JOB EVALUATION AS A DETERMINANT OF JOB WORTH: A CONCEPTUAL AND COMPARATIVE ANALYSIS

By

Robert Michael Madigan

A DISSERTATION

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

DOCTOR OF PHILOSOPHY
School of Labor and Industrial Relations
1982

ABSTRACT

JOB EVALUATION AS A DETERMINANT OF JOB WORTH:
A CONCEPTUAL AND COMPARATIVE ANALYSIS

BY

Robert Michael Madigan

Determination of fair pay is one of the complex and apparently insoluble problems which societies and organizations periodically must confront. The current controversy over the principle of equal pay for work of "comparable worth" or value to the employer introduces a new dimension into debates regarding fair pay. Disparity between average male and female earnings is attributed by comparable worth advocates to valuation of traditionally female occupations based on their sex composition rather than on contribution to organizational objectives. It is proposed that pay be based on the assessed worth of jobs/occupations to the employer, such worth to be determined by means of "non-biased" job evaluation procedures.

The feasibility of non-biased job evaluation is explored here in two ways. First, the concept of worth or value is examined in terms of historical notions of worker/job worth, and through consideration of theoretical and philosophical approaches to defining and determining relative worth. This

discussion serves to underscore the extreme difficulties attendant to definition of the conceptual boundaries and components of job worth in any given situation.

Second, actual job evaluation ratings are analyzed to estimate their measurement qualities and the degree to which different methods generate similar value hierarchies from a common set of jobs. Four raters (analysts) conducted job analyses and evaluations of 20 jobs utilizing three methods - the Position Analysis Questionnaire, a standard three factor guidechart plan, and a locally developed plan. The obtained scores were tested for reliability, bias, dimensionality, and method convergence to develop evidence regarding the technical feasibility of non-biased evaluation.

Findings reported here are basically pessimistic. While differences in measurement quality among the three methods were found, the deficiencies of even the best case (local plan) are such that unacceptable levels of measurement contamination are probable in any instance. Furthermore, significant method divergence in terms of results may also be indicative of job worth construct variation and/or criterion deficiency. Consequently pay grade assignment by means of formal job evaluation processes is likely to vary significantly by situation and method, and the feasibility of non-biased evaluation is highly questionable.

To Judy
whose estraordinary self denial and unfailing support
made this project possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Conflict regarding appropriate bases for establishing the worth of labor is one of the recurring themes of history. Vineyard laborers in the biblical parable are depicted as dissatisfied with pay perceived as disproportional to duration of effort. Medieval societies regulated wage decisions through secular and religious rules specifying "just wages;" market determination of worth was specifically rejected in favor of socially determined wage value (Fogarty, 1961). The development of true market economies in the eighteenth and nineteenth centuries and corresponding acceptance of the tenets of laissez faire capitalism legitimized the market definition of worth, but income disparities reached extreme levels and ultimately became one of the primary forces behind nineteenth century European revolutionary movements. Job worth determination continues to be a central issue in contemporary Western societies as evidenced by the variety of wage setting institutions and procedures and the regularity of pay equity disputes and work stoppages.

Since wages and income are synonymous for the majority of people, issues pertaining to the determination of wage levels for jobs or occupations cannot be divorced from the larger question of the distribution of income and wealth in

the society. Judgments regarding the fairness or justice of income distribution inevitably are made, at times without recognizing the fundamental value conflicts which characterize disputes about distributive justice. Some individuals value efficiency highly, thus their concept of social justice requires distribution according to merit or productivity; others may stress equality or needs. However, extreme emphasis on efficiency typically increases income disparity while equality of distribution eliminates the most powerful incentives to produce. Workable (acceptable) solutions to the problem of distributive justice must recognize the legitimacy of these and other competing values. Consequently, operational definitions of social justice revealed in political party platforms and legislative proposals normally reflect multiple values (efficiency, liberty, equality, quality of life) differing only in emphasis, and appeals to their particular views of social justice are made by all parties to the political debates.

In a number of European countries, beliefs regarding the just distribution of income have been translated into job worth determination policy and statutes specifying wage criteria (Oettinger, 1964) and/or wage relationship objectives (Van Otter, 1975). Job worth decisions in the United States, however, are free of government restraint with the exceptions of minimum wage and anti-discrimination statutes. Actual wage setting decisions within organizations, unilateral or bilateral, normally stress the concept of market

determination of value in combination with the personal, job characteristic and job context criteria considered relevant to the particular situation. These practices are now being challenged. Proposals for direct government intervention into job worth decision making have been advanced under the rationale that current practices are inherently discriminatory against females.

Data apparently supporting the claim of sex based pay discrimination are readily available. For the past twenty years median male earnings have exceeded women's earnings by approximately 70%, a period during which the participation rate of women has risen from 37 percent to 50 percent (U.S.Department of Labor, 1979.) Attempts to explain the differential in terms of employee quality (Fogel, 1979; Gunderson, 1978) and job demands (Halaby, 1979; Englund and McLaughlin, 1979) have been unsuccessful. Nor does the gap disappear when earnings are compared within occupational categories (U.S. Department of Labor, 1979). Consequently, the unexplained portion of the differential, the residual after all "legitimate" reasons have been considered, is typically attributed to sex discrimination.

The persistence of male/female earnings differentials since the 1963 passage of the Equal Pay Ace (EPA) has focused increasing attention on determinants other than direct "within job" wage discrimination (same-job--different pay). Variance in average earnings is also a result of differential distribution across jobs/occupations. Women are concentrated

in relatively low paying occupations and in the lower strata within higher paying occupational groups. Thus, two basic questions are being asked: (1) Why do the male and female occupation distributions differ? (2) Why are women concentrated in the low income distribution, i.e., what is the nature of the occupation distribution-wage level relationship, if any?

Investigations of sex segregation of occupations/jobs take two approaches to the issue. On the one hand, segregation is explained in personal or self selection terms focusing on aspects of female socialization practices thought to channel women toward a limited number of traditional occupations and discourage career achievement. On the other hand, segregation is viewed as a consequence of a complex of institutional and organizational barriers to occupational entry and progression by females. While occupational segregation is undoubtedly a product of both types of forces, the Civil Rights Act of 1964 (CRA) and subsequent regulatory agency rules have embodied the latter view in national policy.

Two basic interpretations can be made of the relationship between female concentration and occupational earnings levels. First, low average earnings may be due to actual low ranking or scores of female occupations on job worth criteria and/or low market assignment of value. Second, causality may be reversed. Lower pay criteria and market worth determinations may result from the fact of high female intensity in

certain jobs or occupations, i.e., "between jobs" wage setting decisions may be based on sex, thus potentially violating the CRA. Both interpretations are probably correct but the current challenge to wage determination practices is based on the belief that a substantial portion of the male-female differential is attributable to systemic discrimination in wage setting practices. If so, changes in job worth decision criteria and processes will have an immediate impact on the earnings gap.

Proponents of change argue that the marketplace has historically discriminated by establishing lower rates of compensation for predominantly female jobs. A pervasive male bias in the structure of society (Blumrosen, 1979), employer motives (Oaxaca, 1977) and/or differing male-female labor supply elasticities caused by restrictions in female access to occupations (Stevenson, 1978) are among the most frequently cited determinants of labor market discrimination. Internal (organizationally determined) criteria of worth are also considered discriminatory since in concept and definition they usually represent male values and orientations to work; job evaluation system design and implementation has been a management (predominantly male) function. Similarly, wage setting under collective bargaining is assumed to be discriminatory in view of historic union exclusion of women and male domination of union leadership.

Proposed solutions to the alleged discrimination problem involve three basic objectives. First, expansion of the EPA

definition of sex based pay discrimination is necessary to incorporate discrimination in "between jobs" situations (dissimilar work). Replacement of the current work equality standard with one requiring only comparability is advocated. Thus discrimination would exist where different rates are paid to males and females occupying jobs of "comparable worth" or work of equal value to the employer. Second, the nullification of discriminatory influences in labor market determinations of job worth is sought. The exclusive use of internal criteria for establishing relative worth and statistical control for market bias have both been suggested as means to accomplish this goal. Solutions to the third objective, elimination of male bias in job evaluation criteria and processes, generally advocate the inclusion of women in evaluation plan design and administration.

In short, critics of current wage setting practices are suggesting that "non-biased" administrative tools and procedures can be developed and employed to objectively assess the worth of jobs. These bias free approaches, hereafter referred to as comparable worth job evaluation systems, will then provide the means to operationalize the comparable worth standard for establishing the presence/absence of sex based pay determination.

This dissertation is a response to the calls for national implementation of the comparable worth job evaluation concept. The scope of discourse is limited to questions of feasibility; lack of data and absence of

specific statutory proposals transform discussions of potential consequences into speculative exercises. More specifically, the focus of the investigation is twofold. The concept of job worth is examined first in terms of past and present approaches to its definition and theoretical explanations for job worth differentials. Of necessity this discussion bridges a number of disciplines presenting a quandary, since compression of arguments and concepts beyond my particular field of expertise are required. The solution adopted is to adhere to the time-honored strategy of parsimony; comments are limited to the minimum amount necessary to convey the essence of the concept or theory. This section provides the ideological backdrop for the subsequent discussion.

Second, the feasibility of using job evaluation tools and procedures as determinants of job worth is explored through assessment of the measurement characteristics and the degree of correspondence of actual comparable worth job evaluation outcomes. This empirical portion of the dissertation is initiated with a conceptual analysis of contemporary wage setting practice as it contrasts with comparable worth job evaluation proposals and concludes with a summary of the major considerations pertaining to the feasibility of adopting a comparable worth job evaluation policy.

