

24575171



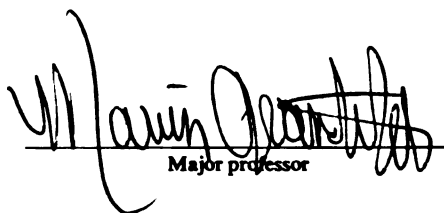
This is to certify that the
dissertation entitled
The Relationship of Sex of Candidate
and Prestige of Institution
To Faculty Performance Evaluation
At the University Committee Level

presented by

Elizabeth A. Hansen

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Education


Major professor

Date February 1, 1990

**PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.**

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

MSU Is An Affirmative Action/Equal Opportunity Institution

**THE RELATIONSHIP OF SEX OF CANDIDATE
AND PRESTIGE OF INSTITUTION
TO FACULTY PERFORMANCE EVALUATION
AT THE UNIVERSITY COMMITTEE LEVEL**

By

Elizabeth A. Hansen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Educational Administration

1989

6054341

ABSTRACT

THE RELATIONSHIP OF SEX OF CANDIDATE AND PRESTIGE OF INSTITUTION TO FACULTY PERFORMANCE EVALUATION AT THE UNIVERSITY COMMITTEE LEVEL

By

Elizabeth A. Hansen

This research provided a means for administrators to compare faculty decision-making against a mathematical model of the perceived rating system. Faculty subjects rated hypothetical applicants for tenure, their ratings were compared to the computer model's decisions. Variables were sex of hypothetical applicant and prestige of the candidate's institution.

The subjects were tenure-track faculty members from Central Michigan University. Using an ANOVA, the results were that male candidates were rated higher than female candidates when the candidates were strong in research, female candidates were rated higher than male candidates when the candidates were strong in teaching; overall, male candidates were rated higher than female candidates, male candidates from high prestige schools were rated higher than male candidates from low prestige schools, and female candidates from low prestige schools were rated higher than female candidates from high prestige schools.

Copyright by

ELIZABETH ANN HANSEN

1989

ACKNOWLEDGMENTS

I wish to thank my advisor, Dr. Marvin E. Grandstaff and the members of my committee, Dr. Samuel A. Moore, Dr. Kenneth L. Neff, and Dr. Eldon R. Nonnamaker for their support.

Special thanks to the faculty members of Central Michigan University who participated in the study. Dr. William Lewis was particularly helpful. I thank him.

Also, I wish to thank my husband, my sons, and my parents on the home front.

TABLE OF CONTENTS

	Page
Acknowledgments.....	iii
Abstract.....	iv
List of Tables.....	vii
Chapter 1 Introduction	
Statement of the Problem.....	1
Purpose of the Study.....	1
Significance of the Study.....	1
Hypotheses.....	2
Setting of the Study.....	3
Scope of the Study.....	3
Chapter 2 Review of the Literature	
Problems with Evaluation.....	4
Two Types of Bias.....	6
Criteria for Evaluation: Teaching, Research and Service.....	8
The University Committee.....	10
Summary.....	11
Chapter 3 Design of the Study	
Introduction.....	12
Model for Faculty Evaluation.....	12
Rationale for Model.....	12
Definitions.....	14
Development of Model.....	15
Selection of Subjects.....	26
Procedure.....	27
Analysis.....	27

	Page
Chapter 4 Presentation of Results	
Introductory Explanation.....	30
Overall Analysis of Data.....	32
Summary of Results.....	44
Chapter 5 Conclusions and Recommendations	
Conclusions.....	46
Recommendations.....	46
Speculations and Recommendations for Further Research.....	47
Appendices	
A. Biographical Material on Hypothetical Candidates for Tenure.....	49
B. Criteria for Tenure.....	58
C. Instructions.....	59
D. The Computer Program.....	61
Bibliography.....	77

LIST OF TABLES

Table	Page
1. Latin Square Design of Study.....	29
2. Data Presented to Subjects.....	31
3. Sum of Scores for Each Hypothetical Candidate.....	32
4. ANOVA: Overall.....	33
5. Sum of Scores for Candidates in High Teaching Category.....	34
6. ANOVA: Sex With High Teaching Category.....	35
7. Sum of Scores for Candidates in High Research Category.....	36
8. ANOVA: Sex With High Research Category.....	37
9. Sum of Scores for Candidates in Low Research Category.....	38
10. ANOVA: Sex With Low Research Category.....	39
11. Sum of Scores for Candidates in Low Teaching Category.....	40
12. ANOVA: Sex With Low Teaching Category.....	41
13. Sum of Scores for Male and Female Hypothetical Candidates.....	42
14. ANOVA: Sex.....	42
15. Sum of Scores of High and Low Prestige Candidates.....	43
16. ANOVA: Prestige.....	44

CHAPTER 1

INTRODUCTION

Statement of the Problem

The problem investigated was whether bias, either based on sex of a candidate or prestige of the candidate's institution, would affect a performance evaluation of that candidate by a university-wide personnel committee.

Purpose of the Study

Results from this research will help administrators determine whether there may be unasked for variables introduced into the decision-making process when members of a tenure and promotion committee are given distinct and measurable performance objectives by which to evaluate a group of faculty. Also, a computerized model for decision-making is offered as a means for comparing committee members' decisions with an unbiased rating.

Significance of the Study

The significance of this research is that it provides a means for administrators to compare faculty decision-making against a mathematical model of the perceived rating system. If the faculty decisions and the decisions produced by the model do not match, then there may be unnamed variables that were not added to the model. These variables may be ones that were not in the model by neglect, in which case the

administrator could easily adjust the computer program to consider them. These variables may be ones that were not in the model because they should not be included in the decision process (such as race or sex) whereby the administrator would have to reconsider the committee decision.

Hypotheses

The following hypotheses were tested. The general hypothesis was: when given information to evaluate faculty for a tenure decision and the faculty are from a field outside one's field of expertise, the members of a university tenure committee will render decisions that are biased in respect to stated measurable criteria for tenure.

The general hypothesis was tested by ascertaining if sex might influence the decisions. The Null Hypothesis was that the sex of an applicant does not significantly affect tenure decisions.

The general hypothesis was tested by ascertaining if perceived prestige of the hypothetical candidate might influence the decisions (prestige was determined by where the hypothetical candidates earned their terminal degrees). The Null Hypothesis was that the prestige of the hypothetical candidate's background does not significantly affect the tenure decisions.

Setting of the Study

The subjects were tenure-track faculty members from Central Michigan University, a large predominantly undergraduate institution. There were eight faculty members from the computer science department, eight faculty members from the education department, eight faculty members from the mathematics department, and eight faculty members from the English department.

Scope of the Study

The subjects were limited to one institution of higher learning to provide a more homogeneous group so as to limit possible effects of extraneous variables. The subjects were limited to tenure-track faculty for the same reason. The departments were chosen for two reasons: on the basis of size; they were large enough to contain at least eight subjects willing to participate; and on the basis of subject matter; mathematics and computer science are more quantitative than history and English. These four departments were chosen because they were not at all related to the hypothetical textile department; this was to simulate a university committee making decisions about a faculty member whose department was not represented on the committee.

CHAPTER 2

REVIEW OF THE LITERATURE

Problems with Evaluation

The issue of evaluation of faculty is a matter that has been commented upon since the beginnings of the university system itself. The roots and rituals of the university go back to the Middle Ages.

"Its hierarchical arrangements are simple and standardized, but the academic hierarchy includes a greater range of skills and a greater diversity of tasks than any business or military organization...It is not easy...to determine the fundamental purposes of a university or the relative importance of different activities in contributing to those purposes."¹

Even though it may be hard to determine the factors and the respective proportions that should be included in the decision-making process, this should be attempted anyway since the decision will be made in any case.

The problem of evaluation of faculty is compounded by problems with communication. As in any organization, there are ways communication can be distorted or even prevented. This can be said to be an information screen: "...an information screen may be defined as a set of social practices, beliefs, and behaviors within an

¹Theodore Caplow and Reece J. McGee, The Academic Marketplace (New York: Arno Press, 1977), p. 4.

organized group which inhibits the communication of certain kinds of information between certain positions or in certain directions..."²

There are many types of information screens. Caplow and McGee identify one information screen as being "erected by the university's administrative officials to shield from the working members of their departments the criteria by which men are officially evaluated. Our data abound in complaints from professors that they were not told exactly either why a given colleague was hired or fired or what he had or did not have that someone else had or did not have. Every university has its legends about certain firings or about recommended promotions that were never made." ³

This problem of communication creates a confusing situation for the faculty member. "In view of the vague and conflicting criteria by which his work is judged, he is uncertain in the allocation of his energies. He knows that he is a competitor, but often is not clear regarding the terms of the competition."⁴ Also, because of the lack of communication, the people doing the evaluation might use procedures that are not standardized or not appropriate. Bias can thus be introduced into the decision-making process.

²Ibid., p. 59.

³Ibid., p. 60-61.

⁴Wilson Logan, The Academic Man: Sociology of a Profession (London: Oxford University Press, 1942), p. 62.

Two Types of Bias

Two types of bias were investigated in this research. One type of bias is based on perceived prestige of the faculty member.

