Clear Logic and Fuzzy Guidance:

A Policy Capturing Study of Merit Raise Decisions

Lizabeth A. Barclay
Kenneth M. York

Policy capturing was used to determine cue weights when a merit raise committee implemented an imprecise directive. Evaluations by three raters of 36 faculty were regressed on actual raises. The committee was consistent in their evaluations, but the policy was similar to that obtained by counting activities in faculty annual reports. This study has implications for organizations that motivate employees through merit pay decisions in ways that are inconsistent with their mission and business objectives.

Merit pay remains an area of controversy. Although merit pay is widely accepted as a way to recognize performance, merit pay systems are problematic.[1] Merit pay systems can de-motivate and generate anxiety, and have been found to be only "marginally" significant as motivators of desired performance.[2,3] For merit systems to be effective, organizations must consider the relationship of such systems to business objectives.

Merit pay is theoretically based on individual performance, although it could also be based on the performance of a group or team.[4] Generally, merit pay is rolled into base salary; therefore, a merit award in one year continues to have an effect on salary in the future.[1] Part of the controversy surrounding merit pay use is related to its dependence on performance appraisal. Organizations considering such systems must believe there is sufficient performance variation, that the variation is measurable, and that employees want to be assessed in this manner.[5]

The merit pay debate occurs not only in private sector organizations; the usefulness of merit pay is also debated in public settings such as academe.[6,7] While universities have performance review systems related to the granting of tenure and promotion, annual merit reviews may — or may not — capture the performance dimensions articulated by the universities for faculty, namely, research, teaching, and service (a university's business objectives). In addition, in the case of merit pay reviews, the outcome has theoretically more variability than a tenure review. In the tenure review, the decision is either to grant or deny tenure, but in an annual merit pay decision, faculty can be assigned a range of increases. Merit pay decisions are typically made by comparing performance across faculty, but rarely would all faculty be

undergoing tenure reviews at the same time. Therefore, in terms of performance, it may be easier to capture the policy of a merit pay committee than a tenure committee, based on the number of individuals being evaluated.

This study uses policy capturing to determine the weights an actual merit raise committee gave to different sources of information. The results will be discussed in terms of effective management, and the relationship of merit evaluations to stated "business" objectives. While this study used data from a public university, we believe that there are implications for other organizations. While many organizations use merit pay, access to actual performance data and ratings for a set of employees is often unobtainable. This data set permits an examination of merit decisions made by real evaluators about real people.

This study was conducted in the business school of a public university with a total enrollment of 14,000 students. Like many universities of its size, it is a research institution with a strong emphasis on undergraduate education. In the university's strategic plan, undergraduate education is described as "central to its mission," graduate and professional education is "responsive to regional and national needs," and research, scholarship, and creative activities are "aggressively encouraged and supported." The business school's criteria for tenure and promotion describe a "balance" among research, teaching, and service, with research receiving the greatest weight for tenure decisions (candidates for tenure must be productive scholars), and an even greater weight in promotion to full professor decisions (candidates for full professor must show evidence of academic maturity, and scholarly activity is the primary focus of the review). Therefore, the performance dimensions are similar to those found in the higher education pay literature and are considered in the current paper to reflect organizational objectives.

The university where this study was conducted has identified performance dimensions for faculty. However, what weight is to be assigned to these dimensions? A survey of business schools found that professors thought research and publication were given 49% of the weight in merit pay decisions, compared to 30% for teaching, 15% for service, and the remainder to other factors.[7] Weights can vary dramatically based on the overall objectives of a particular business school, for example doctoral-granting-versus non-doctoral-granting schools.[6]

This study was designed to answer four questions. First, how consistently (an internal consistency measure of reliability for each rater) can members of a merit committee apply a faculty-ratified directive when making merit raise judgments? Performance evaluation training may not occur, and theoretically, each rater could be inconsistent in his/her evaluations. A university faculty member is typically evaluated on research, teaching, and service, with different universities (or different units within the same university) having varying weights assigned to the three areas.

