5-2008

Characteristics of Faculty Evaluation Formats for Promotion, Tenure, and Annual Review

Angelette Gardner
Western Kentucky University, angelette.gardner@wku.edu

CHARACTERISTICS OF FACULTY EVALUATION FORMATS FOR PROMOTION, TENURE, AND ANNUAL REVIEW

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Arts

By
Angelette Gardner

May 2008

CHARACTERISTICS OF FACULTY EVALUATION FORMATS FOR PROMOTION, TENURE, AND ANNUAL REVIEW

Date Recommended _____

_____
Director of Thesis

_____

_____

_____
Dean, Graduate Studies and Research Date

Acknowledgements

First, I would like to give honor to God who has blessed me throughout this long journey. Second, I would like to thank my parents, Michelle and Ernest Gardner, for all the love and support that they have given me over the years. I would also like to thank Master, Aunt Kim, for her encouragement. Finally, I would like to thank my professors for supplying me with the tools necessary to complete this project, specifically, my thesis committee, Dr. Elizabeth Shoenfelt, Dr. Reagan Brown, and Dr. Steven Haggbloom, who guided me through the thesis process.

TABLE OF CONTENTS

CHARACTERISTICS OF FACULTY EVALUATION FORMATS FOR PROMOTION, TENURE, AND ANNUAL REVIEW

Angelette Gardner May 2008 38 pages

Directed by: Dr. Elizabeth L. Shoenfelt, Dr. Reagan Brown, and Dr. Steven Haggbloom

Department of Psychology Western Kentucky University

The present study attempted to identify common and unique characteristics of faculty performance appraisal formats and procedures by analyzing characteristics of formats and procedures from the psychology departments of 28 universities, including Kentucky Council on Post-secondary Education (CPE) benchmark schools for Western Kentucky University as well as schools that have Industrial/Organizational psychology master's degree programs. It was hypothesized that schools with Industrial/-Organizational Psychology programs would have better formats as defined by eight legal factors. However, this hypothesis was not supported. The hypothesis that graphic rating scales would be the most common method for collecting appraisal data was supported. It was determined that the performance appraisal system used at Western Kentucky University is very similar to systems used at the CPE benchmark schools that participated in this study.

Characteristics of Faculty Evaluation Formats for Promotion, Tenure, and Annual Review

Performance Appraisals

Performance appraisal is an evaluation of employee performance for the purpose of making organizational decisions (Rotchford, 2002). Researchers have defined performance appraisal to include the role of both the individual and the organization. According to The Joint Committee on Standards for Educational Evaluation, performance appraisal is "the systematic assessment of a person's performance and/or qualifications in relation to a professional role and some specified and defensible institutional purpose" (Stufflebeam, 1988, pp. 7-8). Performance appraisal can be useful to organizations. Performance appraisal can improve an organization's ability to make decisions regarding, for example, tenure and promotions (Murphy & Cleveland, 1995). One organizational objective of performance appraisal is to capitalize on the value an employee can add to the organization. Performance appraisal is a very effective way to accomplish this goal because it enables improved individual decisions and employee self-development. Based on appraisal information, individuals can more accurately target needed training and make better self-judgments. Performance appraisal can also serve to increase motivation, organizational commitment, and job satisfaction when employees accept the methods used to make organizational decisions. Finally, performance appraisal helps to justify legally personnel decisions that are made by the organization (Murphy & Cleveland). Thus, evaluations have multiple applications. These include training, wage and salary administration, placement, promotions, discharge, and personnel research

(Muchinsky, 2006). This study will focus on the use of performance appraisals for the purposes of faculty promotion, tenure, and annual review.

The following literature review will discuss the areas of faculty performance that typically are evaluated, followed by a description of the criteria used to define job performance. Next, faculty performance appraisal and methods for collecting performance appraisal data will be addressed. Ethical considerations concerning performance appraisal will be addressed third. Finally, this literature review will conclude with a discussion of the literature related to effective communication of the performance appraisal results back to the faculty members.

Areas of Faculty Performance to be Evaluated

For the most part, faculty evaluations consider teaching, research and creative activities, outreach/professional practice/engagement, and citizenship. Teaching includes all activities related to the development of students and to the faculty member as a teacher (Braskamp, 2005). Research and creative activities include all behaviors associated with conducting research or completing creative projects. Outreach/professional practice/engagement refers to a faculty member's applied work, while citizenship refers to services that the faculty member contributes to the campus and community. Faculty disagree on what aspects of performance are most important, but most agree that research is most easily evaluated and that teaching and service are inadequately measured (Braskamp & Ory, 1994). Evaluations should take into account differences between faculty members with regard to each of these areas. Merit and worth should also be considered during the evaluation process.

Criteria Use in Performance Appraisal

Job performance can be evaluated using several different criteria. Criteria can be classified into three types of data: objective production data, personnel data, and judgmental data. Objective production data are quantitative (Muchinsky, 2006), for example the number of publications a faculty member had in a given time period or the number or student credit hours generated. There are two problems associated with using this type of data. First, differences in performance are not always under the control of the faulty member. For example, a faculty member who teaches an upper-level class is likely to have fewer student credit hours than a faculty member who teaches a lower-level class. This difference is due to the number of students interested in taking each class; it is not due to differences in faculty performance. Second, this type of data typically does not present a complete picture of job performance.