CHAPTER 1

Job Worth Concepts and Perspectives

While the focus of the comparable worth debate is on wage setting and male-female pay differentials, the conflict is rooted in differing underlying concepts of worth or worker value. Defenders of traditional approaches to job evaluation typically define worth in terms of market or exchange value, the rates established through individual or collective employer-employee transactions. Comparable worth advocates tend to reject economic exchange in favor of specific criteria reflecting the relative importance of jobs to the organization; job worth is assumed proportional to the contribution to organizational goal attainment. This organizational or "use" value need not be the same as exchange value, nor do these two concepts exhaust the bases for establishing the value of jobs. Workers, job characteristics per se, and societal goals have also been suggested as foci of value determination.

The purposes of this chapter are threefold. First, basic concepts of value are discussed in order to clearly establish the subjective and monetary aspects of the construct. Second, theoretical perspectives pertaining to job worth, pay differentials, and worth determination processes are examined in terms of their basic assumptions,

components, and implications for wage determination. Third, the normative problem of justice is investigated. Parallels between approaches to the definition of distributive justice and theoretical views of job worth are traced in order to emphasize the philosophical dimensions of the comparable worth debate.

WORTH/VALUE CONCEPTS

Discussions regarding the principle of equal pay for jobs of comparable worth can not sidestep the basic question of what constitutes value. The development of bias-free approaches to assessing comparability among jobs presumes a common understanding of what is meant by value and how it is measured.

Definitions of value (considered synonymous with "worth" here) reveal the quantitative, subjective, and ethical facets of the concept (Webster's Twentieth Century Dictionary, 1978). (1) Value is defined in terms of monetary worth in the market place, hence the ultimate determinants of value are market forces; (2) the source of value may be viewed as embedded in a particular set of tasks, the intrinsic qualities or characteristics of which make the job useful to a particular employer. In this sense value is inherent in the object of valuation and independent of the market place; (3) value is considered as that which is the fair or equitable equivalent (in money or commodities), i.e., value is inherently an ethical concept. The attempt here is to explore some of these facets through a brief review of historical

concepts of value. This, in turn, will facilitate the subsequent reviews of theoretical and philosophical perspectives.

For most of recorded history the concept of job or labor value was inextricably intertwined with social rank or status.[1] Economic and social life were one and the same. The view of labor as an abstract factor of production, a commodity being bought and sold in myriad employer-worker transactions did not exist. In early tribal societies division of labor and private property existed, but economic exchange as the means of meeting human wants was the exception rather than the rule. Material needs were met through the efforts of individuals and primary social units; distribution was governed by social responsibilities. As societies became more complex - division of labor, accumulation of wealth, and social stratification based on private property - distribution continued to be determined by social relationships defined and enforced through military-political processes. The worth of an individual was defined at birth.

From time to time challenges to the existing order, usually the result of growing economic power of commercial interests, stimulated limited examination of the concept of value. Changes in the distribution of economic power threatened existing social relationships, and various aspects of commercial practice such as pricing, interest, and capital accumulation were examined in terms of their impact on the "natural order." Restrictions on these practices were thus justified on moral bases.

Aristotle indirectly provided the first recorded analysis of the concept of value in his discussion of the art of exchange. His focus was not on labor value (class distinctions were assumed) but on articles of commerce; his intent was not to develop a theory of the factors determining exchange value but to develop an ethical base for limiting the scope of commerce. Aristotle developed the distinction between use value and exchange value of an article to explain how the natural purpose of exchange (satisfying wants) can evolve into the unnatural use (usury) of an exchange medium (money) as a source of accumulation. To him trade was an unnatural occupation and the growth in power of the commercial class a major contributor to the social conflict of his time. He also recognized the use of money as a conventional measure of exchange value (in his discussion of justice) but his is essentially an ethical analysis attempting to limit the accepted basis of value to societal use.

Portions of Aristotelian thought have found their way into contemporary economic analysis. His recognition of consumer wants as the ultimate source of exchange value is an early expression of a utility theory of value even though his focus was on the problem of equivalence in exchange. More importantly, his use v. exchange value distinction was developed by classical economists and remains an element in current economic thought.

Justifications of the economic order became more sophisticated and intensive during the Middle Ages as increasing

power of merchants and artisans threatened feudal institutions. Interestingly, while Christ's teaching regarding the worth of labor was diametrically opposed to Greek thought, Aristotle also provided a basis for medieval scholastics' attempts to define a "just wage." Christ proclaimed a brotherhood of man and emphasized the worthiness of all labor in both a spiritual and material sense; Aristotle accepted class distinctions and slavery. This apparent conflict was resolved by emphasizing the spiritual aspects of life. Worldly inequality of men was accepted, indeed it was considered the foundation of the land based and rigidly stratified medieval society. However, the pursuit of wealth was inherently evil. The justification for wages could not, therefore, rest upon supply-demand interactions since imbalances potentially foster avarice. Rather wages were justified by employing Aristotle's distinction between natural and unnatural forms of supply. Wage transactions based on the value of the labor to the community were appropriate; transactions exploiting bargaining leverage were unjust.

Fogarty (1961) summarizes the medieval scholastic philosophers' theory of just wages into three basic principles. The first principle defined value not as measured by the market directly, but through a social process considering value to the community: "(P)ay should be equal to the value of the employees' working capacity." The basic value framework was established by authorities with the specific settlement reached by the parties to the employment relationship. Value

to the community was assessed in terms of capacities, not in how the employer utilized the worker. It was the employer's job to ensure the capacity was effectively utilized. Thus wage value was socially determined; use value to the community justified pay.

The second just wage test considered the employee's standard of living calling for stability and security of earnings. In essence, it was a need criterion of value which differs significantly from the modern concept since inherent inequality among men was the prevailing view. Thus needs (and value) varied by social status.

The third, and perhaps most fundamental principle, was that pay serves the common good. Since the needs to be maintained were social products and pay an element in this social effort, wage transactions could not exceed the limits the community would accept. In effect, this principle called for the explicit determination of community values and distribution in accordance with rules serving that view of the common good.

The just wage principles are interesting here for at least two reasons. First, the concept of value was again considered from an ethical perspective, in this case grounded in moral theology. Second, the justifications advanced those hundreds of years ago are similar to criteria advanced today. If one takes a long enough perspective it appears a case could be made that the historical path is circular, and

perhaps with respect to job/worker value concepts, the past is indeed prologue.[2]

Classical Economics

Where labor or worker value was a peripheral issue to the Greek and Scholastic philosophers, it was central to classical economic thought. The objective of the classical philosopher-economist was not to justify but to understand, to search for the regularities underlying social phenomena pertaining to the creation and distribution of wealth. Grinding poverty of the masses in stark contrast to ruling class luxury was one such phenomenon for which explanations were sought, not in ethical precepts, but in natural forces.

Sir William Petty first sketched some of the basic elements of classical theory pertaining to job or worker value in the late 17th century. Petty's theory of value emphasized labor as the primary source of all wealth and his measure of value in terms of the food requirements of an adult male presaged the development of subsistence theories of wages. Numerous other writers addressed these and other economic issues during the ensuing 100 years, but it was left to Adam Smith to formulate a comprehensive statement of economic principles.

Adam Smith's An Inquiry into the Nature and Causes of the Wealth of Nations describes a system of economic activity which is a self energizing, self regulating engine of progress.[3] With respect to job worth, Smith explained the

nature and mechanics of market determination of value and attributed the poverty of the times to a secular and cyclical interaction between expansion of wealth and population growth which served to depress worker wages toward subsistence levels.

Two elements of Smith's analysis are particularly noteworthy here. First, he employed an exchange value definition of job worth. While he recognized the utility or use value of a commodity, it is the measurement of the natural or real exchange value, its components and relationship to market pricing with which he was primarily concerned. Following Petty, Smith traced the origin of wealth to the labor embodied in the generation of the commodity. Since labor itself is a commodity, the natural value of labor is determined ". . .by what is necessary to maintain the laborer plus an allowance to enable him to rear a family and maintain a supply of labor" (Roll, 1973:164). Differences between this natural value and market price (exchange value) were due to excesses or deficiencies of supply. Smith recognized that the natural or labor value can differ from exchange value and use value but he argued that the trend will be toward exchange and true value equality. Thus he provided a form of moral justification for market determination of worth in the guise of natural law.

Second, Smith's explanation for wage differentiation in terms of worker decisions regarding the net advantage of alternative employments recognized both qualitiative

dimensions of job worth (supply side constraints) and demand side limitations on the competitive model. He asserted that the market is cleared in terms of net advantage, not wages. Exchange value is actually a function of pay and various nonmonetary employment characteristics which workers consider in determining whether to seek or accept specific jobs. Five such job facets were identified by Smith: (1) Disagreeableness in terms of physical conditions or prestige; (2) difficulty and expense of learning; (3) security of employment; (4) trust requirements, an accountability for wealth, reputation, or health concept; (5) risk of failure in the profession. However, Smith warned that these job characteristics result in equality of total advantages and disadvantages only under conditions of perfect freedom when the occupation is well established, demand patterns are normal, and the job is the primary income source (Smith, 1937).

In summary, Smith's discussion of value and wage determination identified most of the elements of modern analysis. General wage level determination was explicated in terms of the market mechanism. The relationship between monetary and nonmonetary rewards was addressed. Wage differentiation was traced to both supply and demand factors with the latter treated as deviations from the natural or normal functioning of the market. Actual wage rates as determined by the market were distinguished from the "real" value of labor and the difficulty of measuring the real value was addressed. Finally, a rationale for the tendency of real value and

monetary rates to equate was advanced, thus supporting the concept of market measure of value.