Second, when a merit committee is given an imprecisely stated merit raise directive that requires it to consider multiple factors with emphasis given to one factor, how does it weight the factors? Although some policy capturing studies have shown

that individual judges can use multiple cues,[8] can a merit committee that must come to a consensus judgment use multiple cues?

Third, do members of a merit committee charged with evaluating faculty generally agree in their evaluations of faculty members' research, teaching, and service (an interjudge agreement measure of reliability)? Just as there may be differences across units of a university, there may be differences within a unit.

Fourth, can the policy of a merit raise committee be captured by assigning values to research, teaching, and service based on "objective data" contained in faculty annual reports and student course evaluation data? Although making merit raise judgments may be a cognitively complex task requiring simultaneous consideration of many pieces of information, a simple linear model might effectively capture the policy of a merit committee.[9]

Method

Participants

Faculty in this business school delegate the responsibility for evaluating each faculty member's job performance over the past year and assigning a merit pay increase to a Merit Committee. The Merit Committee is composed of the four department chairs and the Dean as ex officio. The faculty ***and*** department chairs at this university are unionized (in the same bargaining unit), and have the contractual right to specify how merit pay increases should be allocated and who will make the decisions. The unit's recommendations are forwarded to the administration, which may choose to return them for suggested changes, but may not unilaterally modify the raise amounts based on the unit pool. Typically, the Merit Committee's recommendations are implemented with few or no changes. The charge the faculty have given the Merit Committee is very specific on some details ("faculty returning from a full-year sabbatical leave shall receive an average increase"), but less so on others ("the Merit Committee shall consider each faculty member's contributions in research, teaching, and service, with special consideration for research").

The Merit Committee is provided with standardized annual reports submitted by each faculty member describing their activities for the past year in research, teaching, and service, and any other information they may wish to provide. The Merit Committee also has access to summaries of student course evaluations (means and frequencies for each question for each course, and the means for each question for each department and the School of Business Administration), which are reviewed by committee members prior to the raise meeting.

In a daylong session, the Merit Committee evaluates each faculty member and comes to a consensus judgment about each faculty member's merit pay increase as a percent of his/her current base salary. The amount of money in the merit pay pool is based on the contractually specified annual pay increase. Although the chairs are in the same bargaining unit, they do not make decisions on their own raises; the Dean allocates a separate chair merit pool, based on the Dean's own criteria.

Materials

The business school has a past practice of "partial" pay secrecy. Because it is part of a public university, salary data are publicly available (i.e., the university budget is on reserve in the library), but few faculty seek the information. The ***only*** information the Merit Committee shares directly with the faculty is a letter to each faculty member indicating the range of salary increases, the raise percentage, and his/her new salary for the next academic year. In some years, the Merit Committee also has provided a grouped frequency distribution.

Three days after the raise letters were distributed for the year under study, the first author submitted a Freedom of Information Act (FOIA) request to the university to obtain the Merit Committee's worksheets, the faculty's annual reports, and the actual percentage increases given to each faculty member from the spreadsheets submitted to the university to implement the raises. Worksheets showing the Merit Committee's consensus judgments on research, teaching, and service were obtained, as well as worksheets that were available from three of the four department chairs. Some faculty members also received a "market or compression adjustment" in addition to their merit pay increase (not taken from the merit pay pool); this was not included in the merit pay increase percentage for this study because it falls outside of the faculty directive under which merit is allocated. Because the researchers were former chairs, they were able to identify the raise components on the worksheets and were familiar with the range of performance dimensions used in the process.

Obtaining the Committee's ratings on research, teaching, and service, and the actual merit raise increase percentages for this study required the authors to read the handwritten notes of the department chairs for their individual ratings on research, teaching, and service, and handwritten notations on a spreadsheet for the actual merit pay increase amounts. The consensus ratings were also handwritten on one of the judge's worksheets, therefore, some data recording errors may have been made. However, the authors independently examined the data sheets to minimize problems.