Personnel data include areas of employee performance such as tenure, turnover, absenteeism, or accidents. The main issue with personnel data is relevance. Similar to objective production data, personnel data usually do not capture all aspects of job performance. The final type of data, judgmental data, can be applied to most jobs and attempts to capture a more complete picture of job performance. Judgmental data are most commonly collected using rating scales. Other collection methods include employee comparisons, checklists, and behavioral descriptions. Judgmental data are often subjective and are usually collected by the supervisor (Muchinsky, 2006).

Faculty members are normally evaluated based on teaching, service, and research. Research is considered the most fairly rated because it is the most objective area of faculty performance, while teaching is more often based on student ratings and service is

rarely judged at all (Centra, 1993). Evaluating faculty using these criteria can be difficult if these dimensions are not properly defined with acceptable standards of behavior (Shedd, 2005). Standards refer to the level of performance an employee is expected to achieve (Centra). In most cases, it is the department head's duty to ensure faculty members understand what behavior is expected. Expected performance standards should be specific to the faculty member. Even more important is that the department head is able to communicate these expectations to faculty members at the beginning of the appraisal period to avoid disagreements. Finally, it is important that department heads communicate with faculty members regarding their performance. This communication includes providing documentation regarding the criteria and standards used to evaluate faculty and routine informal evaluations about performance, in addition to the annual formal evaluation (Shedd).

Methods of Collecting Appraisal Data

Faculty evaluations can be conducted using multiple methods, including rating scales, interviews, written essays, observations, and checklists (Braskamp, 2005). Ratings are usually found in the form of a graphic rating scale, employee comparisons, or checklist (Guion, 1998). This study will focus specifically on graphic rating scales, employee comparisons, prescaled checklists, behavioral descriptions, forced choice scales, and distributional rating methods. Each rating method has its own strengths and weaknesses. Therefore, the method chosen should be based on how well it fits the needs of the institution (Guion).

Graphic Rating Scales

Graphic rating scales are the most common type of rating scale (Guion, 1998) and the most commonly used performance appraisal method (Muchinsky, 2006). Although they can be used to evaluate overall performance, graphic rating scales are more often used to rate specific dimensions of performance (Guion). The most common configuration includes five or seven scale points (Muchinsky) with average as the central scale point (Guion). However, there is no set number of scale divisions on a graphic rating scale. Some researchers have limited scale points to nine (Jacobs, 1986; Landy & Farr, 1980), but there is little evidence supporting these decisions (Guion). Research has not shown that the number of scale points is an important factor, so the choice is ultimately guided by the researcher. McKelvie (1978) conducted two experiments that demonstrated that scales with larger numbers of anchor points did not offer any advantage over scales with a smaller number of anchor points. In fact, this study concluded that five to six anchor points is optimal.

Employee Comparisons

Employee comparisons evaluate performance by comparing employees to each other according to some standard (Muchinsky, 2006). These comparisons can be made globally or on specific dimensions of behavior (Guion, 1998). This method helps to eliminate some common rating errors because variance is forced into the appraisals. Specifically, central-tendency and leniency errors are avoided because raters must distinguish between the ratees. There are three types of employee comparison methods: rank-order, paired-comparison, and forced distribution (Muchinsky).

Using the rank-order method employees are ranked from best to worst as defined by particular criteria (Muchinsky, 2006). Alternative ranking is one variation of this method. Using this method the rater is asked to choose the best and worst performers. These names are removed from the list of employees and the rater is asked to choose again the best and worst performers. This process continues until all the employees have been ranked (Guion, 1998). The entire ranking process is relative to some standard of behavior. One problem with this method is that we do not know the absolute degree to which an employee is good or bad at the behaviors; that is, we have only normative data and not criterion referenced data. Another problem with this method is that rankings may become meaningless when conducted for large groups of employees. The best and worst performers may be easily picked out of the group, but as the rater attempts to distinguish between employees with similar performance the accuracy of the rankings declines (Muchinsky).

The method of forced distribution involves assigning each person to a performance level while maintaining a normal distribution across performance levels (Guion, 1998). This method is used when a large number of employees need to be rated at once. Although it can be used to make ratings on multiple criteria, forced distribution is most commonly used to rate employees on only one dimension of performance (Muchinsky, 2006). This method has some similarities to graphic rating scales. Typically, there will be five anchors or categories and each category will be assigned to some percentage of the employees that are to be evaluated. The rater places a certain percentage of employees into each category using the predetermined normal distribution percentages. This method eliminates error by forcing raters to spread ratings across the

entire distribution. One problem with this method is that it is appropriate only when employee performance is normally distributed, not negatively skewed. If most employees are performing well, forcing them into a normal distribution erroneously indicates that some percentage of the employee population is performing poorly.

Finally, the paired comparison method compares each employee with all the other employees in the group being evaluated. Each employee is paired with every other employee and the rater must decide which employee is performing the specific behavior better between the two ratees. Usually the dimension rated while using this method is overall ability to perform the job. The number of times an employee was chosen as the better performer is calculated; then employees are ranked accordingly (Muchinsky, 2006). Important points to remember when conducting paired comparisons is that the same name should not appear in two consecutive pairs and each person should be listed first and second equally often (Guion, 1998). Similar to the rank order method, paired comparisons do not work well with large groups. When the group is large, there are many comparisons to make, which causes the process to be long and cumbersome (Muchinsky).