Marx

The treatment of labor value in the writings of Karl Marx is closer in intent to the Greek or Scholastic than the classical economic tradition. Marx developed his labor theory of value to defend a view of history much as these early writers sought to justify an existing order. He followed Smith and Ricardo both in defining the true value of labor as that amount necessary to ensure perpetuation (subsistance) of the supply of labor, and in recognition of value distinctions. However, he did not attempt to develop a theory of value or wages. Rather, the value concept was used to provide a basis for his theory of exploitation. Marx argued that the units of labor power consumed by a capitalist (use value) exceed the exchange value of labor (subsistance level wages) giving rise to a "surplus value" which accrues to the capitalist when the product is sold at its true or exchange value. This concept of profits as surplus value was then used by Marx to describe the laws of motion in a perfectly competitive capitalist system which inexorably lead to system disintegration and transfer of power to the workers (Heilbroner, 1972).

Marx's contributions to our understanding of value or job worth are nil. His relevance here is as an example of 19th century criticism of classical political economy grounded in a desire for social reform and centered on

distributional issues. As such his surplus value can be seen as an essentially ethical concept in theoretical garb. Injustice in income distribution is to be resolved through complete divorce of wages from production value. The Marxian view of ultimate social determination of income based on need represents an extreme perspective, but other remedies of the times also called for modification of exchange value through some element of social intervention in wage determination.

Marginal Utility

The emergence of marginal analysis marked a significant departure from previous approaches to determination of value (and job worth). The ultimate basis of value shifted from the production perspective of labor (however difficult to measure) to subjective utilities and costs. In marginal analysis the interaction of supply price with aggregated individual utilities (consumer demand) establishes the value of commodities in the market place. Job value is thus a function of rates in a competitive labor market and the marginal utility accruing to the employer from employees. Employer utility, in turn, is sensitive to consumer utility in a competitive product market, thus creating a closed system in which wages are directly responsive to the satisfaction of societal wants. Consequently, the need to reconcile concepts of value is avoided in marginal analysis since use value or utility is equivalent to exchange value.

Marginalist thought provides the rationale for contemporary advocates of market determination of worth. It appears

to offer significant advantages over previous approaches to explaining job value.[4] Paradoxically, the introduction of subjectivity into value theory facilitates the claim of objectivity in wage determination. Wages are seen as the product of impersonal labor and product markets. The problems of determining explicit criteria which plagued the Scholastics and the troublesome abstractions of labor theories of value disappear. Marginal utility theory appears to sidestep value judgments and place job worth determination on an impersonal basis. Wage levels and differentials are the product of market phenomena, not employer or government decisions. The illusory nature of the objectivity will be discussed later.

Discussion of marginal utility concepts of value and job worth also provides the logical point for terminating this brief review of job worth concepts, for while its roots are in the 19th century, marginal utility is an accepted basis of contemporary economic theory. Twentieth century developments have provided refinements rather than new insights into the nature of job value. A summary of the preceding discussion is now appropriate.

Summary

Historical concepts of value clearly reflect the three definitional elements mentioned at the outset. The ethical nature of value determination is explicit in the predominant schools of thought until the 18th century, and implicit but no less significant in classical and subsequent viewpoints. Adam Smith was a moral philosopher who developed a system of

economic principles in consonance with his naturalist beliefs. In so doing, he shifted the touchstone of morality in economic affairs from various socially determined criteria to the "natural order." Economic policy and activities were henceforth to be judged on the basis of their concordance with natural law. Monetary value as established in the free market achieved transcendence over use value in determining the right or just job/worker worth; natural law replaced religious law as the guide to morality in pay determination.

A search for the inherent characteristics or qualities underpinning true or real value is a constant element in the history of efforts to justify or modify earnings distribution patterns. From Aristotle to the present the ultimate source of value has usually been located in some notion of usefulness in meeting societal wants, considered from either a consumption (just price; utility) or a production (labor theory of value) perspective. These hypothesized bases of value have been rationalized through logic systems premised on some view of the right or just order. A clear distinction between concepts of true value and the monetary value resulting from market forces was maintained until a rationale for merging the two concepts was provided by the emergence of classical and neoclassical theory. This distinction is now somewhat blurred, perhaps, but the concepts of use value and exchange value are still in evidence in contemporary pay setting. Employers and workers alike employ use value concepts in establishing pay differentials within broad market parameters.

In conclusion two points deserve emphasis. First, distrust of market determination of value has been the rule rather than the exception over the centuries. But where early concerns were with the power of market forces to realign political/economic structures, modern criticisms center on the many discrepancies between the theory and reality of labor markets. These market "imperfections" are the focus of some of the theorizing described in the following section.

Second, discussions of job worth cannot avoid the issue of values. This fact is readily apparent in regard to use value or value centered in personal characteristics/status. It is less visible with respect to market value where one must recognize the roles, both of societal values in shaping income distributions/demand patterns underlying labor markets, and organizational values in making job worth judgments within the market constraints. These issues will be elaborated upon in the following sections and in chapter 2.

CONTEMPORARY THEORETICAL PERSPECTIVES

Task number one in any review of theoretical formulations is to circumscribe the domain of interest, an undertaking of considerable difficulty in this instance. Job worth determination processes are comprised of economic, social, political, and psychological dimensions. The common element of the frameworks summarized below is their direct focus on the phenomena of central interest, pay differentials and the processes by which they are generated. Basic market

forces as described by traditional economic wage theory[4] are assumed operative (although not in their theoretical purity); discussion here, with one exception, is confined to theories considering relative wages as a function of individual employer/employee decisions. Excluded are the general decision making frameworks and other process schemes such as game or stereotype theory, a discussion of which would be peripheral to the issue of pay differences. For presentation purposes the theories are somewhat arbitrarily categorized as supply or demand oriented based on the primary object of inquiry.

Supply Perspectives

Three different aspects of worker decision behavior affecting pay distribution are considered by the formulations presented in this section. The first employs the classical competitive labor market model to equate pay differentials with variance in marginal productivity. Perspectives two and three introduce supply side qualifications of the wage competition model based on consideration of actual job choice behavior, and a view of workers as equity seeking rather than wage maximizing entities, respectively.

Human Capital-- Human capital theorists apply the concept of return on investment to relative wage analysis.[5] Workers accumulate capital through productivity improving investments in education, skill, training, and work experience with which various direct (tuition, etc.) and indirect (foregone income) costs are associated. Individuals presumably calculate cost/benefit analyses to make their investment

decisions. Since investment actions directly translate into productivity improvements, marginal product value and wages increase correspondingly. Therefore, individual investment behavior is the primary determinant of pay distributions; qualitative differences in labor supplied by individuals or groups (male-female) are reflected in pay structures.

Investigations of male-female pay differences within the human capital framework implicitly define discrimination as unequal pay for equal productivity characteristics. Worth is centered in the person, not the job. Therefore discrimination is operationally defined as the unexplained differential in male-female earnings when human capital variables are controlled - assuming the majority pay structure would prevail in the absence of discrimination. Estimates of sex discrimination obtained by decomposing earnings differentials into explained and residual (discrimination) portions, whether at the macro (Suter and Miller, 1973), occupational (Cohen, 1971), or enterprise level of analysis (Malkiel & Malkiel, 1973) generally indicate moderate to low sensitivity to human capital variables leaving sizeable residuals (20-40%). Even when analysis of differentials is conducted within sex (males only) across occupations (Fogel, 1979), or when job characteristics are added to human capital control variables (Halaby, 1977; Englund and McLaughlin, 1979) large residuals persist.

The human capital explanation of earnings differentials is supportive of two basic components of American ideology,

the superiority of a free market economy and an emphasis on individual effort and responsibility. (Workers who expend the effort and time to improve themselves will be rewarded in the market place). This may partially explain the tenacity with which belief in the competitive labor market is held in the face of conflicting information. To the degree continued adherence to a doctrine of market determination of value ignores the impact on pay of institutional arrangements and behavior, acceptance of market pricing reflects cultural, not economic forces. Thus the implications of human capital theory for the present discussion rest not in its explication of market mechanisms, but in the extent of its inadequacies. Worker characteristics may be only a minor element among the determinants of relative job worth. The perspectives summarized below offer alternative explanations, some of which take direct issue with the view of pay differentials as a function of preparation costs.

Occupational Choice-- Job or occupational choice theories are simply particular applications of human decision making models which offer conflicting views of the mechanisms by which worker choice affects pay differentials. One perspective is represented by the classical economic assumption of rational man, maximizing self interest by carefully weighing all alternatives before making a job choice. Differentials are easily accommodated within this model via nonmonetary rewards. This, of course, was Adam Smith's net advantages rationale, i.e., workers consider not only investment

costs but factors such as security and working conditions in choosing jobs. The inherent subjectivity of this view makes testing difficult, but simple observation indicates it is not entirely in accord with reality. For example, insecure or disagreeable jobs are as likely to be negatively as directly correlated with pay. Job content and context differences undoubtedly do affect pay differentials, but mechanisms other than job choice decisions probably also are involved.

March and Simon (1958) popularized a conflicting model of decision making behavior featuring sequential consideration of alternatives relative to a satisfying (rather than maximizing) criterion. In their view rationality is always limited by incomplete knowledge of options and consequences. Thus workers consider a limited number of jobs accepting the first one deemed minimally acceptable. This outlook is compatible with the obvious limitations on actual self improvement or job choice decisions (poverty, educational limitations, mobility restrictions, and inadequate knowledge of opportunities). Furthermore, other labor market phenomena such as geographic stability and the propensity of children to adopt the parental occupation support the limited rationality model. Thus job choice impact on pay differentials is more likely to be grounded in imperfect knowledge and economic need than in utility or net advantage maximization.

Equity Decisions-- Few things will more quickly arouse individuals and/or move them to action than the perception of

being treated unfairly in some respect. The consequences of inequity perceptions regarding pay are such that managements have devoted a considerable amount of effort to ensuring conformity with the prevailing norms of equity.[6] In this sense equity judgments by workers are a direct determinant of pay differentials.

