (1979). *A laboratory study of the decision-making processes underlying performance appraisal.* Unpublished dissertation, Department of Psychology, University of Minnesota.

Banks CG. (1982). *Cue selection and evaluation.* Unpublished manuscript.

Banks CG, Roberson L. (1985). Performance appraisers as test developers. *Academy of Management Review, 10,* 128-142.

Barnes-Farrell JL, Coutkure K. (1983, August). *Effects of appraisal salience on immediate and memory-based judgments.* Paper presented at the 91st Annual American Psychological Association Meeting, Anaheim, CA.

Bernardin HJ. (1981, August). *Rater training strategies: An integrative model.* Paper presented at the 89th Annual American Psychological Association Meeting, Los Angeles, CA.

Bernardin HJ, Beatty RW. (1983). *Performance appraisal: Assessing human behavior at work.* Boston: Kent.

Bernardin HJ, Smith PC. (1981). A clarification of some issues regarding the development and use of behaviorally-anchored rating scales (BARS). *Journal of Applied Psychology, 66,* 458-463.

Bernardin HJ, Walter CS. (1977). The effects of rater training and diary keeping on psychometric error in ratings. *Journal of Applied Psychology, 62,* 64-69.

Boice R. (1983). Observational skills. *Psychological Bulletin, 93,* 3-29.

Carroll SJ, Schneier CE. (1982). *Performance appraisal and review systems.* Glenview, IL: Scott Foresman.

Cooper W. (1981a). Ubiquitous halo. *Psychological Bulletin, 90,* 218-244.

Cooper W. (1981b). Conceptual similarity as a source of halo in job performance ratings. *Journal of Applied Psychology, 66,* 302-307.

DeNisi A, Meglino B, Cafferty TP. (1984). A cognitive view of the performance appraisal process: A model and research propositions. *Organizational Behavior and Human Performance, 33,* 360-396.

DeVries D, Morrison A, Shullman S, Gerlach M. (1981). *Performance appraisal on the line.* New York: Wiley-Interscience.

Dipboye RL. (1985). Some neglected variables in research on discrimination in apprais-

als. *Academy of Management Review, 10,* 116-127.

Dunnette MD. (1966). *Personnel selection and placement.* Belmont, CA: Wadsworth.

Feldman JM. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology, 66,* 127-148.

Feldman JM. (1983, August). *Practice, practicality, and prospects of training for performance appraisal.* Paper presented at the 91st Annual American Psychological Association Meeting. Anaheim, CA.

Feldman JM. (in press). Instrumentation and training implications for performance appraisal: A perceptual-cognitive viewpoint. IN Rowland KM, Ferris J (Eds.), *Research in personnel and human resources management, 1.* Greenwich, Connecticut: JAI Press.

Flanagan JC. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327-358.

Friedman BA, Cornelius ET III. (1976). Effect of rater participation in scale construction on the psychometric characteristics of two rating scale formats. *Journal of Applied Psychology, 61,* 210-216.

Hays WH. (1981). *Statistics.* New York: Holt, Rinehart and Winston.

Landy FJ, Farr J. (1980). Performance rating. *Psychological Bulletin, 87,* 172-107.

Latham GP, Locke EA. (1979). Goal setting: A motivational technique that works. *Organizational Dynamics, 8,* 68-80.

Lord RG. (1985). Accuracy in development measurement: An alternative definition based on raters' cognitive schema and signal detection theory. *Journal of Applied Psychology, 70,* 66-71.

Lord RG. (in press). An information processing approach to social perception, leadership perception, and behavioral measurement in organizational settings. In Staw B, Cummings LL (Eds.), *Research in Organizational Behavior.* Greenwich, CT: JAI Press.

Mohrman A, Lawler EE. (1983). Motivation and performance behavior. In Landy F, Zedeck S, Cleveland J. (Eds.). *Performance measurement and theory.* Hillside, NJ: Lawrence Earlbaum.

Murphy KR, Balzer W. (1983, August). *Systematic distortions in memory-based ratings: Consequences for rating accuracy.* Paper presented at the 91st Annual American Psychological Association Meeting, Anaheim, CA.

Murphy KR, Balzer W, Kellam K, Armstrong J. (1984). Effects of the purpose of rating on accuracy in observing teaching behavior and evaluating teaching performance. *Journal of Educational Psychology, 76,* 45-54.

Murphy KR, Balzer W, Lockhart M, Eisenman E. (1985). Effects of previous performance on evaluations of present performance. *Journal of Applied Psychology, 70,* 72-84.

Murphy KR, Garcia M, Kerkar S, Martin C, Balzer W. (1982). Relationship between observational accuracy and accuracy in evaluating performance. *Journal of Applied Psychology, 67,* 320-325.

Murphy KR, Martin C, Garcia M. (1982). Do behavioral observation scales measure observation? *Journal of Applied Psychology, 67,* 562-567.

Spool MD. (1978). Training programs for observers of behavior: A review. PERSONNEL PSYCHOLOGY, *31,* 853-888.

Thornton GC, Zorich S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology, 65,* 351-354.