Prescaled Checklists

Prescaled checklists are based on attitude scaling. The two most common types are the method of equal-appearing intervals and the method of summated rating (Likert, 1932). Using the method of equal-appearing intervals, a rater will check statements from a list that apply to the employee being rated. These statements have been previously scaled so that differences occurring a similar number of times are considered equal (Thurstone, 1928). After the statements have been checked a total rating is computed

(Guion, 1998). The method of summated ratings uses a graphic rating scale. For each statement, the rater indicates the level of agreement and frequency of behavior. Each statement has been prescaled to determine its value. The final rating is calculated by summing the values of the ratings given for each statement.

Behavioral Rating Scales

Behavioral rating scales help to eliminate ambiguity by having raters rate specific job behaviors. There is greater agreement between raters about the performance being evaluated, which increases the accuracy of the rating (Muchinsky, 2006). Behavioral based performance appraisals are able to encompass more job complexity and are easily related back to the work done on the job. They also help to eliminate extraneous factors in the appraisal process (Latham & Wexley, 1994). Smith and Kendall (1963) were among the first to postulate using behavioral anchors for rating scales. They proposed a systematic approach, which included domain sampling, involving the raters in the rating process, and the development of clear, meaningful behavioral anchors. Further research was conducted by Bernardin and Smith (1981) that expanded the approach described by Smith and Kendall. There are two commonly used types of behavioral rating scales: behaviorally anchored rating scales (BARS) and behavioral observation scales (BOS; Guion, 1998).

Behaviorally Anchored Rating Scales (BARS). BARS combine critical incidents with a rating scale. Critical incidents are examples of what constitutes good and poor employee performance. The scale is similar to a graphic rating scale but each scale anchor is a critical incident that exemplifies the type of behavior exhibited at a specific level of performance (Muchinsky, 2006). An advantage to the BARS method is that it is

face valid. There are several disadvantages to the BARS method. First, BARS are specific to the job being evaluated. Secondly, job performance may be situational such that no one critical incident could completely exemplify the behavior needed for good performance. Third, BARS contain *examples* of performance at different levels and are not an exhaustive list of descriptors of job behavior.

Behavioral Observation Scales (BOS). BARS give examples of behaviors an employee might have done or could be expected to do. BOS on the other hand provide examples of behavior that are actually observed on the job. The behavioral statements usually are derived using a critical incidents job analysis. BOS uses a graphic rating scale which ranges from 1 (*almost never*) to 5 (*almost always*). The five points represent the frequency at which the employee was observed performing the behavior (Guion, 1998). An advantage to BOS is that it is content valid because the behavioral statements are derived from job analysis (Muchinsky, 2006). The biggest criticism of this method is that the rater must actually observe the employee doing the behavior described on the scale (Latham & Wexley, 1994). If the employee engaged in the behavior but was not observed by the rater, the rater cannot make accurate ratings regarding the behavior.

Other Rating Methods. There are two other types of rating scales that are used to rate performance, forced choice scales and the distributional rating method. Forced choice scales use groups of four descriptive statements. The statements in each foursome are prescaled. Each foursome consists of two pairs of statements that are matched on desirability (i.e., they appear to be equally desirable) but differ on discriminability (i.e., only one statement is related to effective job performance). To rate employees, raters identify the statement that best describes the employee's behavior and the statement that

least describes the employee's behavior. Selecting statements that are related to effective performance as "most" and selecting statements unrelated to effective performance as "least" results in higher scores. Forced choice scales give valid ratings (Guion, 1998) and reduce leniency errors (Latham &Wexley, 1994), but are disliked by raters because there is less control of the evaluation (Guion).

While most rating systems do not take into account that an employee's performance can change over time, the distributional rating method takes into account the variability in employee performance (Guion, 1998). Kane (1986) developed this method, which includes a distribution of outcome efficacy levels. These efficacy levels compose a scale that ranges from least effective outcome to most effective outcome. One other component of this method is a record of employee outcomes. This record of employee outcomes can be created from the evaluator's memory of employee work behavior or by using a diary. Distributional rating is rarely used because implementation is extremely difficult (Guion).

Faculty Performance Appraisal

When undergoing performance appraisal, faculty members are usually required to document their accomplishments since the last time an evaluation was conducted. This report typically contains information on the areas of teaching, research, and service. Those individuals conducting the evaluations, such as administrators, department heads/chairperson, deans, and their respective advisory/executive committee, use a variety of methods to obtain an evaluation of a faculty member. Braskamp, Brandenburg, and Ory (1984) recommended that when making these overall evaluations about faculty

performance based on teaching, research, and service, each area be properly defined and that weighting be flexible to account for situational factors.

Performance appraisals can be either formative or summative. Formative evaluation provides on-going feedback related to strengths and weaknesses that informs faculty development decisions. Summative evaluations focus on outcome measures and are used to make personnel decisions such as promotion and tenure. In essence, formative evaluations may provide faculty members with the knowledge needed to improve before they are judged in a summative evaluation (Centra, 1993).

Performance appraisal data can come from several sources. Colleagues can provide information that cannot be gathered using other sources. Department heads are often considered colleagues because they work alongside the faculty (Centra, 1993). Research on department head evaluations of faculty has shown that evaluations of teaching based solely on observation do not produce the most accurate results (Centra, 1975). Feldman (1989) found that department head evaluations of teaching effectiveness correlated .48 with evaluations made by other colleagues. Centra (1993) recommended that raters utilize course syllabi, assignments, and other documents to make evaluations in addition to observation.