A large body of theoretical and empirical work in social psychology has developed since the 1940's examining the function of social comparisons in the formation of equity/inequity perceptions, particularly with respect to financial compensation. From this perspective, equity decisions, like judgments of deprivation (Runciman, 1966), are always relative; no absolute criterion exists.

The equity model most frequently used in research (Adams, 1965) is a direct descendant of Aristotle's notion of equity as proportionality of ratios. Judgments of inequity are based on expectations which are formed by comparisons of the individual's outcome/input (rewards/investment) ratio to perceptions of a relevant other or others' ratios. Adams postulates that a perceived inequality in ratios induces a tension or drive to reduce or avoid the inequity.

Outcomes, inputs, and referents are all subjectively defined. Outcomes include any monetary or nonmonetary returns from the job; inputs encompass all factors (personal, job content, context) considered relevant to the generation of some return; comparison objects are specific individuals or some abstract composite other. Since perceived ratio

inequalities can be positive or negative, the tension may take the form of guilt or dissatisfaction, with the latter the primary concern for pay differential determination purposes. Perceived pay inadequacies lead to reduced effort, turnover, internal disruptions and other undesired (by management) outcomes.

The equity theory based research is voluminous, well reviewed elsewhere (Admams and Freedman, 1976; Carrell and Dittrich, 1978) and generally supportive of the theoretical propositions regarding underpayment effects. Of most relevance here are not the findings but the problematic aspects. First, the pay criteria (inputs) considered by individuals are apparently multiple, vary among persons and across situations, and in some cases function as both input and outcome (e.g., responsibility). Some guidance regarding the type of inputs considered can be offered (Finn and Lee, 1972) but administrative prescriptions are impossible.

Second, equity theorists do not agree on whether the comparison standard is internal, another person, a group, or some combination (Hills, 1980). Research findings suggest only a tendency to compare with individuals of similar status and of whom the worker has some knowledge (Delafield, 1979).

Third, only broad generalizations about the consequences of inequity perceptions are possible. Adam's (1965) series of propositions and Lawler's (1971) model of the consequences of pay dissatisfaction identify outcome alternatives, but

with the possible exception of underpayment in incentive situations, useful predictions are not possible.

An alternative view of the equity determination process is offered by Elliot Jaques (1961) based on his research program conducted over the past 30 years in the United Kingdom. In essence Jaques argues that widely shared internal norms of fair payment exist for varying levels of work, and that individuals make pay equity judgments through intuitive comparison of actual pay vis-a-vis the norm. Most importantly, he asserts that fair pay norms are directly associated with job autonomy such that measures of relative "time span of discretion" provides a means to identify equitable payment scales. Thus to Jaques, equity judgments are not a result of social comparisons but stem from the job itself; the equitable payment criterion is absolute, not relative.

Jaques, like Adams, suggests that failure to provide equitable payment gives rise to discomfort (neurotic disequilibrium) which energizes the person to restore stability. However, he also specifies the relationship between degree of inequity and corresponding action, e.g., a 20% departure from equity is an "explosive" situation whereas a 10% variance crates only "strong feelings."

Jaques theory is of interest here not because of its empirical support (sparse) but because it is a 20th century variant of just wage doctrine which posits that equitable pay can and should be socially determined based on job content. Relative worth is independent of organization or labor

market, and centered primarily in tasks, not in personal characteristics.

In a general sense, the contributions of equity formulations to our understanding of pay differentials rest in their conceptual contradiction of market explanations. Human beings are characterized as equity seeking rather than income maximizing. The "push" of perceived inequity replaces the "pull" of net advantage. The focus is on reactions to differentials, implying that undesired consequences can be controlled by reflecting underlying norms of equity in administratively determined pay criteria. However, since operationalization of the variables has proved difficult, the practical value of equity frameworks has been limited to retrospective analyses of employee attitudes and behavior.

Demand Perspectives

The scope of the theoretical frameworks summarized in this section varies markedly. Institutional models of labor allocation and wage determination consider the firm as a set of internal labor markets exhibiting significant structural and motivational deviations from the external markets of the classical model. In contrast, discrimination theories center on employer motivation in competitive and/or administered markets, and stratification theory provides a societal outlook in explaining earnings differentials. In each instance pay structures are attributed at least in part to forces beyond the control of the worker.

Internal Labor Markets--The focus of internal labor market theorists is on the impact of employer policies and collective bargaining on human resource allocation and pay determination.[7] These administered labor markets come into being through employer tendencies to hire into a limited number of jobs and allocate people to the other jobs through promotion from within policies, a practice which serves to foster workforce stability and protect employer investment in training. Entry jobs are thus the first rung on "promotion ladders" and the entire ladder or job cluster constitutes an internal market within which training, allocating, and pricing is administratively determined. A typical administrative unit is comprised of a number of such clusters.

Technology plays a major role in internal labor market formulations. The degree of openness to external markets is substantially defined by the skill mix and degree of specialization required. Employers requiring only unskilled people may draw solely from the external market; high technology organizations are likely to severely restrict their entry points. Furthermore, technology constrains job structure design in terms of job content and interrelationships. Within these limitations, employers design vertical and horizontal job relationships which serve to control training costs through facilitating hiring and internal mobility practices. Consequently, ladders normally involve job content progressions, and administrative rules governing internal movement tend to discourage intercluster transfers where content

differences generate increased training costs. Thus broad groupings such as blue collar, clerical, and administrative mobility ladders are the norm, and functional clusters within each of these are common.

Pay allocation within mobility clusters normally is geared informally or formally (job evaluation plans) to job content differences. Overall wage levels within clusters may be affected by market rate influence on entry level and/or certain key jobs in the structure, but differentials are administratively determined or bargained. Job content differences reflecting incremental skill development are the most common job worth criteria since they support the mobility patterns desired by employers and are a major component in the formation of employee expectations. In mature organizations/industries, instances of anomalous "traditional" relationships are often found, but content dimensions are the predominant determinant of pay differentials.

Thurow's (1975) critique of marginal analysis employs an internal labor market framework that differs from the above in two respects significant to this discussion. First, industry and occupational wage differentials are viewed as stable and administered rather than subject to wage competition. Thurow asserts that applicants compete for jobs of varying attractiveness in terms of the cost of being trained to fill the job; the best trained occupy the highest positions in the applicant "labor queue". Employers react to external market supply fluctuations via changes in hiring

specifications (quality differentials) rather than wages. The relative wage level at which employers can compete for talent in the labor queue is, of course, a function of their productivity. Thus, the key constraint on entry rates is product market competition or ability to pay, not labor market competition.

Second, Thurow's explanation for internal pay differentials heavily emphasizes their motivational impact on employee effort and willingness to cooperate. Utilizing economic terminology he essentially paraphrases equity theory arguments regarding the importance of relative differentials to perceptions of "industrial justice." No new insights pertaining to the determinants of proportionality norms are offered, but Thurow emphasizes their historical-cultural origins and rigidity, suggesting that social shocks of great magnitude (e.g., World War II) are necessary to realign equity perceptions. Implied in this discussion is the idea that internal wage structure determination is more a process of discovering norms than rationally assessing job content differences.

The internal labor market perspectives presented by Thurow and others are of particular relevance to this discourse since they were developed from observed job worth determination practices, and the variables and mechanisms suggested directly contradict key contentions of neoclassical thought. The labor market is not conceptualized as a simple manpower bourse, but as numerous external and internal

markets highly differentiated in terms of supply and demand characteristics. Employer and worker motives other than profit/wage maximization are recognized. Work group rather than individual productivity is emphasized as a key determininant of employer wage decisions. Intraoccupational and external market differentials are primarily attributed to product market characteristics, and job content differences in conjunction with social norms are depicted as the primary determinant of internal wage structures. Finally, Thurow even challenges the most basic market concept by asserting that employer reaction to supply imbalances is more likely to take the form of adjustments to hiring criteria than wages.

In short, if the institutionalist views closely approximate reality, wage/income stratification is essentially man-made; pay differentials are a function of social norms operating within variable product market contexts. Consequently the ultimate sources of pay discrimination are those forces shaping contemporary norms of worth and employer decision criteria and processes governing entrance to product and internal markets. Alternative views of the latter will now be reviewed.

Discrimination Theory--Economists' theories of discrimination attempt to explain sex and race related pay differentials in terms of both employer motives and market mechanisms. Two viewpoints considering only motivational aspects are noted here first, followed by those positing some departure from the classical competitive model.

The first category of discrimination theory (Oaxaca, 1977) substitutes tastes or utilities for profits as the primary employer incentive. Employer preference for maintaining social or physical distance from the undesired group or person is thought to create utility in discriminatory behavior offsetting the theoretically diminished profitability due to restricted supply. In effect, the employer is portrayed as paying a premium to indulge in overt discrimination. Since the wages of the majority group are inflated, discrimination should disappear in the long run as competitors exploit this cost disadvantage - an implication which calls the validity of this model into question. The persistence of discrimination points strongly to the likelihood that discrimination is profitable.

The concept of statistical discrimination offers a more benign explanation of discrimination viewing it as a consequence of employer cost control motivation. This theory is simply an economic rationale for the phenomenon of stereotyping in selection decisions. Employers associate greater costs and/or lower probability of success with visible demographic characteristics such as race or sex. Given equal scores on "objective" predictors, employers will therefore reduce the pay offered members of these categories to offset perceived greater risks/costs or assign them to lower wage jobs. Since the former is now blatantly illegal, differential assignment is the likely outcome today if this model is correct.[8] Thus statistical discrimination is one

possible explanation for continued occupational segragation by sex or race.

Occupational segregation per se provides a basis for other theories of pay discrimination. The "overcrowding hypothesis" (Stevenson, 1978) explains male/female differentials under competitive labor and product market assumptions. Sex segregation (typically attributed to employer actions) restricts the occupational choices of females creating different male-female supply functions and equilibrium positions, resulting in lower marginal productivity for female occupations. The labor market, in this view, is segmented into noncompeting male and female groups within which wage competition prevails.