Analysis

Policy capturing is a widely used application of social judgment theory, and was used in this study to examine merit pay increase judgments. The process is based on Brunswik's probabilistic functionalism theory of perception.[10] Brunswik advocated that experiments should be designed to be representative of the natural ecology in which the behavior occurs. Policy capturing has been used to investigate a wide variety of decision making processes, including the rating of livestock, the effect of social desirability response bias on judgments, decisions to be absent from work, judgments about sexism, and the accuracy of venture capitalists' judgments when given different amounts of information.[8,11,12,13,14] In the typical policy capturing study, judges are asked to make multiple judgments based on multiple sources of information, called cues. The values of the cues are systematically varied to produce a representative set of cue combinations, called profiles. If the complete factorial of cue combinations is given to each subject, the cues will have zero intercorrelations. Each subject makes

judgments on enough profiles to allow a regression analysis of ***each*** judge, regressing their judgments on the cues, i.e., the judgments are the dependent variable, and the cues are the independent variable. Figure 1 shows part of the data set for the Merit Committee. This regression analysis produces a judgment policy, with the beta weights indicating the relative importance of each of the cues, and the multiple correlation indicating the individual's consistency (an internal consistency measure of reliability for each judge) in making their judgments. When a complete factorial is not used (e.g., when the value of the cues are taken from the environment rather than set by the researcher) the cues are usually intercorrelated, and a better indication of the weights given to the cues are usefulness indices (the incremental multiple correlation squared when the cue is added to the regression equation last).

Figure 1. Merit Raise Committee Consensus Judgments

Faculty	Raise	Research	Teaching	Service
1	0.028	2	4	3
2	0.033	3	3.5	3.5
...				
36	0.046	5	4	4

In this study, three judges who made merit pay increase decisions for 36 faculty in the business school were examined. Each judge made his/her individual ratings on each faculty member's research, teaching, and service prior to the Merit Committee meeting, then at the meeting came to consensus on the ratings and merit pay increase percentage for each faculty member. Using their consensus judgments on merit pay increase percentages as the dependent variable, the Merit Committee's policy can be captured in two ways: 1) Using their consensus ratings of each faculty member on research, teaching, and service as the cues, and 2) Using data from the faculty annual reports and student course evaluation data to assign values to the cues.

Results

Policy Capturing

The merit pay policy of the Merit Committee was obtained by regression analysis using the merit pay percentage increase as the dependent variable, and the Merit Committee's consensus judgments on research, teaching, and service (on a scale from 0 to 5) as the independent variables. Faculty on full-year sabbaticals were excluded from the analysis, resulting in 36 cases for analysis. The Merit Committee was very consistent (internal consistency) in their judgments ($F(3, 32) = 46.69$, $p < .001$), with a multiple correlation of .90, indicating that 81% of the variance in merit pay percentage increases could be accounted for from the Merit Committee's consensus judgments on research, teaching, and service. The Merit Committee gave the most weight to

research, less weight to teaching, and very little weight to service, with beta weights for research, teaching, and service of .68, .33, and .09.

The cues were moderately intercorrelated: research-teaching .42, $p = .03$; research-service .25, $p = .43$; and teaching-service .26, $p = .40$. The usefulness indices (the increase in the multiple correlation when that cue is added to the regression analysis last) for research, teaching, and service were .37, .09, and .01. As a percentage of the total unique variance accounted for, research accounted for 80%, teaching for 19%, and service for 1%.

The policies of each member of the Merit Committee were similar to the policy of the Merit Committee. All three judges were consistent, with multiple correlations for Judges A, B, and C of .70, .82, and .91. The weights given to each of the cues were similar to the Merit Committee, with each judge giving the most weight to research and the least weight to service. The percentages of total unique variance for research, teaching, and service were 91%, 4%, 4% for Judge A, 70%, 28%, 2% for Judge B, and 79%, 20%, and 1% for Judge C.