The higher the rank of the department in the disciplinary prestige system, the more it serves its individual members by conferring a derivative reputation on them. This reputation tends to make them more desirable to other universities, more independent of their own, and more inclined to mobility....the higher the prestige of a department, the greater will be the tendency for its members to be oriented to the discipline rather than to the university....in the high-prestige institutions men are hired on an estimate of how much research they are likely to do. When their tenure is decided, direct utility to the university hardly enters as a factor in the decision to keep them. The measurement of their worth is haunted by quite another problem- their usefulness in future staff procurement.⁵

Prestige is an important factor in faculty career patterns and can influence decision-making in terms of new faculty members. Where a faculty member received his/her degree can be more of an influential factor than other factors such as achievement, talent or evaluation results.⁶ This factor of prestige is thought to be a dominant consideration in professorial concerns. "Professors wish to be number one--if not for themselves then for their department; if not in their own area of investigation, then as teachers; and certainly for their institution as a whole and its ranking with other colleges and universities."⁷ Research has shown that the norms and values critical

⁵Ibid., p. 107.

⁶David W. Breneman and Ted I. K. Youn (eds.), Academic Labor Markets and Careers (Philadelphia: Falmer Press, 1988).

⁷Robert Blackburn, "The Meaning of Work in Academia," New Directions for Institutional Research, (San Francisco: Jossey-Bass, 1974) I, p. 80.

to high performance are cultivated in graduate programs, therefore doctoral prestige can be used as a predictor of research performance.⁸

Another type of bias that is investigated is sex-bias. Communications expert Patricia King comments, "Some people rate minorities and women lower than others; some expect so little of them that anything they accomplish seems like a miracle and gets very high marks; others bend over backward to give them a break and rate them higher than they deserve." ⁹ There are many reports of court cases involving sex-bias in tenure decisions. One of the problems in doing a study investigating bias is that the confidential files must be opened. The Supreme Court is having a difficult time getting access to confidential decision-making information.¹⁰ Research which can reveal sex-bias by using dummy files and comparing decision-making results of a faculty committee against a computer program is non-invasive to confidential files.

Another problem in determining sex-bias is that men produce more research than women but comparisons between them on just a count basis are inappropriate because men are more likely to have begun their careers earlier and have them interrupted less frequently, to have lighter teaching loads, to be employed full-time and to be paid more.¹¹

⁸David Dill, "Research as a Scholarly Activity: Context and Culture," New Directions for Institutional Research, (San Francisco: Jossey-Bass, 1986), XIII, p. 13.

⁹Patricia King, Performance Planning and appraisal, (New York: McGraw-Hill Book Company, 1984), p. 57.

¹⁰"In Weighing Sex-bias Case, High Court Will Skirt Issue of Confidentiality of Tenure-Review Records," The Chronicle of Higher Education, January 4, 1989, p. A13.

¹¹Blackburn, loc. cit., p. 87.

Criteria for Evaluation: Teaching, Research and Service

The criteria for the evaluation of faculty must reflect the mission of the university or college. Prestige and recognition are as important as the security and financial rewards that are associated with earned promotion and tenure. Many institutions consider the triad of teaching, research and service as the focus for the evaluation of faculty. Therefore, rewards in higher education tend to be related to these three focal points, with emphasis on one or another changing depending upon the particular institution or the particular period in history.

Teaching is a paramount responsibility of a college or university. During the expansion of higher education in the 1960s, most departments were concerned with recruiting and retaining faculty. Now, many departments are not expanding, so the departments are often forced to make fine distinctions between faculty members. In making these fine distinctions, many administrators are emphasizing measurable evidence for evaluating teaching, such as systematic student ratings and the content of course syllabi and examinations.¹²

Research has been increasingly emphasized in higher education. From the Morrill and Hatch Acts of the nineteenth century to the investments of private corporations in the development of new technology, it is evident that external influences are changing the face of institutional reward systems. "Campus after campus has been moving aggressively to upgrade the importance of scholarly productivity as a criterion for academic personnel decisions..."¹³

¹²John A. Centra, "Using Student Assessments to Improve Performance and Vitality," New Directions for Institutional Research, (San Francisco Jossey-Bass, 1978), p. 40.

¹³ J. N. Schuster and H. R. Bowen, "The Faculty at Risk," Change, 1985, 17, pp. 15-16.

Measuring research performance has been problematic. Typically, research performance has been measured by counting the number of publications. In a comment by a professor: "It's very simple. We look at all the publications. Then the committee gets together and we have a gut reaction."¹⁴ Quantitative measures are popular, such as number of journal articles (refereed and nonrefereed), books, unpublished research, citations, and grants. The use of these quantitative, written products of research to assess faculty performance "fits well with the assertion of Etzioni and Lehman (1967)¹⁵ that assessments tend to be based on the attributes that are the easiest to measure."¹⁶

Service, the third tier of faculty performance, tends to be poorly conceptualized and inconsistently expressed in higher education. Service can be thought of as campus committee assignments, membership in a local church, consulting with local businesses and serving on professional associations. The idea that the university should provide public and community service came from the idea of the commons concept in land use. "It seems to be the logical development from the land-grant idea of institutional commitment to services that reach beyond students who are enrolled in degree programs on campus. This service mission emerged out of a conviction that knowledge is useful beyond the classroom and that the university had a responsibility to the society to extend the benefits of learning to the larger public."¹⁷

¹⁴Peter Seldin, Successful Faculty Evaluation Programs (Crugers, N.Y.:Coventry, 1980) p. 34.

¹⁵A. Etzioni and E. W. Lehman, "Some Dangers in 'Valid' Social Measurements," The Annals, 1967, 373, pp. 1-15.

¹⁶J. M. Braxton and A. E. Brayer, "Assessing Faculty Scholarly Performance," New Directions for Institutional Research, (San Francisco: Jossey-Bass, 1986), p. 28.

¹⁷Durward Long, "The University as Commons: A View from Administration," New Directions for Institutional Research (San Francisco: Jossey-Bass, 1977), V, p. 75.

The University Committee

In a study conducted by Peter Seldin, it was found that the policies and practices used to evaluate faculty performance are "becoming more structured and systemized. Chairman and dean evaluations, while still very important, are losing ground to formal faculty committees, self-evaluation, and colleagues' opinions...The trend to decentralization and the sharing of decision-making seems clear, as does the growing effort toward the reliability of the evaluative process."¹⁸ This trend would seem to indicate that the university promotion and tenure committee might exert more influence in the future. The university-wide committee represents university perspectives and standards and insures appropriate consideration of the long term academic priorities of the university and the fiscal situation of the university.¹⁹ This implies that the denial of tenure or promotion could be on the basis of fiscal priorities or academic planning rather than the qualities of the individual.

Research which facilitates the work of the committee by increasing the reliability of the evaluations from the committee can be very useful. The standardization of some of the decision-making should help lessen the confusion in the minds of faculty who are being evaluated and should increase the reliability of the committee decision. All of this in turn can decrease the time and cost involved with the evaluation process.

¹⁸Peter Seldin, Successful Faculty Evaluation Programs, (Crugers, New York: Coventry Press, 1981), p. 34.

¹⁹Donald K. Smith, "Faculty Vitality and the Management of University Personnel Policies," New Directions for Institutional Research, (San Francisco: Jossey-Bass Inc., 1978), V, p. 9.

University committees tend to look at folders which contain written information about the faculty member under discussion. Biographical data have historically been a tool in the selection process. One of the best predictors of what individual's are capable of is the record of what they have done in the past.²⁰

Summary

The issue of evaluation of a faculty member is problematic. The evaluation can be biased and error prone but since a decision has to be made, it is important to understand where bias can appear. The literature supports the trend in higher education to quantify decision-making about faculty evaluation. This research points out where bias can appear and offers a computerized model which can highlight bias and help an administrator toward more informed decision-making.

²⁰Michael Nash, Making People Productive, (San Francisco: Jossey-Bass, 1985) p. 43.

CHAPTER 3

DESIGN OF THE STUDY

Introduction

In order to compare faculty responses to an unbiased response, a mechanism to obtain an unbiased response was put into place. Since calculations based on the tenure criteria can be very tedious to perform, a computer program which allows users to implement the model was used in this study. (See Appendix D). The computer program allows the user to do three different things. The first is to devise an evaluation criteria, that is, choose high-level characteristics such as teaching, research and service and choose corresponding primitive characteristics, such as teaching evaluation scores, number of journal articles, and how many times the candidate for tenure lectured in the local community. The second is to weight any evaluation criteria, such as giving 60% weight to research. The third is to evaluate an individual. Thus, if a faculty member evaluates a hypothetical candidate for tenure on specific criteria and gets a different ranking than the computer program produces, then the faculty member allowed an unasked for criteria to influence the decision-making. The study was designed so that the only unasked for differences among the hypothetical candidates were sex of candidate and the candidates's Ph.D. granting institution.

Model for Faculty Evaluation

The model allows you to choose which broad areas you use in evaluating faculty and in these broad areas allows you to specify sub-areas. Everything can be weighted and a single number will result so that comparisons can be made across faculty. The weights can be modified so that administrators can predict the results of a policy change on the number of faculty attaining promotion or tenure.