When the competitive labor market assumption is relaxed (monopsonistic employers) sex segregation is thought to generate differentials via employer exploitation of disparate male and female supply elasticities (Blau and Jusenius, 1976. Female occupations are presumed to be less wage elastic due to restricted work alternatives or geographic mobility, thus creating the opportunity for employers to pay females a lower wage than paid males of equal productivity.

The overcrowding and monopsonist explanations for the male-female pay gap utilize modified traditional economic frameworks. The former attributes wage determination to within segment employee wage competition, and the latter attributes differentials to employer exercise of economic power. A third category of segmentation theory explains the

same phenomenon more in political power than economic terms. Radical viewpoints (Stevenson, 1978) tend to cast the discussion into class conflict and conspiracy terms, explaining differentials as a product of employer "divide and conquer" and wage collusion tactics.

Common to all discrimination theories is the linkage of pay differentials to employer decisions. In this sense they are compatible with internal labor market perspectives. A variety of employer motives are hypothesized, ranging from sinister attempts at subjugation and exploitation to innocent desire to conform with social norms. All probably are true to some degree.

Labor segregation plays a central role in all of the discrimination formulations, either independently or in conjunction with market forces. Thus whether wage determination is considered in terms of market or administrative processes, the theories imply that occupational segregation and sex based pay differentials are two facets of a common problem. Furthermore, if a restricted range of job opportunities for female/minorities does translate into lower wages through any or all of the hypothesized market mechanisms, the "exchange value" for female positions reflects discrimination in job access.

As noted earlier, a major attraction of the concept of market determination of value is apparent objectivity in worth determination; wage levels and differentials are theoretically linked to productivity and serve to allocate

human resources to meet societal needs. However the preceding two sections offer compelling contradictions in which wages are primarily determined by employers and differentials reflecting socially unacceptable criteria are attributed to a variety of possible employer motives and mechanisms. The following section continues the general theme of social determination of income by reviewing a societal level explanation for economic stratification

Stratification Theory--While the preceding frameworks explain pay differentials in terms of employer/employee decisions and perceptions, stratification theory attempts to provide a societal level explanation. The perspective taken is that of functionalism which holds that social structures are best understood in terms of their function within society. The central issue in functional analysis is the way in which cultural items meet the needs of the greater collectivity, the society or social system. Of particular interest here is the phenomenon of social stratification, specifically, the functionalist explanation for why positions within a society receive differing shares of resources and prestige.[9]

Functional theories of stratification were developed in the late 30's and 40's by a number of sociologists but stratification theory is most closely associated with Kingsley Davis and Wilbert Moore. The basic principle of the "Davis and Moore Hypothesis" can be stated as follows: "The more functionally important positions in a system, and the

less available qualified personnel to fill those positions, the greater the inducements necessary to attract qualified personnel, and hence, the greater the rewards attached to these positions" (Turner and Maryanski, 1979:59). To Davis and Moore, all societies face the problems of manning positions which vary in terms of functional importance, and motivating people to perform in them. Differential rewards serve this function in their view, while also giving rise to stratification systems. Thus stratification becomes the mechanism by which societies allocate people to necessary positions, and the wealth associated with an occupation is symbolic of its social ranking.

Two stratification (reward differential) criteria are suggested by Davis and Moore. First, social ranking varies with the functional importance of a position to society, i.e., the existence of some generalized criterion of worth to society is hypothesized. However, importance is not operationally defined. The two sub-dimensions suggested, the uniqueness of a position and the number of positions dependent upon it, are open to varying interpretations and applicable only to intra-organizational assessments. They are an inadequate basis for evaluating the relative importance of different societal components. Thus inability to identify and define common criteria limits the stratification hypothesis as it has all attempts to determine or explain relative worth.

Davis and Moore also noted that rewards and social ranking are affected by the scarcity of personnel possessing the required talents and skills, thus the role of supply/demand in the determination of relative wealth and status is recognized. However, they did not define the nature of the scarcity-functional importance relationship; the impact of various combinations on status/award rankings is not specified.

Criticisms of stratification theory are numerous (Abramhamson, 1978), but two deserve particular attention with respect to male-female pay differentials. First, and perhaps most fundamental, the theory is logically flawed. Functionalists fall into a tautological trap by assigning both cause and consequence of reward differentials to societal needs for differential role assignments. Stratification is caused by the need to fill functionally important roles; the consequence of stratification if fulfilled social system needs. Circular logic such as this offers no explanations for low female occupation status/rewards other than the view that they must be low in functional importance given their current status in society. Thus it is not surprising that some women's rights advocates summarily dismiss functionalism as the male sociologists' rationale for maintaining the status quo (Morris, 1978).

Stratification theory is also seriously deficient in terms of its basic functionalist assumption. Whether extrinsic rewards provide the only motive for mobility is highly questionable; numerous other motivational assumptions

are potentially valid. In addition, access to mobility channels may not be equal, thus the consequences to society may be less than positive. To the degree factors other than abilities (sex, race, family resources, etc.) determine access to positions, stratification must be dysfunctional to society. In other words, stratification may be functional only for certain groups in a society.

Stratification theory can be characterized as a societal level "use value" approach to the explanation of occupational worth, in which value is considered proportional to position importance in meeting collective needs. Because the theory specifies nether distribution criteria nor mechanisms, it has no practical value. It does, however, evidence a continuing desire to explaing differentials in social rather than market terms.

Conclusions

One conclusion to be drawn from this review of selected theories is inescapable. The determinants of relative value in our society are multiple. Figure 1 summarizes this fact by depicting the two basic approaches to explaining pay differentials. On the one hand, pay structures are viewed primarily as a dependent variable constrained or determined by network of environmental and organizational forces. On the other hand, structures are both independent and dependent, subject to administrative decisions attempting to influence pay consequences. Both approaches doubtless are

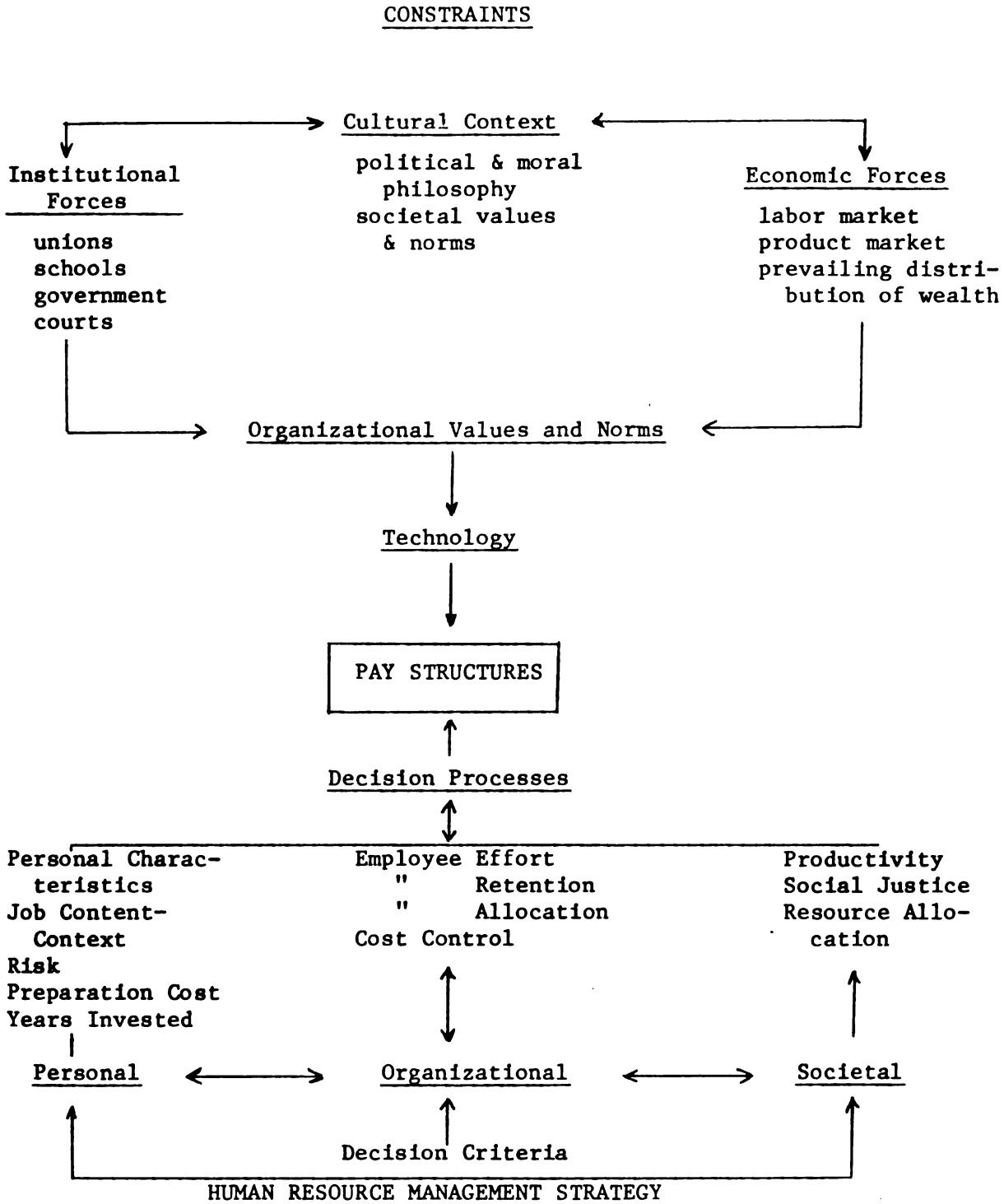


Figure 1. Pay Structure Determinants

correct, with their relative importance varying substantially across situations.