Capturing the Merit Committee's Policy Using Objective Data

A second policy capturing analysis was done to determine whether the Merit Committee's judgments could be captured directly from the data they used to make their judgments. Instead of using the Merit Committee's ratings for research, teaching, and service, tallies were made of items listed in the faculty annual report under the categories of research, teaching, and service. For this regression analysis, research was operationally defined as the number of publications and presentations listed in the annual reports; teaching was defined as the mean rating across six items on the student course evaluation form; and service was defined as the number of department, SBA, university, and professional service activities listed in the annual reports. Using this "objective" data, the policy of the Merit Committee was captured, with a multiple correlation of .67, indicating that 45% of the variance in merit pay percentage increases could be accounted for ($F(3, 32) = 8.60$, $p < .01$). The beta weights for research, teaching, and service were similar to the policy capturing based on the Merit Committee's ratings, .52, .26, and .03. The cue intercorrelations were also similar, with research-teaching .36, $p = .09$, research-service .30, $p = .22$, and teaching-service .34, $p = .12$. The usefulness indices also showed a heavy emphasis on research (.22), and little weight given to teaching (.05) and service (.00).

The degree of policy similarity (i.e., similarity in the linear models rather than similarity in actual judgments) between the merit committee's consensus policy and the objective data policy based on annual report and student course evaluation data, can be measured by the correlating the predicted scores from the regression equation for the merit committee's consensus policy with the predicted scores for the objective data policy. The correlation between predicted judgments is corrected for attenuation due to the judge's unreliability and, therefore, is "error-free" and provides an index by which several judges' policies can be compared.[15] The correlation between the consensus and objective policies was .73, $p < .01$, indicating a substantial degree of similarity between the Merit Committee's policy and the objective policy.

The policy capturing results based on the objective data can be improved slightly ($F(3, 32) = 17.29$, $p < .01$) by counting only academic publications (not practitioner publications or books or chapters in books) and published conference proceedings (not conference presentations). This "academic research only" policy had a multiple correlation of .79, indicating that 62% of the variance can be accounted for using this more narrow definition of research. The results of the policy capturing analyses are shown in Table 1.

Table 1. Policy Capturing Results Using Merit Committee Judgments and Objective Data

Policy	R	R^2	F-Ratio	Mean (SD)	Beta Weights	Usefulness Indices
Merit Committee	.90	.81	46.69*			
Research				2.7 (1.2)	.682*	.373
Teaching				3.4 (0.7)	.331*	.088
Service				3.5 (0.9)	.087	.007
Judge A	.70	.49	10.29*			
Research				3.1 (1.5)	.700*	.368
Teaching				3.4 (0.7)	−.180	.018
Service				3.5 (1.2)	.170	.018
Judge B	.82	.68	22.12*			
Research				3.1 (0.9)	.575*	.201
Teaching				3.7 (0.8)	.385*	.081
Service				3.5 (0.7)	.095	.005
Judge C	.91	.82	48.14*			
Research				2.7 (1.2)	.676*	.249
Teaching				3.5 (0.6)	.365*	.064
Service				3.5 (0.9)	.085	.004
Objective Data	.69	.45	8.60*			
Research				2.8 (2.2)	.515*	.221
Teaching				4.2 (0.3)	.259	.052
Service				8.6 (3.6)	.032	.001
Academic Research	.79	.62	17.29*			
Research				0.6 (0.8)	.649*	.393
Teaching				4.3 (0.3)	.300	.016
Service				8.6 (3.6)	.044	.001

Note. * $p < .05$.

Pre-Meeting Agreement Among Merit Committee Members

To determine the level of agreement among the Merit Committee members before the merit raise meeting, intraclass correlation coefficients were calculated on the research, teaching, and service judgments for the three Merit Committee members for which data was available. The level of pre-meeting agreement on research and service was high, with intraclass correlations of .73 for research and .68 for service. The level of agreement for teaching, however, was only .35. To determine if there were significant differences across raters (i.e., disagreement), three Analyses of Variance were calculated using research, teaching, and service as dependent variables. There were no significant differences across raters for research, teaching, or service. The level of pre-meeting agreement with the Merit Committee's consensus judgments was determined by correlating each judge's pre-meeting ratings with the committee's consensus ratings. For research, the correlations ranged from .99 to .78; for service .94 to .67; and for teaching .96 to .14.