Rationale for Model

One takes measurements of system components to see if the components are performing as the system requires. We may view college professors as a component in a very complex system called a university. Often people argue that it is not fair to take measurements of a subset of a college professor's performance and then base a promotion or tenure decision on these measurements. They usually base their arguments on the fact that any model of performance will be too simple to capture the complexity of the job. What these people forget is that even if measurements are not taken, a decision will be made as to how well a college professor performs. By forcing a university to come up with a mathematical model for faculty evaluation, two things might happen. Sometimes the very act of building the model and seeing how the model evaluates college professors will cause the institution to admit that it has been saying something it did not mean to faculty, such as "we value teaching above all" while results from the model consistently show it is research that gives an instructor a higher rating. Having the evaluation model also allows faculty to evaluate themselves by the model at times other than formal university evaluation periods and make any mid-career adjustments that are necessary.

Definitions

The word "system" is used in many contexts: there are educational systems, biological systems and political systems. A system consists of interrelated subsystems that work together to convert inputs to outputs.

...we can define a system as a relation between inputs and outputs. No matter how simple or complex the system, its successful use depends on an understanding of its structure. This is the main task of system analysis. The job of the analyst is to study how an organized whole can process the inputs. The analyst wants to know the character of the system in order to forecast its future. Destiny means evolution over time, and this evolution is governed by the character of the system--that is, by its structure...we can model the structure by fixing our attention on the state of the system. The state is defined as the minimum information required to describe the system's condition in such a way, if the inputs are known, then the condition at any time is completely determined.²¹

It will be difficult to achieve the goal of improvement of quality unless we can define and measure the components of quality.

In a restricted sense, quality is often considered synonymous with reliability...If we can define and measure these characteristics of quality with some degree of precision, then managers and customers can use such quality metrics either to set goals to be achieved by a software product, or as the basis for rejection or acceptance of a completed project...each high-level characteristic is decomposed into primitive characteristics...If a metric can be defined for each primitive, then measurements of these primitives can be combined to produce a single metric for reliability...we would like to define metrics for each primitive and combine these metrics in some way to produce a single figure of merit for the overall product.²²

²¹Constantin V. Negoita and Dan Ralescu, Simulation, Knowledge-Based Computing, and Fuzzy Statistics, (New York: Von Nostrand Reinhold Co., 1987), p. 1.

²² S.D. Conte, H.E. Dunsmore, and V.Y. Shen, Software Engineering Metrics and Models, (Menlo Park, California: Benjamin/Cummings Publishing Co., 1986), pp. 7-8.

One way of quantifying performance of a college professor is to specify it as high-level characteristics such as:

- * teaching
- * research
- * service

The problem is to measure these with some degree of precision. It seems difficult to measure the above characteristics with a single metric.

One way of dealing with this would be to break the high level characteristics into primitive characteristics. For example, service, a high level characteristic, might be broken down into:

- * departmental committee service
- * college committee service
- * university committee service
- * service to professional groups
- * service to student groups
- * service to community

Development of Model

Mathematically, we might describe this as follows:

Consider n high level characteristics denoted $H_i (i=1, \dots, n)$. We will assume that each high level characteristic is made up of m primitive characteristics (M_i). We will use S_{ij} to denote the j th primitive characteristic of the i th high level characteristic.

The score SC_i for the i th high level characteristic is

$$SC_i = \frac{M_i}{\sum_{j=1} S_{ij}}$$

The total score would be

$$TSC = n \sum_{i=1} \frac{M_i}{\sum_{j=1} S_{ij}}$$

The above formula suffers from a number of drawbacks. The first and foremost drawback is that if one high level characteristic has more primitive characteristics than another, the formula will weight it more heavily. One way to do away with this is to divide each SC_i by M_i . Recall that M_i was the number of primitive characteristics associated with each H_i .

This gives us an average score for each of the H_i .

$$TASC = \sum_{i=1}^n \frac{1}{M_i} \sum_{j=1} S_{ij}$$

Suppose we wanted to weight the various high level characteristics. We could associate a weight W_i with each high level characteristic such that

$\sum W_i = n$. (Note there are n high level characteristics.)

$$TWASC = \sum_{i=1}^n \frac{W_i}{M_i} \sum_{j=1} S_{ij}$$

For example, let us assume the high level characteristics are teaching, research, and service. If we want to weight them we will choose three weights whose sum is 3. If

all were to be weighted equally, then $W_1 = 1$ (teaching), $W_2 = 1$ (research), and $W_3 = 1$ (service). Suppose we wanted to weight teaching so that it was twice as important as research and weight research so that it was twice as important as service. This causes the following equations.

$$W_1 = 2 * W_2$$

$$W_2 = 2 * W_3$$

$$W_1 + W_2 + W_3 = 3 \quad \text{the initial constraint}$$

Now with a little algebra we can calculate the weights:

$$W_1 + W_1/2 + W_1/2 = 3$$

$$W_1 + W_1/2 + W_1/4 = 3$$

$$(1 + 1/2 + 1/4)W_1 = 3$$

$$W_1 = 3/1.75 = 12/7$$

$$W_2 = 12/14 = 6/7$$

$$W_3 = 12/28 = 3/7$$

Hence the weights $12/7$, $6/7$, and $3/7$ give the appropriate weighting for the high level characteristics.

We may wish to weight the low level characteristics associated with a given high level characteristic. For example, we might wish for departmental committee service to count less than service to professional groups. This is done in a manner similar to how

we weighted the high level characteristics only now for each primitive characteristic of the high level characteristic (i) choose weights L_{ij} such that

$$\sum_{j=1} M_i L_{ij} = M_i .$$

Using these weights instead of evenly weighting as the above formula implies yields the new formula:

$$TWWSC = \sum_{i=1}^n \frac{W_i}{M_i} \sum_{j=1} M_i S_{ij} * L_{ij}$$

To make it easier to compare differences, we would want to normalize a score. A difference of .1 between schools with different characteristics would mean the same. We can compute a normalized score by taking TWASC and dividing it by the maximum possible score, $TWASC_{max}$.

$$NWASC = \frac{TWASC}{TWASC_{max}}$$

In like manner we could also normalize TWWSC as follows:

$$NWWSC = \frac{TWWSC}{TWWSC_{max}}$$

The normalized score has the property $0 \leq NWASC \leq 1$. The closer this score is to 1, the higher the desirability of the instructor is under our weighting system. If the total weighted score is divided by the highest possible weighted score, the best someone could do is 1.

So far we have not discussed the actual values the score for each primitive attribute may take on. Although we could use the weights to compensate for different

maximum scores for different primitive attributes we will assume that they are all on the same scale. This makes the choice of weights more meaningful. We assume each range is from 0 to 5.

If the instrument that measures a primitive attribute gives values on a different scale, we will use a simple linear transformation to scale it to 0 to 5.

EXAMPLE

Suppose our model has three high level characteristics with the indicated primitive characteristics.

H₁ Teaching

- * student evaluation
- * peer evaluation
- * chairman's evaluation

H₂: Research

- * grants and contracts
- * refereed papers
- * unrefereed papers and presentations

H₃: Service

- * departmental committees
- * college committee service
- * university committee service
- * service to professional groups
- * service to student groups
- * service to community

In teaching, suppose we choose to weight student evaluations 3 times as heavily as peer evaluations and to give peer evaluations equal weight with chairman's evaluation.

$$L_{11} = 3 * L_{12}$$

$$L_{12} = L_{13}$$

$$L_{11} + L_{12} + L_{13} = 3 \quad \text{the initial constraint}$$

Now we can calculate the weights for the primitive characteristics of teaching.

$$L_{11} + (L_{11})/3 + (L_{11})/3 = 3$$

$$(5*L_{11})/3 = 3$$

$$L_{11} = 9/5$$

$$L_{12} = (L_{11})/3 = 3/5$$

$$L_{13} = L_{12} = 3/5$$

In research, suppose we choose to weight refereed papers twice as heavily as grants and contracts. Suppose we also decide to weight grants and contracts three times as heavily as unrefereed papers and presentations.

$$L_{22} = 2 * L_{21}$$

$$L_{22} = 3 * L_{23}$$

$$L_{21} + L_{22} + L_{23} = 3 \quad \text{the initial constraint}$$

Now we can calculate the weights for the primitive characteristics of research.

$$L_{21} + 2 * L_{21} + (L_{21})/6 = 3$$

$$11/3 * L_{21} = 3$$

$$L_{21} = 9/11$$

$$L_{22} = 18/11$$

$$L_{23} = 6/11$$

In service, suppose we choose to weight all six primitive characteristics evenly. This gives us the following:

$$L_{31} = 1, L_{32} = 1, L_{33} = 1, L_{34} = 1, L_{35} = 1, L_{36} = 1 .$$

Now suppose we want to weight research (H_2) twice as much as teaching (H_1).