A second, and critical, conclusion for the present study is the importance of societal, organizational, and individual values for wage determination. Productivity and utility values are subsumed in the competitive market model; competing notions of internal equity based on content or career investment are often advanced by workers/unions; employers frequently stress efficiency and progression incentives, and minimum wage legislation reflects a need or adequacy criterion. Each of these value dimensions in turn represents a potential basis for equity judgments. While the importance of fair distribution is emphasized by the behavioral theorists, the two central questions with respect to distributive justice are ignored: (1) The process issue - how should values be determined, and (2) the content issue - which of the potentially relevant criteria should be considered in determining the "fair" distribution. Philosophical approaches to these two questions are briefly reviewed in the final section of this chapter.

THE NORMATIVE PROBLEM OF JUSTICE

Centuries of debate and accumulated wisdom have not served to resolve or reduce the issues relating to determining the justice of income distribution. If anything, the

correct, with their relative importance varying substantially across situations.

A second, and critical, conclusion for the present study is the importance of societal, organizational, and individual values for wage determination. Productivity and utility values are subsumed in the competitive market model; competing notions of internal equity based on content or career investment are often advanced by workers/unions; employers frequently stress efficiency and progression incentives, and minimum wage legislation reflects a need or adequacy criterion. Each of these value dimensions in turn represents a potential basis for equity judgments. While the importance of fair distribution is emphasized by the behavioral theorists, the two central questions with respect to distributive justice are ignored: (1) The process issue - how should values be determined, and (2) the content issue - which of the potentially relevant criteria should be considered in determining the "fair" distribution. Philosophical approaches to these two questions are briefly reviewed in the final section of this chapter.

THE NORMATIVE PROBLEM OF JUSTICE

Centuries of debate and accumulated wisdom have not served to resolve or reduce the issues relating to determining the justice of income distribution. If anything, the

intensity of disputes is increasing. Greater realization now exists that relative income determination is not simply a function of uncontrollable forces. Furthermore cultural pluralism and increasing uncertainty about the objectivity of traditional values fuel the debate over distributive formulas as various societal factions proclaim the justice of their cause and attack the injustice of opposing views.

Philosophers traditionally have attempted to provide answers to this unsettled problem of justice. As noted earlier, Aristotle's general principle of distributive justice utilized a proportional equality formula; scholatics appealed to higher authority basing just wage principles on their view of the supernatural and natural order, and naturalists substituted the authority of the market (natural law) for that of the state and/or church in their explanation of the just or market rate. Socialists, utopians, and other reformers offered yet other principles of just distribution. The objective of this discussion is not to compare and contrast these historical and contemporary formulas, but to identify the central issues of distributive justice in order to highlight the philosophical facets of the comparable worth debate.

A Definition

Justice is a concept with a bewildering variety of uses and meanings, sometimes relating specifically to treatment under the law, at other times to the general ideas of fairness and/or equality in daily interactions with others.

Common to all of the uses is a concern with societal relationships. Justice is a regulative principle, a moral rule of action prescribing the treatment of others by which they are "accorded their due." It is a legalistic concept defining the necessary quid pro quos in human relationships. Justice is often contrasted with the other great moral principle, beneficence, which goes beyond bare minimalities to prescribe obligations arising out of altruism or the belief that men should help one another. The line of demarcation between the two is unclear; one continues where the other leaves off. But they may conflict. In situations involving property, for example, (such as the distribution of rewards) the just action may fall short of that required by beneficence, or the just action may be difficult to ascertain while the prescription under beneficence is clear.

The classical dictum of justice, "giving to each his due," incorporates the fields of application often referred to as distributive and retributive justice (Sabine, 1956). The just social order imposes restraints or penalties on individuals violating rights of others (retribution). Similary, the just society gives to each his "fair share" of the goods or evils (distribution). It is the latter with which we are primarily concerned here.

The problem of distributive justice is one of adjusting competing claims in a context of scarcity and conflicting values. Individuals press for their rights to a commodity based on their duties, need, contributions and other such

criteria of moral worth. Justice requires striking a proper balance under varying circumstances while avoiding arbitrary (irrelevant) distinctions. For example, the just distribution of family resources to children in one instance may be one of equality; in another, particular needs of a member may justify inequality. The dominant value underlying both of these distributions may be equality of opportunity but in the second case one child requires more resources to achieve that end. If, however, efficiency of resource utilization is also valued, a different distribution might be required based on relative capability to profit from the resource. In this situation, as is frequently the case, values are in conflict and distributive justice requires prioritizing of values before considering claims and conditions.

Value Conflicts

Justice consists of giving each person his due in the context of a particular relationship. Human relationships, however, fall into two classes which generate different obligations on the part of the participants (Baldwin, 1966). First, relationships exist between individuals as fellow, equally important human beings. On this level, the equality demands of justice are paramount. Equality of consideration as a human being creates the obligation to claim for ourselves, and to allow others, equal liberty to live and do as they please. Equal liberty clearly encompasses obligations such as equal political rights, or equality of opportunity in economic affairs. The liberty to live our own lives as

responsible persons is perhaps the fundamental right of human beings. It is also the major force creating the second class of relationships, those involveing morally relevant differences.

All men may be created equal, but the situation changes radically once they begin to act. People engender merit and demerit, and these differences give rise to demands for equity, for treatment in proportion to the good or bad behavior. Inequality of effort of lawful behavior demands inequality of reward or retribution by the wrongdoer. Thus equity requires discrimination, in the sense that morally relevant differences in actions must be reflected in treatment. The essence of equity, therefore, is not equal but differential or proportional treatment.

Differential treatment may also serve the goal of equality. As in the earlier example, all the children may justifiably expect to attain a reasonable level of physical health, but the fulfillment of that expectation may require unequal distribution of family resources to the children if the need of one is greater. This in turn may limit other members' choices of pursuits, i.e., equitable treatment may enhance equal rights for the one but restrict liberty or economic opportunity for others. Similarly within organizations, differential treatment of individuals or groups undertaken to serve the goal of equality of opportunity limits the options of others.

Interrelationships among basic goals become even more complex as additional values are taken into account. For example, distributions intended to maximize economic progress are often viewed as incompatible with equality consideratons. The problem of defining the just distribution is thus best characterized at its most fundamental level as one of reconciling values. Dissonance is unavoidable, particularly between equality and equity. "Many sit down at the table of society who do not deserve to be there, and many eat from it who have not made a contribution" (Boulding, 1962:83).

It is the equality-equity tension to which Gardner (1961) is referring when he mentions the contesting philosophies of equalitarianism and competitive performance. Overemphasis on equality of treatment can result in mediocrity due to lack of incentives for individual performance. On the other hand, emphasis on individual performance can bring about excellence but it also can lead to exploitation of those less able to compete. The fact that the two values are in conflict does not, in Gardner's view, mean one has to choose. The tension between equal and differential treatment is simply a fact of life which will never, and should never, be resolved. Competing claims must constantly be held in balance based on clear understanding of the implications of each in a particular situation.

Value conflicts actually come into play at two levels in pay distribution. First, value judgments shape one's vision of the desired "just" distribution, the goal or end to be

attained. Beliefs regarding such features as necessary minimums, allocations to "nonproductive" members, and extent of differentials are rooted in what might be called ultimate norms, beliefs concerning the nature of man and social relationships. The preservation of human dignity, maximum individual freedom, and economic progress, in varying combinations, comprise the core set of goals underlying most notions of just pay. Second, operational or proximate distribution norms also are a constant source of conflict. Such disputes are predictable since these criteria are means to differing ends, i.e., they derive their validity from ultimate norms. However, their relative effectiveness for achieving a specified goal is also a matter of contention. The venerable merit pay debate evidences this fact (impact of merit versus equality allocations on effort). In addition, over a period of time some pay criteria appear to take on value independent of ultimate norms. Strong commitment to payment relative to various merit criteria, organizational level, credentials, etc., often develops in the absence of clear linkage to the fundamental values. Thus judgments among competing claims often require the consideration of not only ends but prevailing values regarding the means. Competing beliefs regarding ultimate and proximate norms and their relationships are the stuff from which just or equitable pay allocaiton decisions must be made.

The complexity of fair distribution decisions is even further increased by the fact that the ultimate norms often

incorporate a number of dimensions generating multiple and conflicting means-ends relationships. Liberty, for example, can be defined and assessed in terms of the relative freedom of choice to pursue economic self interest. From this perspective utility and efficiency distribution criteria are most appropriate to the preservation of freedom since personal actions and desires govern economic activity. However, one of the primary criticisms of economic utilitarianism is that of its restriction on the freedom of those with low incomes (Bowie, 1971). The utilitarian view considers only consumption related freedom, and to the poor, consumer sovereignty is a mockery. Worker choice of job, hours, location, etc. are also elements of personal liberty affected by distribution principles. Consequently egalitarians argue for competing pay criteria under the same banner of liberty, albeit a different dimension.

Operational norms reflect a similar dimensionality problem. While productivity criteria are often stressed, true productivity measures are seldom possible. Thus components such as ability or skill, effort, credentials, and experience (seniority) are employed to justify distributions. The relationship between these pay norms and output is often tenuous. Effort can be in vain or misguided as well as productive; qualifications are often not relevant; abilities must be properly applied, and the moral worth of experience is as frequently formulated in loyalty or life investment terms as it is linked to productivity. Attempts to operationalize a

"need" criterion have encountered even greater difficulties given its obvious subjectivity and numerous possible dimensions.