Influence in the Consensus Process

The members of the Merit Committee made their judgments on research, teaching, and service prior to the Merit Committee meeting, then during the meeting came to a consensus judgment about each faculty member. The correlation between each committee member's pre-meeting judgments and the committee's consensus judgment is an indication of the degree of influence of each committee member. Judge C appeared to be the "opinion leader" of the Merit Committee, with pre-meeting individual with consensus judgments correlations for research, teaching, and service of .99, .96, and .94. Judge A's correlations were .87, .86, and .66. Judge B appeared to be the most influenced by the merit raise meeting, with correlations of .77, .14, and .79. The pre-meeting judgment correlations, and the means and standard deviations for each judge's ratings are shown in Table 2 (next page).

Table 2. Correlations Between Each Judge's Pre-Meeting Ratings on Research, Teaching, and Service, and the Merit Committee Consensus Ratings, and Means and Standard Deviations of Ratings

	Research	Teaching	Service
Correlation with Consensus			
Judge A	.77	.14	.79
Judge B	.87	.86	.66
Judge C	.99	.96	.94
Means (Standard Deviations)			
Judge A	3.12 (0.9)	3.65 (0.8)	3.47 (0.7)
Judge B	3.07 (1.5)	3.36 (0.7)	3.51 (1.3)
Judge C	2.67 (1.2)	3.47 (0.6)	3.46 (0.9)
Consensus	2.68 (1.3)	3.43 (0.7)	3.54 (0.9)

Note. * $p < .05$.

To test for adverse impact on race (for Asians only, there were too few blacks, Hispanics, and native Americans to calculate adverse impact on these groups) in the raise decisions, three ANOVAs were calculated for each judge's pre-meeting judgments and the Merit Committee's judgments using research, teaching, and service judgments as the dependent variable; a MANOVA was also calculated across all three variables. For this test, there were 7 Asians and 28 whites. There were no significant effects of race for any of the judges or for the Merit Committee, indicating that there was no adverse impact on Asians. Similar results were found for sex discrimination (7 females and 29 males), with one exception. Judge A had a significant overall effect for sex ($F(3, 32) = 2.97$, $p = .047$), but only the test for adverse impact in teaching judgments was significant ($F(1, 34) = 7.04$, $p = .012$), and Judge A gave higher ratings on teaching to women (mean of 3.9, standard deviation 0.9) than men (mean of 3.2, standard deviation 0.6), indicating no adverse impact on women.

Discussion

In this study, a merit raise committee was charged with assigning merit pay increases to a group of employees based on data provided in faculty annual reports and other available data. A Freedom of Information Act request was used to obtain data on the Merit Committee's judgments, to determine how consistently the Merit Committee made their decisions, what weight they gave to the different sources of information, and whether they followed the merit pay directive they were charged with implementing. This study showed that the Merit Committee and individual Merit Committee members were very consistent in using the data they had been provided to make merit pay increase decisions. The Merit Committee was so consistent that the committee could have generated the pay increase percentages by applying a formula to their consensus ratings.

It is clear by the time and effort the faculty put into completing their annual reports and that the Merit Committee puts into evaluating each faculty member's data, that it is assumed that deciding on merit pay increase percentages requires careful thought and judgment. However, the policy capturing analysis based on "objective" data only (number of publications, mean score on six items taken from courses evaluations, and number of service activities) accounted for a substantial amount of the variance in merit pay percentage increases (45% using the objective data, compared to 81% using the Merit Committee's consensus ratings), and mirrored the weights given to the cues (usefulness indices of .37, .09, and .01 compared to .22, .05, and service .00). When only academic research was counted, the variance accounted for was 62%, with usefulness indices for research, teaching, and service of .39, .02, and .00. Furthermore, the members of the Merit Committee showed a high level of agreement among themselves before the meeting on the ratings for research and service, but only a moderate level of agreement for teaching. Except for one judge's ratings on teaching, there also were high levels of agreement between Merit Committee members and the consensus judgments of the Merit Committee. In short, not only did the Merit Committee rely almost exclusively on research to make its merit pay judgments, there was a high level of agreement among the committee members on their ratings for research before the Merit Committee meeting.