Suppose we also want to count research three times as much as service.

$$W_2 = 2 * W_1$$

$$W_2 = 3 * W_3$$

$$W_1 + W_2 + W_3 = 3 \quad \text{the initial constraint}$$

Now we can calculate the weights for the high level characteristics.

$$(W_2)/2 + W_2 + (W_2)/3 = 3$$

$$W_2 = 18/11$$

$$W_1 = 9/11$$

$$W_3 = 6/11$$

The model is now completely specified and all that remains is to collect data for each faculty member in each of the primitive characteristics and scale it to the 0 to 5 scale which we have assumed.

Professor A

Teaching

Student evaluation	4.3
Peer evaluation	5
Chairman's evaluation	4

Research

Grants and contracts	1
Refereed papers	4
Unrefereed papers and presentations	2

Service

Departmental	4
College	3
University	3
Professional groups	0
Student groups	4
Community	0

Recall that:

$$\begin{aligned}
 \text{TWWSC} &= \sum_{i=1}^n \frac{W_i}{M_i} \sum_{j=1}^{M_i} S_{ij} * L_{ij} \\
 &= \frac{9(11)}{3} * \left(4.3 \left(\frac{9}{5} \right) + 5 \left(\frac{3}{5} \right) + 4 \left(\frac{3}{5} \right) \right) \\
 &\quad + \\
 &\quad \frac{18(11)}{3} * \left(1 \left(\frac{9}{11} \right) + 4 \left(\frac{18}{11} \right) + 2 \left(\frac{6}{11} \right) \right) \\
 &\quad + \\
 &\quad \frac{6(11)}{6} * (4*1 + 3*1 + 0*1 + 4*1 + 0*1) \\
 &= 9.4679
 \end{aligned}$$

To find NWWSC we need to calculate $TWWSC_{\max}$.

$$TWWSC_{\max} = 15$$

$$NWWSC = 9.4679/15 = .6312$$

Professor A's strongest point is teaching. Now let us consider Professor B whose strongest point is research and who is a poor teacher.

25

Professor B

Teaching

Student evaluation	1
Peer evaluation	2.5
Chairman's evaluation	3

Research

Grants and contracts	4.3
Refereed papers	5
Unrefereed papers and presentations	4

Service

Departmental	4
College	3
University	3
Professional groups	4
Student groups	0
Community	0

Let us see how Professor B does.

$$\text{TWWSC} = 10.2357$$

$$\text{NWWSC} = .6824$$

It would appear that at a school which rated research twice as much as teaching, one could do a pretty poor job of teaching and still come out with a good overall score. This might argue for choosing weights that are relatively close to each other unless you really do not care about a particular component.

This model does not allow for unasked for variables such as sex or age of the candidate to be considered. If any unasked for variables are influencing a decision, then the model's rating score would be different from the score someone obtained from another rating method. University administrators can see empirically the consequences of policy or criteria and hopefully be able to modify the policy or criteria to correspond with reality or defend the criteria and add it to the high-level or primitive characteristics.

Selection of Subjects

Subjects were tenure-track faculty from English, mathematics, education, and computer science at Central Michigan University. They were randomly selected from the campus telephone directory and asked if they would participate. If they were willing they were given an instruction sheet. (See Appendix C).

Procedure

Subjects (32 faculty members, eight from each of the four departments) were given a folder on each of four hypothetical candidates. (See Appendix A). These folders contained enough detail to evaluate the candidates according to the given criteria for tenure. The criteria for tenure were described in a separate handout. (See Appendix B). The subjects were then asked to evaluate the applicants giving them a number between one and ten. (See Appendix C). Overall, the folders were designed so that according to the criteria the men and women were equal. (Each female applicant was paired with a male whom the evaluation program (See Appendix D) gave an equal score but not every evaluator received equal pairs.) The statistical test used was an analysis of variance with a Latin square design. Because of the design of the folders, the average for the males (as calculated by the computer model, see Appendix D) was equal to the average for the females and the average for the high prestige pairs was equal to the average for the low prestige pairs.

Analysis

The experiment tested to see if two factors which should not enter into a tenure decision (sex of candidate and prestige of Ph.D. granting institution) do in fact enter into such a decision. In order to do this subjects were asked to look at folders for candidates and evaluate them. For this to be a success there were a number of clearly distinct outcomes. The four folders were designed so that the subjects would produce clearly dissimilar evaluations provided that the subjects followed the stated criteria. The subjects were people who are tenure-track faculty at Central Michigan University and the experiment did not take up an inordinate amount of time. It was assumed that if the subjects were asked to do a time consuming task then the level of participation would perhaps drop to a point where a statistical analysis would be

impossible. Also, according to Caplow and McGee, it is debatable whether evaluators actually spend more than a few minutes in decision-making.²³ The four folders together with cover sheets indicating sex and prestige of institution gave sixteen combinations. It was assumed that if a factorial experiment were done, two problems would arise. The experiment would take too long. At five minutes per folder each subject would spend 80 minutes evaluating the folders. More importantly, there would be a need to generate four different versions for each of the rating levels. While this could be done the chance of biasing the experiment by something as simple as choice of titles of publications would be great.

The experimental design chosen was the Latin square design. It is a special design that permits the researcher to assess the relative effects of treatments when a double type of blocking restriction is imposed on the experimental units. This research involved testing of the effect of sex and prestige of institution on evaluation score. Thus the set up uses the levels as columns and sex-prestige pairs as rows. Using such a design four subjects can test all possibilities with each person getting one folder of each level. The following is a classic Latin square design from Snedecor and Cochran.²⁴

²³Theodore Caplow and Reece J. McGee, The Academic Marketplace (New York: Basic Books, Inc., 1958), p. 127.

²⁴George W. Snedecor and William G. Cochran, Statistical Methods (Ames,Iowa: The Iowa State University Press, 1967, p.312.

TABLE 1
LATIN SQUARE DESIGN OF STUDY

	high research	high teaching	low teaching	low research
Prestige:				
male high	A	B	C	D
male low	C	A	D	B
female high	D	C	B	A
female low	B	D	A	C

For example, subject A was given the following four people to evaluate: 1) a male from a high prestige school with high research, 2) a male from a low prestige school with high teaching, 3) a female from a high prestige school with low research, 4) a female from a low prestige school with low teaching.

Thirty-two subjects participated in the study, eight subjects from each of the four departments. In each department two subjects were assigned to each letter. The total score for a cell was the sum of the eight scores.

The category of service was held constant during this experiment since according to the literature, service is not a major focus in this type of decision-making.

CHAPTER 4

PRESENTATION OF RESULTS

Introductory Explanations

There were four achievement levels:

1. high teaching scores, mid-level research score and good service score
2. high research score, mid-level teaching scores and good service
3. low research score, mid-level teaching scores and good service
4. low teaching scores, mid-level research score and good service.

These achievement levels were chosen so that using the criteria given out in the experiment to the faculty subjects, the hypothetical candidates should be ranked in this order. Indeed, if you set the program to evaluate according to the criteria, it will rank in this order. Hence, if the faculty subjects do not rank them in this order, either they are disregarding the criteria or they are showing bias for sex or prestige of institution since these were the only other varied information in the hypothetical candidate's folder.

The following are the scores for the hypothetical candidates. The teaching scores represent student, peer and chair evaluation scores, the research score was derived from the formula on the criteria sheet and the calculations were done for each of the subjects (so there would be no calculation error introduced on the part of the subjects), and service consisted of six items and was above the level stated by the criteria for full credit. (See Appendix A and B).

Following is the data presented to the subjects. These four sets were used in combination with different names (male and female) and different prestige schools (high or low).

TABLE 2
DATA PRESENTED TO SUBJECTS

1. High Teaching	: Teaching (3.28, 3.2, 3.4) : Research (3.0) : Service (six items)
2. High Research	: Teaching (2.98, 3.0, 3.2) : Research (3.2) : Service (six items)
3. Low Research	: Teaching (2.98, 3.1, 3.1) : Research (2.67) : Service (six items)
4. Low Teaching	: Teaching (2.72, 3.0, 2.8) : Research (3.0) : Service (six items)

Overall Analysis of the Data

RAW DATA (Each score is the sum of eight evaluations. Each evaluation is between 0 and 10.)

TABLE 3
SUM OF SCORES FOR EACH OF THE CANDIDATES

high research	high prestige	male	65.1
		female	60.1
high research	low prestige	male	66.56
		female	61.1
high teaching	high prestige	male	67.685
		female	68.785
high teaching	low prestige	male	67.2
		female	69.3
low teaching	high prestige	male	59.715
		female	55.165
low teaching	low prestige	male	52.5
		female	57.9
low research	high prestige	male	57.3
		female	59.9
low research	low prestige	male	57.1625
		female	60.7525

The following is the output from an ANOVA program called UNISTAT that was run on an IBM AT compatible machine.