Research has shown that colleagues are able to evaluate research activity more accurately than teaching effectiveness (Centra, 1993). Kremer (1990) found greater agreement between colleagues when they were evaluating research than when they were evaluating teaching effectiveness. In a study using colleague evaluations based on dossiers of individual faculty members, Root (1987) found high reliabilities for teaching,

research, and service. According to Centra, the Root study supports the use of colleague evaluations as a basis for tenure and promotion decisions.

Centra (1993) indicated that most faculty members are dissatisfied with their current performance appraisal system. There are several explanations for this dissatisfaction with faculty performance appraisal. First, faculty appraisal often fails to clearly define the entire spectrum of faculty work performed (Stake, 1970). Appraisals should give employees the information and understanding needed to improve their performance. Poorly defined performance appraisal dimensions make it harder for employees to understand what behaviors they need to improve and how to improve those behaviors. Shulman (1993) indicated that more time should be spent describing faculty work.

Second, faculty appraisal is often viewed rigidly with no discussion about the appraisal process. House (1993) indicated that evaluations would be better understood if the process were more flexible and allowed for more discussion and debate. The lack of communication between the assessor and the faculty member can cause some faculty to be reluctant to be open and candid during the appraisal process (Centra, 1993). Finally, in some cases more attention has been placed on the methods used to assess faculty performance than on properly describing and judging the actual work. A performance appraisal method is only effective if faculty members are able to communicate their achievements to others (Eisner, 1993). This is difficult to do if the performance appraisal method does not adequately capture the complexity of faculty work. The use of more qualitative performance methods instead of the more commonly used quantitative methods may help to encompass the complexity of faculty work (Centra).

Communicating Results/Making Appraisal Meaningful

The results of the performance appraisal should be communicated to employees. Performance appraisal systems are often meaningless to employees if it is not understood how the data will be used (Dilts, Haber, & Bialik, 1994). This is especially important if the purpose of the appraisal is performance improvement (Braskamp et al., 1984). The communication step can be as important as data collection. The feedback process should increase the employee's knowledge and understanding of what is expected of their work (Braskamp & Ory, 1994). Burkhalter and Buford (1988) suggested creating an organizational climate that is conducive to effective communication. A supportive climate can reduce anxiety that may be experienced by the employee. Communication is an ongoing process that is often forgotten (Braskamp et al.). Success or failure of a performance appraisal system is contingent on how the information is communicated back to employees (Burkhalter & Buford).

Ethics/Legal Issues and Performance Appraisal

Performance appraisals are covered by Equal Employment Opportunity law under Title VII of the Civil Rights Act of 1964 and the Civil Rights Act of 1991. Accordingly, performance appraisal data can come under legal scrutiny any time it is used to make a personnel decision (Bernardin & Beatty, 1984). After reviewing a number of court cases, Malos (1998) concluded that discrimination cases were often related to performance evaluation. Institutions should avoid legal issues by working to prevent them before they occur. Eight guidelines have been derived from case law and are prescriptive for

avoiding legal challenges to performance appraisal (Malos, 1998, 2005; Werner & Bolino, 1997). These guidelines include basing the appraisal system on a job analysis, communicating the performance standards to the employees and raters, evaluating employees based on specific dimensions of job performance rather than on an overall measure, defining performance dimensions behaviorally, training raters to use the performance appraisal system, documenting the evaluation, establishing a formal appeals process, and providing some type of corrective guidance to assist poor performers to improve their performance. When performance appraisals are done correctly, legal issues are not a concern because the practical implications and legal implications of performance appraisal are closely related. These same eight practices also lead to the perception that appraisals are fair (Greenberg, 1986).

The Present Study

Faculty performance appraisal includes a review of the faculty member's work relative to institutional goals and includes feedback to foster faculty development. Faculty members typically are evaluated on teaching, research, and service (Centra, 1993). Well-constructed faculty performance appraisal systems should be based on the behaviors and outcomes necessary for the faculty member to adequately accomplish both individual and organizational goals (Latham & Wexley, 1994). The current study attempted to identify common and unique characteristics of faculty performance appraisal formats and procedures by analyzing characteristics of formats and procedures from the psychology departments of 28 universities, including Kentucky Council on Post-secondary Education (CPE) benchmark schools for Western Kentucky University as well

as schools that have Industrial/Organizational psychology master's degree programs. Two hypotheses were evaluated.

Hypothesis 1: Schools with Industrial/Organizational Psychology programs will have better formats as defined by the eight legal factors discussed earlier.

Hypothesis 2: Some form of a graphic rating scale (e.g., BARS, BOS, or pre-scaled checklists) will be the most common method for collecting appraisal data.

Method

Performance appraisal instruments and guidelines were collected from Western Kentucky University's CPE Benchmark Universities and schools with I/O psychology master's programs via an email that was sent to department heads and I/O program directors, respectively. The email included a letter explaining the purpose of the study. A copy of each letter may be found in Appendix A. A questionnaire asking the respondent to identify characteristics of the performance appraisal instrument used in his/her department was attached to the email. A copy of this questionnaire may be found in Appendix B.

The items on the questionnaire addressed the following issues:

- Type of method used to collect appraisal data
- Standards based on job analysis
- Performance standards communicated to employees and raters
- Employees evaluated on specific dimensions of job performance
- Performance dimensions defined behaviorally and supported by objective, observable evidence
- Raters trained to use the performance appraisal system properly
- Documentation of evaluation
- Formal appeals process established
- Corrective measures are set up to improve poor performance
- Written instructions for implementing all phases of the system are provided
- Raters have direct knowledge of the individuals they are evaluating
- Feedback provided to employees on an on-going basis
- Employees may review appraisal results

Eight schools out of 18 (44.4%) CPE benchmarks responded; 16 schools with I/O psychology master's programs responded. Five other schools which were neither CPE benchmarks nor I/O program schools also responded. A list of these schools may be found in Appendix C.