Why did the Merit Committee rely almost entirely upon research when making its merit raise decisions? One possible explanation is that faculty raters may have more confidence making evaluations of research than of teaching or service because they can both count and read the publications, and they may have already made judgments about the quality of journals. Faculty raters may have less confidence in their ratings of teaching or service when they must rely primarily upon information provided by others, such as student course evaluations. These sources of information may be considered more subjective in nature.

The difference in variance accounted for between the merit committee's consensus ratings of research, teaching, and service compared to the "objective" data may be evidence of an underspecified model. In this case, the Merit Committee may be taking into consideration the quality of the research (first versus second tier journals, proceedings versus presentations), the level of teaching (large sections of survey courses versus elective courses in the major), and the quality of service performed (tenure committee versus faculty senate), or other factors such as visibility or citizenship. It did not appear that ratee race or sex was a variable that should have been included in the model; there was no evidence of race or sex discrimination in the Merit Committee's consensus judgments or in each individual judge's ratings.

Not only did the Merit Committee tend to rely primarily on only one cue in making its raise decisions, one judge tended to heavily influence the Merit Committee's consensus judgments, with few differences between that judge's judgments and the consensus judgments, and with correlations between that judge's pre-meeting and the committee's consensus judgments for research, teaching, and service all above .94. This result may indicate that one member of the Merit Committee was better prepared

and had better arguments to support his/her initial ratings, or had less of a departmental bias, or induced deference from the other committee members.

This study also illustrates the problems of doing "real-world" research, where the situation is not entirely under the control of the researchers. Much of the policy capturing literature is done under laboratory conditions. A set of students or experts evaluates paper people. The cues used in the evaluation are controlled, good versus poor performance may be quite clear. This type of study has a high degree of experimental control; however, this research may not capture the complexities of an actual decision situation. The current study involves evaluators charged with a merit pay decision. The cues used in the study are not under "tight" control. For example, the faculty publish in a wide range of outlets. Deciding what is good versus poor performance may not be clear and may be subject to interpretation. Brunswick calls this type of study an "representative" design[10]. That is, decision-makers make decisions in the same way they ordinarily would. Hence, this research contributes to our understanding of the process.

According to a survey of professors in business schools, merit pay systems generate the traditional types of dissatisfaction with merit pay plans, including disagreements about the weights to be attached to the evaluation criteria, the procedures for implementing the plan, and a lack of balance among the criteria of research, teaching, and service.[7] These attitudes are not unlike those in other public and private sector organizations. This study illustrates the use of policy capturing to determine whether the merit pay policy implemented by the Merit Committee was, in fact, technically in accordance with the wishes of the faculty. The faculty merit raise directive specifies that each faculty member's contributions in research, teaching, and service should be considered, with "special consideration" for research. The weight the Merit Committee actually gave to research, teaching, and service was about 80%, 19%, and 1%. Although this complies with the vague directive to consider all three areas in making merit pay decisions, and to give the greatest weight to research, the merit pay increase judgments were dominated by research. Relative weights such as 34%, 33%, 33%, or even 98%, 1%, 1% would also comply with the directive. Absent the FOIA request and the policy capturing analysis, however, the faculty never knew how their directive was actually being implemented.

Based on the previously mentioned survey results[7], the faculty may have thought the relative weights used by the Merit Committee were approximately 50%, 30%, and 20%. Now that the faculty have seen the Merit Committee's actual judgment policy, it remains to be seen how they will react to the information. The implications of letting the current merit pay directive remain unchanged are that faculty will put most of their time into research, and minimize teaching and service opportunities. Lawler (1990), discusses "influenceability" of pay systems or an individual "being able to affect his or her performance measure through behavior (p. 14)." The short-term strategy for maximizing salary in this unit based on the cue weights would be to put one's effort into research and reduce the amount of investment in service and teaching. This would be counter to the organization's stated objectives for faculty to provide teaching, research and service. Expecting faculty to continue to take on service assignments