Table 4
ANOVA: OVERALL

SOURCE: grand mean					
sex	prestige	N	MEAN	SD	SE
		16	61.6391	5.1761	1.2940
SOURCE: sex					
sex	prestige	N	MEAN	SD	SE
male		8	61.6528	5.7291	2.0256
female		8	61.6253	4.9586	1.7531
SOURCE: prestige					
sex	prestige	N	MEAN	SD	SE
	high	8	61.7188	4.9191	1.7392
	low	8	61.5594	5.7619	2.0371
SOURCE: sex prestige					
sex	prestige	N	MEAN	SD	SE
male	high	4	62.4500	4.7760	2.3880
male	low	4	60.8556	7.2168	3.6084
female	high	4	60.9875	5.6766	2.8383
female	low	4	62.2631	4.9054	2.4527
FACTOR:	level	sex	prestige	score	
LEVELS:	4	2	2	16	
TYPE:	RANDOM	WITHIN	WITHIN	DATA	
SOURCE	SS	df	MS	F	p
=====					
mean	60789.9838	1	60789.9838	552.216	0.000 ***
level	330.2511	3	110.0837		
sex	0.0030	1	0.0030	0.000	0.989
sex/level	39.6695	3	13.2232		
prestige	0.1016	1	0.1016	0.046	0.843
prestige/level	6.5569	3	2.1856		
sex/prestige	8.2369	1	8.2369	1.448	0.315
s/p/levels	17.0617	3	5.6872		

These results appear to show that sex and prestige of school do not make a difference. The only significant thing is level of achievement and clearly, ranking on level of achievement is what ought to be done in a faculty evaluation. The overall average score was 61.6391. Males had an overall average score of 61.6528 and females had an overall average score of 61.6253. People from high prestige schools had an overall score of 61.7188 and those from low prestige schools an overall score of 61.5594. There is one anomaly that is worth noting. Males from high prestige schools averaged 62.45 while males from low prestige schools averaged 60.8556. (See Table 4 ANOVA Source sex prestige). The results were almost reversed for females. Females from high prestige schools averaged 60.9875 while females from low prestige schools averaged 62.2631. (See Table 4 ANOVA Source sex prestige)

The above anomaly led to running the tests within achievement level.

Sex Within Achievement Level : High Teaching

RAW DATA (each score is the sum of eight evaluations.)

TABLE 5
SUM OF SCORES FOR CANDIDATES IN HIGH TEACHING CATEGORY

high prestige	male	67.685
	female	68.785
low prestige	male	67.2
	female	69.3

The ANOVA was run on the computer again with the above data.

TABLE 6
ANOVA: SEX WITH HIGH TEACHING CATEGORY

SOURCE: grand mean					
prestige	N	MEAN	SD	SE	
	4	68.2425	0.9679	0.4839	
SOURCE: prestige					
prestige	N	MEAN	SD	SE	
high	2	68.2350	0.7778	0.5500	
low	2	68.2500	1.4849	1.0500	
FACTOR:					
		sex	prestige	score	
LEVELS:		2	2	4	
TYPE:		RANDOM	WITHIN	DATA	
SOURCE	SS	df	MS	F	p
=====					
mean	18628.1554	1	18628.1554	7276.568	0.007**
sex	2.5600	1	2.5600		
prestige	0.0002	1	0.0002	0.001	0.981
pres./sex	0.2500	1	0.2500		

TABLE 6
ANOVA: SEX WITH HIGH TEACHING CATEGORY

There is a significant difference in the way males and females are treated if they have high teaching scores. Prestige of institution does not make a significant difference. People in this level are treated significantly better if they are female than if they are male.

Sex Within Achievement Level: High Research

RAW DATA (Each score is the sum of eight evaluations.)

TABLE 7
SUM OF SCORES FOR CANDIDATES IN HIGH RESEARCH CATEGORY

high prestige	male	65.1
	female	60.1
low prestige	male	66.56
	female	61.1

The ANOVA was run on the computer again using the above data.

TABLE 8
ANOVA: SEX WITH HIGH RESEARCH CATEGORY

SOURCE: grand mean					
prestige	N	MEAN	SD	SE	
	4	63.2150	3.1048	1.5524	
SOURCE: prestige					
prestige	N	MEAN	SD	SE	
high	2	62.6000	3.5355	2.5000	
low	2	63.8300	3.8608	2.7300	
FACTOR:	sex		prestige	score	
LEVELS:	2		2	4	
TYPE:	RANDOM		WITHIN	DATA	
SOURCE	SS	df	MS	F	p
=====					
mean	15984.5440	1	15984.5440	584.382	0.026 *
sex	27.3529	1	27.35290		
prestige	1.5129	1	1.5129	28.599	0.118
pres./sex	0.0529	1	0.0529		

There is a significant difference in the way males and females are treated if they have high research scores. Prestige of institution does not make a significant difference. People in this level are treated significantly better if they are male than if they are female.

Sex Within Achievement Level: Low Research

RAW DATA (Each score is the sum of eight subjects)

TABLE 9
SUM OF SCORES FOR CANDIDATES IN LOW RESEARCH CATEGORY

high prestige	male	57.3
	female	59.9
low prestige	male	57.1625
	female	60.7525

The ANOVA was run on the computer again using the above data.

TABLE 10
ANOVA: SEX WITH LOW RESEARCH CATEGORY

SOURCE: grand mean					
prestige	N	MEAN	SD	SE	
	4	58.7787	1.8213	0.9107	
SOURCE: prestige					
prestige	N	MEAN	SD	SE	
high	2	58.6000	1.8385	1.3000	
low	2	58.9575	2.5385	1.7950	
FACTOR:		sex	prestige	score	
LEVELS:		2	2	4	
TYPE:		RANDOM	WITHIN	DATA	
SOURCE	SS	df	MS	F	p
=====					
mean	13819.7656	1	13819.7656	1442.710	0.017 *
sex	9.5790	1	9.5790		
prestige	0.1278	1	0.1278	0.522	0.602
pres./sex	0.2450	1	0.2450		

There is a significant difference in the way males and females are treated if they have low research scores. Prestige of institution does not make a significant difference. People in this level are treated significantly better if they are female than if they are male.

Sex Within Achievement Level: Low Teaching

RAW DATA (Each score is the sum of eight evaluations.)

TABLE 11
SUM OF SCORES FOR CANDIDATES IN LOW TEACHING CATEGORY

high prestige	male	59.715
	female	55.165
low prestige	male	52.5
	female	57.9

The ANOVA was run on the computer again using the above data.

TABLE 12
ANOVA: SEX WITH LOW TEACHING CATEGORY

SOURCE: grand mean					
prestige	N	MEAN	SD	SE	
	4	56.3200	3.1596	1.5798	
SOURCE: prestige					
prestige	N	MEAN	SD	SE	
high	2	57.4400	3.2173	2.2750	
low	2	55.2000	3.8184	2.7000	
FACTOR:	sex	prestige	score		
LEVELS:	2	2	4		
TYPE:	RANDOM	WITHIN	DATA		
SOURCE	SS	df	MS	F	p
=====					
mean	12687.7699	1	12687.7699	70243.330	0.002 **
sex	0.1806	1	0.1806		
prestige	5.0176	1	5.0176	0.203	0.731
pres./sex	24.7506	1	24.7506		

There is a significant difference in the way males and females are treated if they have low teaching scores. Prestige of institution does not make a significant difference. People in this level are treated significantly better if they are female than if they are male.

Because of the sex-prestige anomaly noted in the first look at the data in Table 4, it is important to look at sex alone and prestige alone.

SEX Over the Whole Experiment

RAW DATA (Each score is the sum of 64 evaluations.)

TABLE 13
SUM OF SCORES FOR MALE AND FEMALE HYPOTHETICAL CANDIDATES

male	495.3225
female	491.4025

The ANOVA was run on the computer again using the above data.

TABLE 14
ANOVA: SEX

SOURCE: grand mean					
N	MEAN	SD	SE		
2	493.3625	2.7719	1.9600		
FACTOR:	sex	score			
LEVELS:	2	2			
TYPE:	RANDOM	DATA			
SOURCE	SS	df	MS	F	p
=====					
mean	486813.1188	1	486813.1188	63360.290	0.003 **
sex	7.6833	1	7.6833		

Men get a significantly higher rating than women over all. Note that even though women appear to be favored in three of the four achievement levels (See Tables 6, 10 and 12), overall men do significantly better. This suggests that the bias against women doing well in research is weighty.

PRESTIGE Over the Whole Experiment

RAW DATA (Each score is the sum of 64 evaluations)

TABLE 15
SUM OF SCORES OF HIGH AND LOW PRESTIGE CANDIDATES

high	494.25
low	492.475

The ANOVA was run on the computer again using the above data.

TABLE 16
ANOVA: PRESTIGE

SOURCE: grand mean					
N	MEAN	SD	SE		
2	493.3625	1.2551	0.8875		
FACTOR:	prestige		score		
LEVELS:	2		2		
TYPE:	RANDOM		DATA		
SOURCE	SS	df	MS	F	p
=====					
mean	486813.1188	1	486813.1188	309028.505	0.001 **
p/	1.5753	1	1.5753		

People from high prestige institutions get an overall higher rating than people from low prestige institutions. Therefore, prestige really was significant even though it appeared earlier that it was not. (See Table 4).

SUMMARY OF RESULTS

Faculty achievement was significant at the .01 level ($p < .0001$). This means that when faculty are evaluated for tenure by a university-wide committee, achievement plays a significant role. For example, people with high scores in teaching received higher ratings than people with low scores in teaching. This was to be expected because the decision was supposed to be based on achievement.