Four schools opted to send performance appraisal materials rather than answering the questionnaire. The materials from these schools were coded to determine answers to the questionnaire items. Each performance appraisal document was independently coded by two I/O psychology master's degree candidates. An I/O psychologist served to break ties in coding decisions. For these four schools (i.e., Texas State San Marcos, University of South Alabama, Marist College, Southern Illinois University Edwardsville), inter-rater reliability was calculated using a chi-square analysis, $X^2(16) = 507.52$, $p = .00$, which indicated a high level of rater agreement. Three schools submitted forms completed by two different raters. Inter-rater agreement for these three schools, calculated using Cohen's kappa coefficient, was .83 for Western Kentucky University, .80 for Murray State University, .60 for California State University San Bernadino.

Results

The first hypothesis, that schools with Industrial/Organizational psychology master's programs would have better formats, as defined by a positive response to questionnaire items 1 to 17, was determined by first summing the number of times each university format received a positive score across the items. The results of this process may be found in Appendix C. Eight schools (4 I/O and 4 non-I/O) had insufficient responses to include in this analysis. Next, an independent samples *t*-test was conducted to test for statistical significance between the mean for I/O-program schools (M = 10.23, SD = 3.85) and non-I/O schools (M = 11.0, SD = .87). The independent samples *t*-test conducted to test Hypothesis 1 was not significant, $t\ (17) = -.43$, $p = .67$.

The second hypothesis, that graphic rating scales (GRS; e.g., BARS, BOS, or pre-scaled checklists) would be the most common method for collecting performance appraisal data, was evaluated by first counting the number of times each type of format was used. Each format was coded as a GRS (1) or not a GRS (0). A *z*-test for proportions (test value = .5) was conducted. The *z*-test for proportions conducted to test Hypothesis 2 was significant, $z = 1.70$, $p < .05$, one-tailed. Eighteen of the 28 formats (64.30%) were some form of a graphic rating scale.

The extent of consistency of Western Kentucky University's performance appraisal system compared to the CPE benchmark universities was evaluated even though no hypotheses were made regarding this question. First, the percentage of yes and no responses given by the eight CPE benchmark universities was calculated for each item on the questionnaire. A typical profile was then created using the majority response for each questionnaire item. Each characteristic was either typical (1; more than 50% of the

responding CPE schools had this characteristic in their performance appraisal system) or not typical (0; fewer than 50% of responding CPE schools had this characteristic) of a CPE benchmark. The percentage of yes/no responses for each characteristic across CPE benchmarks (not including WKU) may be found in Appendix D, along with the responses for WKU. The number of characteristics on which WKU's performance appraisal was consistent with the typical CPE characteristic was then determined. Western Kentucky University questionnaire responses matched the typical CPE profile on 15 out of 17 questionnaire items, indicating that the WKU performance appraisal system is very similar to the typical CPE profile. The two items on which WKU did not match the CPE profile were items which CPE respondents endorsed 50% of the time (i.e., a job analysis was used and department head/chair assigns rating) and the WKU Psychology Department had present in its appraisal system.

Discussion

It was hypothesized that departments with Industrial/Organizational programs would have better performance appraisal formats, as defined by the questionnaire items, because industrial/organizational psychologists would be expected to be more familiar with the requirements for a good performance appraisal instrument as they are an important component of the discipline of I/O psychology. The results of the analysis did not support this hypothesis. There was no significant difference between the schools that had Industrial/Organizational programs and those that did not. This could be due to several factors. First, even if there is no Industrial/Organizational program there still may be an industrial/organizational psychology faculty member on staff. Secondly, the department could be cognizant of the practical and legal implications for performance appraisals regardless of the presence of an Industrial/Organizational program or faculty member.

The second hypothesis, that graphic rating scales (e.g., BARS, BOS, or pre-scaled checklists) would be the most common method for collecting performance appraisal data, was based on the fact that graphic rating scales are the most common type of rating scale (Guion, 1998). Of the 28 schools in the sample, 18 used graphic rating scales and 10 did not. This finding is consistent with the literature that, in practice, graphic rating scales are the most commonly used performance appraisal format. Of the 10 schools that did not use graphic rating scales, four used a written essay format, five responded other, and one did not respond. This finding further supports the prevalent use of graphic rating scales in performance appraisal.

When comparing Western Kentucky University questionnaire responses to the typical CPE profile, the responses matched for 15 out of 17 of the questionnaire items. The items that did not match Western Kentucky University's response on the typical profile were present in the WKU performance appraisal process but were endorsed by only 50% of the CPE benchmark schools. These items concerned the use of a job analysis and the use of department heads/chairs as performance appraisal raters. Western Kentucky University utilizes both job analysis and the department head serves as the rater.

The CPE benchmark schools for Western Kentucky University are schools that represent a standard by which WKU should be evaluated. In theory and, as this study demonstrates, in practice the performance appraisal system used by Western Kentucky University's Psychology Department should be consistent with the performance appraisal systems of the psychology departments of the CPE benchmark schools. There is a great deal of consistency across the performance appraisal systems used in the psychology departments of the CPE benchmark schools.