In the absence of a value consensus the problem of distributive justice is far too complex to be resolved by a simple principle. Aristotle recognized this long ago when he noted that:

> . . .all men argue that what is just in distribution must be according to merit in some sense, though they do not all specify the same sort of merit, but democrats identify it with the status of freement, supporters of oligarchy with wealth (or with noble birth), and supporters of aristocracy with excellence. (Nicomachean Ethics: V.3)

Aristotle may also have been implying in this passage that the search for principles of justice in distribution will always be frustrated by that most powerful of ultimate norms, self interest. The value laden nature of distributive justice has led many contemporary moral philosophers to the conclusion that the issue is not amenable to the canons of logic. Others attempt to develop what appear to be hopelessly complex contingency theories (Bowie, 1971) or focus on the characteristics of moral judgment processes. The latter approach is particularly relevant here.

Justice as Process

Some philosophers argue that universal maxims are of more use as questions than answers. From this point of view distributive justice is defined in terms of its process, the depth of the analysis of a particular situation and the balance exhibited between equality of consideration and

recognition of different needs and deserts. Thus Frankena (1980)[10] emphasizes rationality and respect or concern for others as the basic characteristics of morally responsible judgments, and Tsanoff (1956:16) defines justice as the ". . .the principle of thorough and balanced recognition of all the factors and values involved in a complex personal situation, as opposed to any abstractly rigid or onesided adjudication."

The process or "fair game" approach to justice attempts to establish the basic rules for participating in economic activity which, if followed, would justify any outcome. Rawls' (1971) theory of the justice of a market system, for example, posits the development of just social institutuions of distribution abiding by the two principles that justice (1) requires equal right to the most extensive liberties for all individuals, but (2) allows inequalities in social position and wealth if income possibilities are open to all and if the inequality serves to improve the prospects of the least advantaged. Within these tests, a wide variety of potentially equitable distributions could occur.

Defenders of market determination of worth often adopt a fair game approach to justice stressing market impartiality and responsiveness to aggregated individual choices. However, the market mechanism does not measure up to either of Rawls' principles. Inequality of opportunity is the norm, and losers in the economic game tend to become increasingly disadvantaged. More fundamentally, the notion of a fair game

is fatally flawed. Rawls assumes that members of society can, at some point, determine the relevant bases for inequality and embody them in institutional rules. Similarly, the market argument assumes that at some previous point in time the distribution was just, because market distribution of income reflects demand patterns established by previous distributions. Even if the rules of the game are fair, in market economies an initial distribution of resources must be certified as equitable for any subsequent distribution to be so characterized. Alternatively stated, a fair game approach to justice cannot escape the necessity for an initial set of value judgments!

Conclusions

Exploration of the concept of distributive justice serves to emphasize the philosophical dimensions of the comparable worth debate. From this perspective the normative nature of wage-setting is more clearly perceived, i.e., pay allocation is a subset of the larger social issue of the justice of income and wealth distribution. The array of pay rates within an organization may not always be a question of social justice in the normal sense of that phrase, but the cogent issues in equity determination are identical at the societal and organizational levels. Thus the lessons learned in the search for principles of justice should directly apply to organizational efforts to establish relative worth.

Four general conclusions for worth determination are readily drawn from the philosophical formulation of justice

as balancing equality of consideration with recognition of relevant differences. First, multiple criteria are required to recognize both equality and equity based claims. In this respect modern "just wage" decisions differ dramatically from their historical predecessors since contemporary belief in the equality of man places the burden of proof on arguments for differential treatment. Equality considerations tend to be the most morally compelling. Thus the right of all members to some minimum share in the economic pie is an essential principle in any equity specification. The challenge to organizations is to define what this means in their particular situation, and then to spell out the proximate criteria for differential allocations and distribution assessment.

Second, justice is centered as much in the process as in the outcome. Balanced consideration of all claims and consequences can only be ensured through formal and rational decision processes. Continued attention to design and operational characteristics of pay determination procedures is required for wage setting to be a "fair game."

Third, norms or principles of worth are perceptual and dynamic rather than objective and static. Sufficient value system uniformity within organizations may exist at a point in time to implement a set of generally accepted pay criteria. However, such a "working consensus" is subject to challenges stemming from basic individual differences regarding ultimate and proximate norms, resource fluctuations, and a

host of environmental forces.[11] Consequently value system hierarchies and composition are constatly evolving, and wage determination justice requires periodic formal reassessment of the governing value structure.

Finally, and most importantly, the philosophical discussions of justice dispel any notion that social determination of worth can be avoided. Labor market determination of worth is no less a social decision than installation of a national job evaluation or incomes policy. The central issue is one of degree or intrusiveness of direct decisions regarding relative worth vis-a-vis the criteria reflected by an imperfect market. As noted earlier, national policy in the United States has stressed minimum governmental interference but administered structures are the rule in primary labor markets. Comparable worth proponents directly challenge the fairness of both market and administrative criteria and the processes by which they are applied. In short, they assert that criteria are irrelevant and/or deficient and the game is not fair. Chapter two addresses these claims through examination of wage setting mechanisms relative to the criticisms.

CHAPTER NOTES

1. The discussion of value concepts draws heavily upon Heilbroner, R. L. The Worldly Philosophers (New York: Simon and Schuster, 1978) and Roll, E., A History of Economic Thought (London: Faber and Faber, 1973).

2. The intense interest in Japanese compensation management systems which eschew individual productivity criteria, and the emphasis on the worker as the ultimate source of value in the recent papal encyclical (Laborem Exercens) are two indicators of increasing concern with the social impact of wage determination processes (in addition to comparable worth).

3. Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations (New York: Random House, 1937).

4. See for example, Allan M. Cartter, Theory of Wages and Employment (Homewood, Ill.: Richard D. Irwin, 1959) or J. R. Hicks, The Theory of Wages (New York: McMillan & Co., Ltd, 1963).

5. See for example, Jacob Mincer, "The Distribution of Labor Incomes: A Survey, with Special Reference to the Human Capital Approach," Journal of Economic Literature, Vol. 8, No. 6 (1970) pp. 1-26.

6. Stieber's (1959) discussion of the steel industry wage structure provides an excellent example of the problem and its management in one context.

7. While numerous expositions of internal labor market theory are available, (e.g., P. Doeringer and M. Piore, Internal Labor Markets and Manpower Analysis. Lexington, Mass.: Heath, 1975) this section is most directly influenced by E. Robert Livernash, "The Internal Wage Structure" and Peter B. Doeringer, "The Structure of Internal Labor Markets" as reprinted in Thomas A. Mahoney, Compensation and Reward Perspectives (Homewood, Ill.: R. D. Irwin, Inc., 1979).

8. Note the consistency of this view with Thurow's job competition model.

9. This portion of the paper reflects the views of M. Abrahamson (1978) and Turner and Maryanski (1979).

10. William K, Frankena, Thinking About Morality, 1980 as cited by K. E. Goodpaster and J. B. Matthews, "Can a Corporation Have a Conscience?", Harvard Business Review (Jan.-Feb. 1982) p. 134.

11. Both Selznick (1969) and Rohrbaugh, McClelland, and Quinn (1980) report on the extent and correlates of individual differences in equality v. utility definitions of justice.

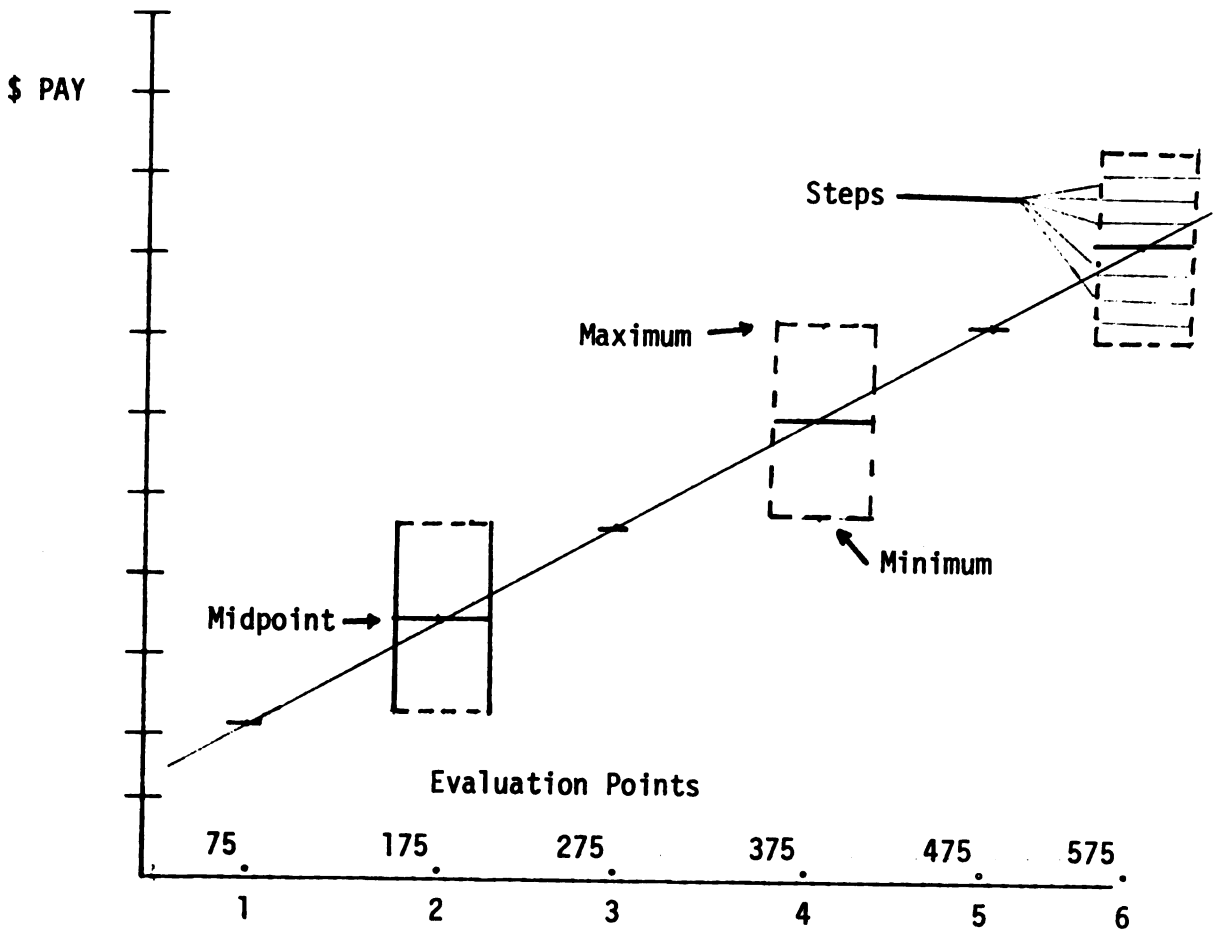
CHAPTER 2

Pay Administration, Job Evaluation, and Comparable Worth

In this chapter the focus shifts from theory and philosophy to reality. An understanding of the basic characteristics of tools and procedures employed in making pay decisions is prerequisite to any analysis of wage-setting criticisms or comparable worth proposals. Therefore, the state of the art in job worth decision making is briefly examined to establish the how and why of relative worth decision. The concept of comparable worth job evaluation is then examined in terms of its implications for compensation administration practice, and the critical issues in the comparable worth debate are identified.