and invest time and effort to improve their teaching would be an example of the "folly of rewarding A while hoping for B."[16] However, in the long run, such a strategy could work against a faculty member in a tenure review. Tenure and promotion criteria at this university specify acceptable standards in teaching and service as well as research. In addition, it may be that a vaguely worded merit raise directive implemented as it was in this case could work counter to a more balanced and "business" objective approach. If faculty put all of their efforts into research, the university would have difficulty providing community service or excellent teaching, two stated business objectives. The reputation of the organization, while possibly increasing in the research environment, would diminish in other sectors of the environment. In the long run, this could result in fewer donations as well as a decrease in the number of students applying for admission.

In this particular situation, the faculty must decide whether they should give more specific instructions on the relative weights, or continue to allow the Merit Committee to exercise its own judgment about how much weight to give to research, teaching, and service. In other organizations, evaluators must be aware of cue weightings and should review the links between cues and organizational objectives prior to any evaluation.

Notes

1 Lawler, E. E. III (1990). *Strategic pay*. San Francisco, CA: Jossey-Bass.

2 Bassett, G. (1994). Merit pay increases are a mistake. *Compensation and Benefits Review, 26* (2), 20-22.

3 Eskew, D., and Heneman, R. L. (1996). A survey of merit pay plan effectiveness: End of the line for merit pay or hope for improvement? *Human Resource Planning, 19*, 12-19.

4 Dessler, G. (2000). *Human resource management* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.

5 Bergman, T. J., and Scarpello, V. G. (2001). *Compensation decision making*. (4th ed.). Fort Worth, TX: Harcourt College Publishers.

6 Gomez-Mejia, L. R., and Baekin, D. B. (1992). Determinants of faculty pay: An agency theory perspective. *Academy of Management Journal, 35*, 921-955.

7 Prewitt, L. B., Phillips, J. D., and Yasin, K. (1991). Merit pay in Academia: Perceptions from the School of Business. *Public Personnel Management, 20* (4), 409-417.

8 Phelps, R. H, and Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance, 21*, 209-219.

9 Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582.

10 Brunswik, E. (1955). *The conceptual framework of psychology*. Chicago, IL: University of Chicago Press.

11 Mazen, A. M. (1990). The moderating role of social desirability, age, and experience in human judgment: Just how indirect is policy capturing? *Organizational Behavior and Human Decision Processes, 45* (1), 19-40.

12 Martocchio, J. J., and Judge, T. A. (1994). A policy-capturing approach to individual's decisions to be absent. *Organizational Behavior and Human Decision Processes, 57*, 358-386.

[13] Brant, C. R., Mynatt, C. R., and Doherty, M. E. (1999). Judgments about sexism: A policy capturing approach. *Sex Roles, 41* (5-6), 347-374.

[14] Zacharakis, A. L., Meyer, G. D. (2000). The potential of actuarial decision models: Can they improve the venture capital decision? *Journal of Business Venturing, 15(4)*, 323-346.

[15] Naylor, J. C., and Schenck, E. A. (1966). $_m$ as an "error-free" index of rater agreement. *Educational and Psychological Measurement, 26*, 815-824.

[16] Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal, 18*, 769-783.

Authors

Kenneth M. York
School of Business Administration, Oakland University
(248) 370-3272
york@oakland.edu

Kenneth M. York received his Ph.D. in Industrial/Organizational Psychology from Bowling Green State University and is an Associate Professor of Management in the School of Business Administration, Oakland University. He specializes in the application of behavioral decision theory, and the creation of experiential learning exercises for development of management skills.

Lizabeth A. Barclay
School of Business Administration, Oakland University
(248) 370-3293
barclay@oakland.edu

Lizabeth A. Barclay, Professor of Management, received her Ph.D. from Wayne State University. She has been at Oakland University since 1980. Her interests include international OB/HRM, teaching effectiveness, self-efficacy, and union-management relations.

Author Note

The authors would like to thank Alan R. Bass and Terry Beehr for their helpful comments.