There are at least two limitations to the present study. First, the questionnaire responses were self-reported by the department heads or faculty members of the various schools that responded to the request to participate. While department heads and faculty should have considerable knowledge of the performance appraisals in their own departments, there was no check on the validity of the reported data. Second, only eight of the 18 CPE benchmark schools responded to the request to participate and only 20 of the 40 or so other schools contacted responded. While this is 48% response rate is

acceptable for survey research, there was no check on the representativeness of the survey respondents.

Faculty evaluation is an important process. A practical, legal performance appraisal system can help ensure appropriate management of faculty performance and that the university retains the best faculty members. It was of interest to evaluate the performance appraisal system of the Psychology Department at Western Kentucky University against those of comparable schools, such as the CPE benchmark schools and the Industrial/Organizational psychology schools used in this research. Thus, the purpose of the present study was to determine if the WKU Psychology Department appraisal system contained the components necessary for a successful performance appraisal and if those components were consistent with those found in comparable schools.

In summary, this study analyzed faculty performance appraisal formats and procedures of Kentucky CPE benchmark schools for Western Kentucky University as well as schools that have Industrial/Organizational psychology master's degree programs to identify common and unique characteristics. It was hypothesized that schools with Industrial/Organizational Psychology programs would have better formats as defined by the eight legal factors identified earlier. This hypothesis was not supported. It also was hypothesized that a graphic rating scale (e.g., BARS, BOS, or pre-scaled checklists) would be the most common method for collecting appraisal data, which was supported. It was determined that the performance appraisal system used at Western Kentucky University is very similar to systems used at the CPE benchmark schools that participated in this study.

References

Bernardin, H. J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored rating scales (BARS). *Journal of Applied Psychology, 6*, 205-212.

Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent Publishing.

Braskamp, L. (2005). Faculty assessment serves institutional, individual needs. *Academic Leader, 21*, 8-8.

Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. San Francisco: Jossey-Bass Publishers.

Braskamp, L. A., Brandenburg, D. C., & Ory, J. C. (1984). *Evaluating teaching effectiveness: A practical guide*. Beverly Hills: Sage Publications, Inc.

Burkhalter, B. B., & Buford, J. A., Jr. (1988). *Performance appraisal: Concepts and techniques for postsecondary education.* Washington, DC: The American Association of Community and Junior Colleges.

Centra, J. A. (1975). Colleagues as raters of classroom instruction. *Journal of Educational Measurement, 46*, 327-337.

Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass Publishers.

Dilts, D. A., Haber, L. J., & Bialik, D. (1994). Assessing what professors do: An introduction to academic performance appraisal in higher education. Westport, CT: Greenwood Press.

Eisner, E. (1993). *Forms of understanding and the future of education research.*

Presidential address at the annual meeting of the American Education Research Association, Atlanta.

Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators and external (neutral) observers. *Research in Higher Education, 30,* 137-189.

Greenberg, J. (1986). Determinants of perceived fairness of performance appraisal. *Journal of Applied Psychology, 71*(2), 340-342.

Guion, R. M. (1998). *Assessment, Measurement, and Prediction for Personnel Decisions.* Mahwah, NJ: Erlbaum.

Jacobs, R. (1986). Numerical rating scales. In R. A. Berk (Ed.), *Performance assessment: Methods and application.* Baltimore: Johns Hopkins University Press.

House, E. R. (1993). *Professional Evaluation: Social Impact and Political Consequences*. Newbury Park, CA: Sage.

Kane, J. S. (1986). Performance distribution assessment. In R. A. Berk (Ed.), *Performance assessment: Methods and applications* (pp. 237-273). Baltimore: Johns Hopkins University Press.

Kremer, J. (1990). Constant validity of multiple measures in teaching, research, and service and reliability of peer ratings. *Journal of Educational Psychology, 82,* 213-218.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87,* 72-107.

Latham, G. P., & Wexley, K. N. (1994). *Increasing Productivity Through Performance Appraisal* (2nd Edition.). Reading, MA: Addison Wesley.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology,*

140, 44-53.

Malos, S. B. (2005). The importance of valid selection and performance appraisal: Do management practices figure in case law? In F. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives*. San Francisco: Jossey Bass.

Malos, S. B. (1998). Current legal issues in performance appraisal. In J. W. Smither (Ed.), *Performance appraisal* (pp. 49-94). San Francisco: Jossey-Bass.

McKelvie, S. J. (1978). Graphic rating scales—how many categories? *British Journal of Psychology, 69,*185-202.

Muchinsky, P. M. (2006). *Psychology applied to work*. Belmont, CA: Thomson Wadsworth.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage.

Root, L. S. (1987). Faculty evaluation: Reliability of peer assessments of research, teaching, and service. *Research in Higher Education, 26,* 71-86.

Rotchford, N. L. (2002). Performance management. In J. W. Hedge & E. D. Pulakos (Eds.), *Implementing organizational interventions: Steps, processes, and best practices* (pp. 167-197). San Francisco: Jossey-Bass.

Shedd, P. (2005). Reducing ambiguity, awkwardness in evaluating faculty. *Academic Leader, 21,* 5-8.

Shulman, L. S. (1993). Assessments for teaching: An initiative for the profession. *Phi Beta Kappan, 69,* 38-40.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: an approach to the

construction of unambiguous anchors for rating scales. *Journal of Applied Psychology, 47,* 149-155.