Within the achievement level characterized by high teaching scores, sex was significant at the .01 level ($p = .007$). This means that people with high teaching scores are treated significantly better if they are female.

Within the achievement level characterized by high research scores, sex was significant at the .05 level ($p=.026$). This means that people with high research scores are treated significantly better if they are male.

Within the achievement level characterized by low research scores, sex was significant at the .05 level ($p=.017$). This means that people with low research scores are treated significantly better if they are female.

Within the achievement level characterized by low teaching scores, sex was significant at the .01 level ($p=.002$). This means that people with low teaching scores are treated significantly better if they are female.

In the study overall, sex was significant at the .01 level ($p=.001$). This means that men receive a significantly higher rating than women overall.

In the study overall, prestige was significant at the .01 level ($p=.001$). This means that people from high prestige institutions received a higher rating than people from low prestige institutions. When women were looked at separately, this reversed and women from low prestige institutions were given higher scores than women from high prestige institutions.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The study shows that sex of the candidate and prestige of the candidate's Ph.D. awarding institution do have a significant effect on the candidate's performance evaluation by a university committee. This was not apparent at first glance of the data because men were given less credit at doing a good job teaching than women and women were given less credit for doing a good job in research than men. This balance of bias can deceive the observer into thinking no bias is present. The prestige bias was not apparent because of another balance of bias. Men from high prestige universities were ranked higher than men from low prestige universities while women from low prestige universities were ranked higher than women from high prestige universities.

Recommendations

Tenure decisions in which a woman is turned down primarily for not having enough research should be reviewed very carefully. This study seems to indicate that on the basis of equal research, women are judged lower than men by university committees. The same can be said about a man's being turned down because of teaching scores. It is indicated that on the basis of equal teaching scores, men are judged lower than women by university committees. In general, women are rated lower than men for the same scores. This might argue for careful review of all tenure decisions on female faculty.

A good way of quickly reviewing such decisions would be to use the software discussed in this study. In fact, the act of using the software forces a person to quantify the factors in a tenure decision.

Speculations and Recommendations for Further Research

Why do women from low prestige universities get higher ratings than women from high prestige universities? In any study, there is always the chance that the results are just random error. This same study can be repeated at other institutions and the results can be compared. Another reason for this particular result is the possibility that most faculty consider it unseemly for a woman to get a degree from a high prestige university. Maybe faculty expect more from a woman from a high prestige university and are more generous to a woman from a low prestige university. Possible the faculty feel threatened by a woman from a high prestige university.

Why do men get less credit than women for teaching well and women get less credit than men for good research? Perhaps faculty view teaching as a woman's work and research as a man's job. Because of the way the study was constructed, teaching received more weight than research, which makes these findings all the more dramatic. This implies that a faculty committee might feel that a man can make up for mediocre teaching with excellent research but women can not do that. Further research can be done by replicating the study and changing the instructions. There could be a case where research received more weight and a case where teaching and research received equal weight.

Another implication is that faculty (mostly men in this sample) were more satisfied with the image of a woman getting her degree possibly near home so as not to disrupt family life and becoming a teacher. Further research could be done adding the variables of sex of rater and background of rater (prestige of institution, views on research and teaching, and age).

Further research using the model and the computer program could involve a study looking back at past tenure decisions. Using stated criteria, the program could

rank candidates and this could be compared to actual committee ranking. The main drawback to this kind of study is that it would involve confidential personal files.

Another use of the model and the computer program might be as a simulation tool for administrative decisions. It could be used to show the consequences of several different criteria before policies were set. For example, before announcing a possible new emphasis on research, administrators could estimate how many present faculty would earn tenure based on their current productivity rate.

This model and software would be useful in determining if bias is present in decision-making provided the institution has objectively stated the major criteria in such decision-making. The university committee, when making decisions about candidates in a field other than their own, need objective guidelines and even when given these guidelines there is the possibility of bias entering into the decision-making. There is an implication that if the computer model could provide an unbiased decision, why have human committee members be involved at all? This is an area for further investigation. If a department and its chairman provide input to a candidate's qualifications, why should a dean get input from a university committee made up of human prone to biased decision making and who are not familiar with the candidate's subject area? This is a philosophical question besides one of efficiency and bears further investigation.

APPENDICES

APPENDIX A

Biographical Material on Hypothetical Candidates for Tenure

The sequence letters at the top represent the four achievement levels. W is high teaching, Y is high research, X is low teaching, and Z is low research. The sequence numbers at the top of the other pages represent prestige of institution and sex of candidate. Sequence 1 is female high prestige, Sequence 2 is female low prestige, Sequence 3 is male low prestige, and Sequence 4 is male high prestige.

The four numbers were combined with the four letters to give 16 combinations. These combinations were used in the Latin square design.

SEQ: W

Teaching:**Student Evaluations (0 to 4 scale)**

Year	Average	Department Average
84/85	3.5	2.3
85/86	3.0	2.2
86/87	3.4	2.4
87/88	3.2	2.2
88/89	3.3	2.1
Avg.	3.28	2.24

Peer Teaching Evaluation (0 to 4 scale): 3.2

Chair Teaching Evaluation (0 to 4 scale): 3.4

Research:

"Use of Solar Powered Looms" Journal of Modern Weaving Vol. 11, No. 8, 1985.

"Design of Efficient Injectors" Textile Engineering Journal Vol. 2, No. 3, 1985.

"Weaving Large Plant Economics" Journal of Textile Economics, Vol. 23, No. 4, 1986.

"Injection Molds Have a Place in Cloth Production" Textile Engineering Journal, Vol. 3, No. 4, 1987.

"Looms in Everyday Use" Journal of Modern Weaving, Vol. 15, No. 3, 1987.

"Design of Super Injectors" Textile Engineering Journal, Vol. 9, No. 7, 1988.

Six publications. Since the departmental average over this period was 2 papers the formula yields a score of 3.0 .

Dr. Sandra Ralston

Education: Ph.D. Harvard 1983

Service:

Committees:

- Departmental Laboratory Supply Committee**
- Departmental Curriculum Committee**
- College Standards Committee**
- College Committee on Committees**

Other:

- Introduced a new course: TXS 315 Dynamic Loom Design.**
- Represents the department at freshman orientation.**

SEQ: X

Teaching:**Student Evaluations (0 to 4 scale)**

Year	Average	Department Average
84/85	2.2	2.3
85/86	3.1	2.2
86/87	2.6	2.4
87/88	2.6	2.2
88/89	3.1	2.1
Avg.	2.72	2.24

Peer Evaluation (0 to 4 scale): 3.0

Chair Evaluation (0 to 4 scale): 2.8

Research:

"Cloth Regeneration Techniques" Journal of Modern Weaving, Vol. 11, No. 5, 1986.

"Plant Layout" Textile Engineering Journal, Vol. 3, No. 2, 1986.

"Non-Woven Cloth Production" Journal of Textile Production, Vol. 18, No. 3, 1987.

"Cloth Production: It Can Be Increased Without New Equipment" Textile Engineering Journal, Vol. 3, No. 4, 1987.

"Weaving a Major Factor in Cloth Production" Journal of Modern Weaving, Vol. 16, No. 5, 1988.

"Injection Molds and You" Textile Engineering Journal, Vol. 7, No. 5, 1989.

Six publications. Since the departmental average over this period was two papers the formula yields a score of 3.0 .

SEQ: 2

Dr. Susan Powers

Education: Ph.D. North Texas State University 1983

Service:

Committees:

**Departmental Undergraduate Committee
Departmental Honors Committee
College Planning Committee
College Parking Committee**

Other:

**Introduced a new course: TXS 313 Textile Technology.
Represents the department at parents day.**

SEQ: Y

Teaching:**Student Evaluations (0 to 4 scale)**

Year	Average	Department Average
84/85	3.2	2.3
85/86	2.7	2.2
86/87	3.0	2.4
87/88	3.1	2.2
88/89	2.9	2.1
Avg.	2.98	2.24

Peer Evaluation (0 to 4 scale): 3.0

Chair Evaluation (0 to 4 scale): 3.2

Research:

"Water Powered Looms" Weaving History Journal, Vol. 10, No. 4, 1985.

"Design of Energy Efficient Looms" Textile Engineering Journal, Vol. 3, No. 3, 1986.

"Weaving the End of an Era" Journal of Textile Economics, Vol. 23, No. 3, 1986.

"Cloth Production in Underdeveloped Nations" Textile Economics Journal, Vol. 2, No. 5, 1986.

"Looms of the Future" Journal of Modern Weaving, Vol. 16, No. 2, 1987.

"Design of Cost Effective Textile Delivery Systems" Textile Engineering Journal, Vol. 8, No. 4, 1987.

"Weaving vs Spinning" Journal of Textile Economics, Vol. 26, No. 2, 1988.

"Injection Molds a Thing of the Past" Textile Engineering Journal, Vol. 7, No. 5, 1989.

Eight publications. Since the departmental average over this period was two papers the formula yields 3.2 .