Stake, R. E. (1970). Objectives, priorities, and other judgment data. *Review of Educational Research, 40,* 181-212.

Stufflebeam, D. L. (1988). *The Personnel Evaluation Standards: How to Access Systems for Evaluating Educators*. Newbury Park, CA: Sage.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33,* 529-554.

Werner, J. M., & Bolino, M. C. (1997). Explaining U.S. court of appeals decisions involving performance appraisal: Accuracy, fairness, and validation. *Personnel Psychology, 50,* 1-24.

APPENDIX A:

Letter Sent to CPE Benchmark Schools and Schools with I/O Psychology Master's Degree Programs

CPE Letter

Our Psychology Department at Western Kentucky University is in the process of revising our performance appraisal system. As part of that process I am trying to collect information about appraisal systems in Psychology Departments at our Kentucky Council on Post-Secondary Education Benchmark universities. Your school is one of our benchmark institutions. **Attached is a brief questionnaire about performance appraisal in your department. The questionnaire should take less than 10 minutes to complete.**

If possible, I would appreciate receiving:
1. Your response to the attached, brief multiple-choice questionnaire about your appraisal system.
2. A copy of your performance appraisal format
3. Any guidelines your department or university has for the implementation of the appraisal

Departmental performance appraisal formats are preferred; however any appraisal format your department uses (e.g., college or university format) would also be appreciated.

If you have your format and guidelines electronically, I would appreciate it if you would attach them along with the completed questionnaire in a reply email to:
steven.haggbloom@wku.edu
(Or, if your appraisal format is available online, you can send the url along with the completed questionnaire.)

If they are not available in electronic format, I would appreciate receiving a hard copy at:

Dr. Steven Haggbloom
Head, Department of Psychology
Western Kentucky University
1906 College Heights Blvd. # 21030
Bowling Green, KY 42101-1030

Even if you are not able to send a copy of your formant, please return the completed questionnaire. I would appreciate your response by March 14th.

Thanks so much for your help with this effort.

Steve Haggbloom

I/O Program Letter

Our Psychology Department at Western Kentucky University is in the process of revising our performance appraisal system. As part of that process I am trying to collect information about appraisal systems in Psychology Departments – especially those with I/O graduate programs. My assumption is that departments with I/O programs should have I/O faculty who should know something about performance evaluation and, as a consequence, should have solid PA systems. (That is why you received this email rather than your department head).

I would very much appreciate receiving:
1. Your response to the attached, brief multiple-choice questionnaire about your appraisal system.
2. A copy of your performance appraisal format
3. Any guidelines your department or university has for the implementation of the appraisal

Departmental performance appraisal formats are preferred; however any appraisal format your department uses (e.g., college or university format) would also be appreciated.

If you have your format and guidelines electronically, I would appreciate it if you would attach them along with the completed questionnaire in a reply email to:
betsy.shoenfelt@wku.edu
(Or, if your appraisal format is available online, you can send the url along with the completed questionnaire.)

If they are not available in electronic format, I would appreciate receiving a hard copy at:

Dr. Betsy Shoenfelt
Department of Psychology
Western Kentucky University
1906 College Heights Blvd. # 21030
Bowling Green, KY 42101-1030

I would appreciate it if you could return the completed questionnaire and the appraisal materials by March 14th.

Thanks so much for your help with this effort.

Betsy Shoenfelt

APPENDIX B: Questionnaire

Name of School: _____

Please consider the performance appraisal used to evaluate faculty in your department. Respond to each item by marking either the "Yes" or "No" column. Thank you!

No.	Items	Yes	No
1.	Are written instructions provided for the use of the appraisal system?		
2.	Is the appraisal system based on a job analysis? A job analysis is a comprehensive study of a job (i.e., faculty member) that identifies important job responsibilities/tasks and/or underlying knowledge, skills, and abilities needed for successful job performance.		
3.	Are faculty evaluated on specific dimensions of job performance rather than on a global overall measure?		
4.	Are performance dimensions (e.g., teaching, research, service, etc.) defined in behavioral terms?		
5.	Are performance standards (i.e., the level of performance needed to meet expectations) clearly communicated to employee prior to the beginning of the appraisal period?		
6.	Do raters have direct knowledge of the individuals they are evaluating?		
7.	Is the person who assigns the ratings given guidance on the use of the instrument used to evaluate faculty?		
8.	Is feedback given on an on-going basis (i.e., year round)?		
9.	Is documentation (e.g., an annual activity report) provided by the faculty member to be used in the evaluation process?		
10.	Are the ratings assigned by the department chair/head?		
11.	Are performance results reviewed with the faculty member?		
12.	Is there an improvement plan for employees who perform below standard?		
13.	Does the department head document the basis for appraisal ratings?		
14.	Is there a formal appeals process available to employees?		
15.	Is your performance appraisal system determined by a union contract?		
16.	Did faculty participate in the development of your performance appraisal system?		
17.	Is this a department-specific performance appraisal system?		
	-or is it college-wide?		
	-or is it university-wide?		
18.	Does your department have an Industrial/Organizational Psychology Master's level program?		
19.	Are you an Industrial/Organizational Psychologist?		