Dr. William Watson

Education: Ph.D. Denver University 1983

Service:

Committees:

**Departmental Laboratory Supply Committee
Academic Senate
College Grievance Committee
College Committee on Committees**

Other:

**Introduced a new course: TXS 330 Textile Plant Design.
Represents the department at freshman orientation.**

Teaching:**Student Evaluations (0 to 4 scale)**

Year	Average	Department Average
84/85	3.1	2.3
85/86	3.1	2.2
86/87	3.0	2.4
87/88	2.9	2.2
88/89	2.8	2.1
Avg.	2.98	2.24

Peer Evaluation (0 to 4 scale): 3.1

Chair Evaluation (0 to 4 scale): 3.1

Research:

"Cloth Production in Industrial Nations" Textile Economics Journal, Vol. 2, No. 5, 1986.

"A Survey of Textile Delivery Systems" Textile Engineering Journal, Vol. 7, No. 6, 1987.

"Cloth Production as a Measure of Wealth" Journal of Textile Economics, Vol. 26, No. 2, 1988.

"Cloth Production in the Underground Economy" Journal of Textile Economics, Vol. 27, No. 3, 1989.

Four publications. Since the departmental average over this period was two papers the formula yields 2.67 .

SEQ: 4

Dr. James Anderson

Education: Ph.D. University of California - Berkeley 1983

Service:

Committees:

**Departmental Laboratory Supply Committee
Departmental Hiring Committee
College Parking Committee
College Honors Committee**

Other:

**Introduced a new course: TXS 380 Textile Delivery
Systems.
Represents the department at graduate orientation.**

APPENDIX B

Criteria for Tenure

You are a member of the tenure committee of the Textile Science Department of Mid-America University. You have been given the task of evaluating four tenure applicants. All of the applicants are assistant professors with five years of service.

Mid-America University rates applicants for tenure on three criteria: teaching, research, and service. The University has carefully studied the problem of weighting these three criteria and has agreed that the relative weights in percent are: teaching - 60% , research - 30% , and service 10%.

Within the criteria of teaching, student evaluations are given a weight of 50% and chair and peer evaluations are given a weight of 25% each. In the area of research, publications and grants are both taken into consideration. One funded is counted as half a paper. The department has come up with a formula for scoring based on number of papers. The formula is as follows:

$$\text{rating} = (1 - (1/(x/a+1))) * 4$$

where

x is the number of acceptable papers published over the time period in question.

a is the average number of acceptable papers published in the department over the time period in question.

Note that if the person publishes no papers the rating is 0, if the person publishes at the department average the rating is 2, and that the rating approaches 4 as the number of papers published grows large. It is very hard to get much above three with this formula.

Service is generally rated by counting the number of service items listed in the folder. If the applicant has at least one item for each two years of service, then full credit is usually given for service.

You are to study the following tenure folders and evaluate the applicants. Evaluate the four tenure applicants by assigning each a real number between 0 and 10. (e.g., 5 or 7.6) Ten indicates a perfect candidate and zero indicates a candidate with no redeeming qualities. Neither ten or zero are common scores for a candidate to receive. Tenure decisions are made by the dean who relies heavily on the rankings of the tenure committee.

APPENDIX C

Instructions

Time Estimate: This will take you approximately 20 minutes.

You indicate your voluntary willingness to participate by completing the attached survey.

This research is being done by Elizabeth A. Hansen as part of a doctoral dissertation in Educational Administration for Michigan State University. The purpose is obtain survey data on decision-making about tenure from faculty members. All results will be treated with strict confidence and the subjects involved will remain anonymous. On request and within these restrictions results will be made available to subjects.

Assume that you are on a university-wide tenure/promotion committee. You will read the "Criteria for Tenure" sheet and then look at the information provided for four hypothetical applicants. You will give each applicant a score from zero to ten and record it on a score sheet which is supplied for each applicant.

From the above explanation of the research and your understanding of the ramifications of it, know that you are free to participate or discontinue participation at any time without recrimination.

Score Sheet

SEQ:

Name of candidate for tenure _____

Score for candidate (must be a number between 0 and 10)

Comments (optional)

APPENDIX D

The Computer Program

-MAIN PROGRAM-

```
DECLARE SUB FirstScreen ()
DECLARE SUB Menu (a$)
DECLARE SUB MakeSC ()
DECLARE SUB MakeCT ()
DECLARE SUB Eval ()
DECLARE SUB Scale (low, high, real, new)
OPEN "DUMMY.DES" FOR OUTPUT AS #1
CLOSE #1
OPEN "DUMMY.WGT" FOR OUTPUT AS #1
CLOSE #1
CALL FirstScreen
response$ = "n"
WHILE response$ = "n"
    CALL Menu(response$)
    IF response$ = "1" THEN
        CALL MakeCT
        response$ = "n"
    ELSEIF response$ = "2" THEN
        CALL MakeSC
        response$ = "n"
```



```
ELSEIF response$ = "3" THEN
    CALL Eval
    response$ = "n"
ELSEIF response$ = "4" THEN
    response$ = "y"
END IF
IF response$ = "n" THEN INPUT "Hit enter to continue ", junk$
WEND
END
```

-EVALUATION SUBPROGRAM-

SUB Eval

REM

REM This subprogram is used to evaluate a person.

REM

CLS

PRINT "Which weighting system do you want to use?"

PRINT "DUMMY is not a valid weight file. "

PRINT "Your choices are: "

FILES "*.WGT"

PRINT "ENTER NAME ONLY NOT THE EXTENSION."

INPUT wfile\$

OPEN wfile\$ + ".wgt" FOR INPUT AS #1

INPUT #1, des\$

OPEN des\$ FOR INPUT AS #2

INPUT #2, n

DIM m(n), h\$(n), w(n)

FOR i = 1 TO n

INPUT #2, m(i)

NEXT i

REM find max

max = m(1)

FOR i = 2 TO n

IF max < m(i) THEN max = m(i)

```

NEXT i
DIM ll$(n, max)
DIM l(n, max)
FOR i = 1 TO n
  INPUT #2, h$(i)
  FOR j = 1 TO m(i)
    INPUT #2, ll$(i, j)
  NEXT j
NEXT i
FOR i = 1 TO n
  INPUT #1, w(i)
NEXT i
FOR i = 1 TO n
  FOR j = 1 TO m(i)
    INPUT #1, l(i, j)
  NEXT j
NEXT i
CLOSE #1
CLOSE #2
DIM sc(i, max)
res$ = "$"
WHILE res$ <> "Y" AND res$ <> "N"
  PRINT "Remember the scores must be between 0 and 5. If you choose any
other"
  PRINT "range you will be prompted for a range with each entry."

```

```

PRINT "Enter Y if all your scores are between 0 and 5."
PRINT "Enter N for individual range prompt at each entry"
INPUT res$
res$ = UCASE$(res$)
WEND
FOR i = 1 TO n
  FOR j = 1 TO m(i)
    sc(i, j) = -1
    sc$ = "q"
    WHILE ((sc(i, j) <= 0) OR (sc(i, j) > 5)) AND NOT sc$ = "0"
      IF res$ = "Y" THEN GOTO skip
back:    PRINT "For "; h$(i); " : "; ll$(i, j)
      INPUT "Enter the lowest possible score ", low
      INPUT "Enter the highest possible score ", high
      IF high < low THEN GOTO back
skip:    PRINT "Enter score for "; h$(i); " : "; ll$(i, j);
      INPUT sc$
      sc(i, j) = VAL(sc$)
      IF res$ = "Y" THEN GOTO skip2
      IF sc(i, j) = low THEN sc$ = "0"
      CALL Scale(low, high, sc(i, j), new)
      sc(i, j) = new
      PRINT "The score was scaled as "; sc(i, j); "on a 0 to 5 scale."
skip2:   WEND
    NEXT j
  NEXT i

```

```
NEXT i

osum = 0

FOR i = 1 TO n
    insum = 0
    FOR j = 1 TO m(i)
        insum = insum + sc(i, j) * l(i, j)
    NEXT j
    osum = osum + (w(i) / m(i)) * insum
NEXT i

PRINT "The total weighted score is : "; osum

PRINT "Normalized weighted weighted score is : "; osum / (5 * n)

END SUB
```

-FIRST SCREEN SUBPROGRAM-**SUB FirstScreen****REM This subprogram displays a title and copyright notice.****CLS****LOCATE 5, 20****PRINT "FACULTY EVALUATOR"****LOCATE 18, 20****PRINT "(c) 1989 Elizabeth A. Hansen"****FOR k = 1 TO 5000****NEXT k****END SUB**