20. Please mark the type of format used to conduct faculty evaluations.

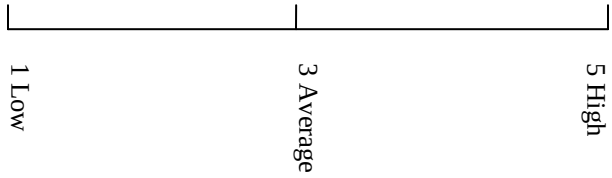
_____Graphic Rating Scales
_____Employee Comparisons
_____Written Essays
_____Checklists
_____Behaviorally Anchored Rating Scales
_____Behavior Observation Scales
_____Observational Checklist
_____Other

_____

Examples of different formats follow on the next page.

GRAPHIC RATING SCALE

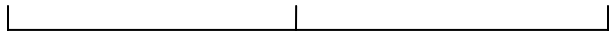
Teaching



BEHAVIORALLY ANCHORED RATING SCALE

Behavioral anchors describe expectations at different levels of performance on a specific dimension.

Teaching

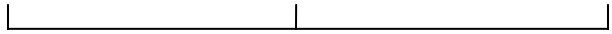


1: Behavioral Description of Low Job Performance: e.g., Rarely has syllabus prepared for each course.	3: Behavioral Description of Average Job Performance: e.g., Has a syllabus prepared for each course prior to the start of the semester.	5: Behavioral Description of High Job Performance: e.g. Has a syllabus posted on-line for each course that provides learning objectives and comprehensive overview of course, timetable, and other information relevant to student success.

BEHAVIOR OBSERVATION SCALE

Behavioral anchors describe frequency with which behaviors have actually been observed on the job.

Teaching: Has syllabus describing specific course objectives.



1:Rarely	3: Often	5: Almost Always

EMPLOYEE COMPARISONS

Employees are compared to every other employee on dimensions of job behavior.

Teaching: Circle the faculty member in each pair who is better at teaching.

Jones vs. Smith	Brown vs. Jones
Brown vs. Smith	Smith vs. Moore
Moore vs. Jones	Moore vs. Brown

APPENDIX C: List of Schools with Sum of Item Scores

List of Schools with Sum of Item Scores

School	Type of School		Sum Score
	CPE	I/O	
Avila University	No	No	Too many missing variables
Ball State University	Yes	No	9
California State University- San Bernardino	No	Yes	12
Creighton University	No	Yes	7
Eastern Michigan University	Yes	Yes	15
Indiana State University	Yes	No	10
IUPUI	No	Yes	Too many missing variables
Marist College	No	No	Too many missing variables
Middle Tennessee State University	Yes	No	12
Minnesota State University	No	Yes	5
Missouri State University	Yes	Yes	14
Missouri University of Science of Technology	No	Yes	9
Murray State University	No	No	13
Northern Kentucky University	No	Yes	12
Radford University	No	Yes	6
Southern Illinois University- Edwardsville	No	Yes	10
Texas State- San Marcos	No	No	Too many missing variables
Towson University	Yes	No	12
University of Connecticut	No	No	Too many missing variables
University of Idaho	No	Yes	Too many missing variables
University of Northern Iowa	Yes	Yes	6
University of Redlands	No	Yes	Too many missing variables
University of South Alabama	No	No	Too many missing variables
University of Tennessee at Chattanooga	No	Yes	Too many missing variables
University of Wisconsin	No	Yes	4
Western Illinois University	Yes	No	15
Western Kentucky University	Yes	Yes	14
Xavier University	No	Yes	14

APPENDIX D:

Summary of Comparison of Faculty Appraisal Process at WKU to CPE Benchmark Schools

Summary of Comparison of Faculty Appraisal Process at WKU to CPE Benchmark Schools

No.	Items	Yes	No	WKU
1.	Are written instructions provided for the use of the appraisal system?	100	0	1
2.	Is the appraisal system based on a job analysis? A job analysis is a comprehensive study of a job (i.e., faculty member) that identifies important job responsibilities/tasks and/or underlying knowledge, skills, and abilities needed for successful job performance.	50	50	1
3.	Are faculty evaluated on specific dimensions of job performance rather than on a global overall measure?	75	25	1
4.	Are performance dimensions (e.g., teaching, research, service, etc.) defined in behavioral terms?	75	25	1
5.	Are performance standards (i.e., the level of performance needed to meet expectations) clearly communicated to employee prior to the beginning of the appraisal period?	75	25	1
6.	Do raters have direct knowledge of the individuals they are evaluating?	100	0	1
7.	Is the person who assigns the ratings given guidance on the use of the instrument used to evaluate faculty?	62.5	37.5	1
8.	Is feedback given on an on-going basis (i.e., year round)?	62.5	37.5	0
9.	Is documentation (e.g., an annual activity report) provided by the faculty member to be used in the evaluation process?	100	0	1
10.	Are the ratings assigned by the department chair/head?	50	50	1
11.	Are performance results reviewed with the faculty member?	75	25	1
12.	Is there an improvement plan for employees who perform below standard?	75	25	1
13.	Does the department head document the basis for appraisal ratings?	87.5	12.5	1
14.	Is there a formal appeals process available to employees?	87.5	12.5	1
15.	Is your performance appraisal system determined by a union contract?	37.5	62.5	0
16.	Did faculty participate in the development of your performance appraisal system?	87.5	12.5	1
17.	Is this a department-specific performance appraisal system?	75		1
	-or is it college-wide?	12.5		0
	-or is it university-wide?	12.5		0
18.	Does your department have an Industrial/Organizational Psychology Master's level program?			
19.	Are you an Industrial/Organizational Psychologist?			

Note: Entries in the Yes/No columns refer to the percentage of 8 CPE schools that responded to each option. Entries in the WKU column indicate yes (1) or no (1).