-MAKE CRITERION SUBPROGRAM-

```

SUB MakeCT
REM
REM This subprogram is used to make a criterion file
REM
CLS
n = 0
WHILE n <= 0
    PRINT "How many high level characteristics does this criterion have";
    INPUT n$
    n = VAL(n$)
WEND
DIM h$(n), m(n)
FOR i = 1 TO n
    PRINT "Enter name for "; i; " high level characteristic ";
    LINE INPUT h$(i)
    h$(i) = UCASE$(h$(i))
NEXT i
FOR i = 1 TO n
    m(i) = 0
    WHILE m(i) <= 0
        PRINT "For high level characteristic "; h$(i)
        PRINT " enter the number of primitive characteristics";
        INPUT m$
    
```

```

        m(i) = VAL(m$)
    WEND
NEXT i

REM Find maximum
max = m(1)
FOR i = 2 TO n
    IF max < m(i) THEN max = m(i)
NEXT i

DIM l$(n, max)
FOR i = 1 TO n
    PRINT "Enter primitive characteristics for"
    PRINT "high level characteristic "; h$(i)
    FOR j = 1 TO m(i)
        PRINT "Enter primitive characteristic"; j; " ";
        LINE INPUT l$(i, j)
    NEXT j
NEXT i

PRINT
PRINT "It is now time to select a name for the description file."
PRINT "The name must consist of only letters and numbers."
PRINT "Case is ignored. Other descriptions if any are listed."
PRINT "DUMMY is not a real file. All descriptions have the extension DES."
FILES "*.DES"

PRINT "Caution use of a listed file will destroy that description file."
PRINT "ENTER ONLY THE FILE NAME NOT THE EXTENSION"

```



```
PRINT "Enter file name: ";  
INPUT des$  
OPEN UCASE$(des$) + ".DES" FOR OUTPUT AS #1  
  
WRITE #1, n  
FOR i = 1 TO n  
WRITE #1, m(i)  
NEXT i  
FOR i = 1 TO n  
WRITE #1, h$(i)  
FOR j = 1 TO m(i)  
WRITE #1, l$(i, j)  
NEXT j  
NEXT i  
CLOSE #1  
END SUB
```

-MAKE SCORING FILE SUBPROGRAM-

SUB MakeSC

REM

REM This subprogram weights the criterion in a description file.

REM

CLS

PRINT "Which description file do you want to use?"

PRINT "DUMMY is not a valid description file. "

PRINT "Your choices are: "

FILES "*.DES"

PRINT "ENTER ONLY THE NAME NOT THE EXTENSION"

PRINT "Enter file name: ";

INPUT it\$

OPEN it\$ + ".DES" FOR INPUT AS #1

INPUT #1, n

DIM m(n), h\$(n)

FOR i = 1 TO n

INPUT #1, m(i)

NEXT i

REM Find maximum

max = m(1)

FOR i = 2 TO n

IF max < m(i) THEN max = m(i)

NEXT i

```

DIM ll$(n, max)
DIM l(n, max)
FOR i = 1 TO n
  INPUT #1, h$(i)
  FOR j = 1 TO m(i)
    INPUT #1, ll$(i, j)
  NEXT j
NEXT i
CLOSE #1

PRINT "We need to establish the weights to give high level characteristics."
PRINT "For each characteristic enter the per cent weight it should have."
PRINT "Remember there are"; n; " characteristics and the weights must total 100%"

DIM w(n)
w(n) = -1
WHILE w(n) < 0
  total = 0
  FOR i = 1 TO n - 1
    PRINT "Enter per cent for "; h$(i);
    INPUT w(i)
    total = total + w(i)
  NEXT i
  w(n) = 100 - total
  PRINT "To add up to 100 % "; h$(n); " must be weighted "; w(n); " %."
  IF w(n) < 0 THEN PRINT "Weights must total 100%. Too much weight given to
the first items. Try again. "

```

```

WEND

FOR i = 1 TO n
    w(i) = (w(i) / 100) * n
NEXT i

PRINT "We also need to establish weights for the primitive characteristics."

FOR i = 1 TO n
    PRINT h$(i)

    PRINT "There are "; m(i); " primitive characteristics for "; h$(i)

    PRINT "Remember the weights must add up to 100%."

    l(i, m(i)) = -1

    WHILE l(i, m(i)) < 0

        total = 0

        FOR j = 1 TO m(i) - 1
            PRINT "Enter per cent for "; ll$(i, j);

            INPUT l(i, j)

            total = total + l(i, j)
        NEXT j

        l(i, m(i)) = 100 - total

        PRINT "To add up to 100% "; ll$(i, m(i)); " must be weighted "; l(i, m(i)); " %."

        IF l(i, m(i)) < 0 THEN PRINT "Weights must total 100%. Too much weight
given to first items. Try again."

    WEND

    FOR j = 1 TO m(i)
        l(i, j) = (l(i, j) / 100) * m(i)
    NEXT j

```

```

NEXT i

PRINT "It is now time to select a name for the weight file. The name"
PRINT "must consist of only letters and numbers. Case is ignored. Other"
PRINT "descriptions if any are listed. Dummy is not a real file. All "
PRINT "weight files have the extension WGT."

FILES "*.WGT"

PRINT "Caution use of a listed file will destroy that weight file."

PRINT "ENTER ONLY THE NAME NOT THE EXTENSION."

PRINT "Enter file name: ";

INPUT wgt$

OPEN UCASE$(wgt$) + ".WGT" FOR OUTPUT AS #1

WRITE #1, it$ + ".DES"

FOR i = 1 TO n
    WRITE #1, w(i)
NEXT i

FOR i = 1 TO n
    FOR j = 1 TO m(i)
        WRITE #1, l(i, j)
    NEXT j
NEXT i

CLOSE #1

END SUB

```

-MENU SUBPROGRAM-**SUB Menu (a\$)****REM This subprogram displays a menu and waits for a response of****REM 1, 2, 3, or 4.****REM****WHILE a\$ <> "1" AND a\$ <> "2" AND a\$ <> "3" AND a\$ <> "4"****CLS****PRINT "Make your selection: "****PRINT****PRINT "1 -- Devise an evaluation criterion"****PRINT " Choose high level characteristics and corresponding "****PRINT " primitive characteristics "****PRINT****PRINT "2 -- Weight an evaluation criterion"****PRINT " Weight the characteristics in terms of per cent"****PRINT****PRINT "3 -- Evaluate an individual"****PRINT " Calculate a score based on a particular weighting of characteristics"****PRINT****PRINT "4 -- END"****PRINT****INPUT a\$****WEND****END SUB****-SCALE SUBPROGRAM-**

SUB Scale (low, high, actual, new)

REM

REM This subprogram scales scores so that they run from 0 to 5.

REM

REM low -- holds lowest possible score

REM high -- holds highest possible score

REM actual -- holds actual score value on old scale

REM new -- holds score scaled as a number between 0 and 5

REM spread -- holds the range of scores on old scale

REM

spread = high - low

new = ((actual - low) / spread) * 5

END SUB

BIBLIOGRAPHY

BIBLIOGRAPHY

- Blackburn, Robert. "The Meaning of Work in Academia," New Directions for Institutional Research, San Francisco: Jossey-Bass, 1974, I, p. 80.
- Braxton, J. M., and Brayer, A. E. "Assessing Faculty Scholarly Performance," New Directions for Institutional Research, San Francisco: Jossey-Bass, 1986, p. 28.
- Breneman, David W., and Youn, Ted I. K. (eds.), Academic Labor Markets and Careers, Philadelphia: Falmer Press, 1988.
- Caplow, Theodore and McGee, Reece J. The Academic Marketplace, New York: Arno Press, 1977.
- Centra, John A. "Using Student Assessments to Improve Performance and Vitality," New Directions for Institutional Research, San Francisco Jossey-Bass, 1978, p. 40.
- Conte, S.D., Dunsmore, H.E., and Shen, V.Y. Software Engineering Metrics and Models, Menlo Park, California: Benjamin/Cummings Publishing Co., 1986.
- Dill, David. "Research as a Scholarly Activity: Context and Culture," New Directions for Institutional Research, San Francisco: Jossey-Bass, 1986, XIII, p. 13.
- Etzioni, A. and Lehman, E. W. "Some Dangers in 'Valid' Social Measurements," The Annals, 1967, 373, pp. 1-15.
- "In Weighing Sex-bias Case, High Court Will Skirt Issue of Confidentiality of Tenure-Review Records," The Chronicle of Higher Education, January 4, 1989, p. A13.
- King, Patricia. Performance Planning and appraisal, New York: McGraw-Hill Book Company, 1984.
- Logan, Wilson. The Academic Man: Sociology of a Profession London: Oxford University Press, 1942.
- Long, Durward. "The University as Commons: A View from Administration," New Directions for Institutional Research, San Francisco: Jossey-Bass, 1977, V, p. 75.
- Nash, Michael. Making People Productive, San Francisco: Jossey-Bass, 1985.

Negotia, Constantin V. and Ralescu, Dan. Simulation, Knowledge-Based Computing, and Fuzzy Statistics, New York: Von Nostrand Reinhold Co., 1987.

Schuster, J. N. and Bowen, H. R. "The Faculty at Risk," Change, 1985, 17, pp. 15-16.

Seldin, Peter. Successful Faculty Evaluation Programs, Crugers, New York: Coventry Press, 1981.

Smith, Donald K. "Faculty Vitality and the Management of University Personnel Policies," New Directions for Institutional Research, San Francisco: Jossey-Bass, 1978, V, p. 9.

Snedecor, George W. and Cochran, William G. Statistical Methods, Ames, Iowa: The Iowa State University Press, 1967.

MICHIGAN STATE UNIV. LIBRARIES



31293006059590