PAY STRUCTURE DECISIONS

Wage or salary structures are the most basic administrative tool in formal pay administration programs. While they vary in sophistication, pay structures all serve the same function of defining the relationship between jobs and pay (Figure 2). Thus two components, a system for hierarchically ordering jobs (job evaluation) and a dollar conversion equation, are necessary to build a pay structure. Job evaluation system outcomes (job structures) are usually point distributions and/or categorical assignments, with the latter in



JOB EVALUATION OUTCOMES

Figure 2. Pay Structure Concepts

the form of classification or pay grade levels the most common. Pay equations may involve single rates as indicated by pay grades 1, 3, and 5 in Figure 2, or jobs may equate to a range of potential values as illustrated in classifications 2, 4, and 6. In the latter case assignments to specific values or steps within pay ranges (see grade 6) are determined by personal characteristics of job incumbents such as merit or seniority. Hence, pay structures provide a means to recognize judgments regarding both job worth and person worth (individual differences) in establishing actual pay for a particular person.

The distinction between job and personal criteria of worth is fundamental to the notions of job evaluation and comparable worth and is thus employed throughout this discussion. Yet its limitations should be recognized. Job evaluation is based on the concept that certain variable characteristics, contextual aspects, and/or demands of jobs directly reflect their worth to the organization. These factors are viewed as independent of any particular job incumbent. However, in certain types of jobs, the characteristics, context and demands are constants; the only variables apparently affecting worth to the organization are personal. In such situations job and person worth concepts merge and evaluation systems may take the form of maturity curves (such as among engineers) or rank-in-person approaches as often found in the professions (professors, researchers, lawyers).

Job Evaluation Methods

Job evaluation is a generic label for a variety of formal processes employed by organizations to differentiate among jobs in terms of the contributions or value of the jobs to the organization. Evaluation techniques range from simple rankings of whole jobs to complex systems involving multiple assessments of job content and context components yielding quantitative scores. Whatever their particular form, job evaluation procedures involve judgments, both of job content differences and the relative importance of the content dimensions to the organization.

Formal job evaluation procedures were not popularized until World War II when government regulation provided a strong stimulus. General wage increases were rigidly controlled by the National War Labor Board but increases were permitted to correct demonstrated inequities. Evaluation techniques were often adopted to provide the necessary documentation for such increases. Further, by introducing rationality and stability into previously chaotic wage setting processes, job evaluation systems served to control disruptive wage inequity disputes and ensure continuity of war material production. Thus, Regional War Labor Boards promoted the installation of evaluation systems through liberal wage structure revision allowances in conjunction with the introduction of job evaluation systems (Slichter, Healy, Livernash, 1960).

Job evaluation systems were introduced as a supplement, not an alternative to market determination of job worth. Inconsistencies due to individually negotiated hire rates and continuous bargaining over "inequities" placed pressure on companies and unions to introduce logic into pay structure determination. The use of job content ratings or rankings on factors such as skill and responsibility provided a means to systematize judgments regarding the worth of jobs in many organizations. Market values defined the worth of a frame-work of key jobs or "benchmark" jobs, but non-key job relationships were established by the evaluation system. Conflicts between key job market values and evaluation determinations of worth were informally "adjusted." Thus it is that Chamberlain (1965:407) described the role of job evaluation in pay determination as "(T)idying up after the market."

Job evaluation plans may be unique to an enterprise, employed industry-wide, or broadly adopted by a cross section of organizations. Whatever their particular form they all can be differentiated in terms of two basic characteristics. First, evaluation methods vary in terms of the degree to which the job is factored into components for assessment purposes. Simple ranking systems assess the "whole job" on some global dimension of worth. Job component questionnaires represent the other extreme utilizing responses to hundreds of items to generate job worth scores. The more commonly employed classification, factor comparison and point-factor

methods typically attempt to differentiate among jobs on three to twelve job dimensions or compensable factors.

The second basic characteristic of job evaluation methods is the nature of the job worth assessment standard. In ranking and factor comparison approaches other jobs provide the referent; job worth is clearly defined in relative terms. Monetary or other measurement units are used in factor comparison systems to express judgments regarding the degree of inter-job differences on factors, but the process remains a series of job-to-job comparisons. In contrast, jobs are matched with categorical definitions in classification schemes and scored on rating scales in point factor and questionnaire methods. A job can theoretically be evaluated without references to other jobs since fixed verbal standards ostensibly replace other jobs as evaluation criteria. The use of such standards is thought to facilitate validity and reliability of measurement (Belcher, 1974). However, this distinction in type of standard is more apparent than real; traditional evaluation methods are all essentially based on job-to-job comparisons. This fact is readily apparent from a review of evaluation plan development procedures.

Ranking Systems--Simple or whole job ranking systems require no development effort, per se. One or more judges compare all the jobs to one another--usually in terms of a composite job worth concept incorporating the notions of importance and difficulty--and arrange them in hierarchical order. Where multiple judges are used, differences are resolved

through consensus or averaging. Thoughtfully generated rankings may accurately mirror the accepted job structure but the obvious subjectivity and lack of documentation limit the capability of rankings to rationalize or justify differentials.

Factor comparison systems also involve job ranking procedures, but multiple rankings on prespecified criteria (compensable factors) are developed for each job, and the final evaluated worth of the job is expressed quantitatively. The essential tool in factor comparison approaches is the job comparison scale, an array of benchmark job rankings on each compensable factor along a measurement scale (dollars, points, percent). Figure 3 illustrates such a scale for a four factor system with four benchmark jobs (A,B,C,D). Comparison scales are developed by first determining the "true" worth of benchmark jobs (market value or current rate) and then subjectively allocating portions of that worth to each compensable factor in proportion to the contribution of the factor to total job worth. Job C in Figure 3, for example, has a total worth of 1050 units allocated in amounts of 400, 200, 100, and 350 to the physical demands, skill, responsibility, and working conditions factors respectively. In this manner different values for each factor are identified for all benchmark jobs, thus generating a comparison framework. The remaining jobs in the organization are ranked relative to the benchmarks on each factor, and job worth is operationally defined as the sum of factor values. Thus

Measurement Units	Factors			
	Physical	Skill	Responsibility	Working Conditions
400	Job C	Job A		
-				
-			Job B	Job C
-	Job B		Job A	
300				
-		Job B		
-				Job B
-				
200		Job C		
-				
-	Job A			Job D
-				
100		Job D	Job C	
-				Job A
-	Job D		Job D	
-				

Figure 3. Job Comparison Scale

multiple job-to-job rankings are converted into quantitative "measures" of individual job worth.

Classification--Classification approaches to job evaluation are similar to filing systems. Job characteristics are compared to salary grade (file) definitions and where the match is closest the job is filed or assigned to the salary grade.

Classification systems start with a determination of the number of levels or grades of job worth in the organization. Each of these salary grades is then defined in terms of various compensable factors in an attempt to describe a progression of worth. Since the classification system usually must accommodate a wide variety of jobs, the grade definitions tend to become broad generalized statements permitting considerable latitude in classification judgments. The common solution to this problem is to clarify the definitions by including examples of typical, i.e., benchmark positions associated with each grade. Where classification descriptions do not explicitly contain benchmark referents, they are associated with labor grades through practice. Thus classification systems, in operation, are a form of job-to-job ranking in which the number of levels is constrained by the predetermined set of grades.

Point-Factor Methods--The distinguishing characteristic of point plans is the use of fixed rating scales to measure the degree to which compensable factors pertain to specific jobs. The sum of the scores on each factor

generates a job worth point total which determines the job's relative position in the salary structure.

Compensable factor ratings of point plans appear to provide criteria of job worth independent of market or traditional determinants of worth. This, in fact, would be the case if the rating schemes were developed without reference to market or traditional pay patterns. However, normal point plan development procedure involves adjusting factor content and/or scale scores to obtain correspondence between evaluation results and desired (benchmark) job relationships (Schwab, 1980). Therefore, the true criteria of relative worth are the determinants of benchmark relationships, not the factor measures. Point factor evaluation systems merely model or rationalize economic or traditional forces in compensable factor terms acceptable to organization members, and point ratings are nothing more than indirect comparisons with the benchmark structure, i.e., indirect job-to-job comparisons.

Job component systems (McCormick, 1979; Gomez-Mejia, 1979) are of the same genre as point systems when used for job evaluation, but much more inclusive in the task and/or worker behavior information collected. The best known and most heavily researched of these systems, the Position Analysis Questionnaire (McCormick, Jeanneret, and Mecham, 1972) assesses jobs on 187 components, but only a small number of these components function as compensable factors in typical job evaluation applications of the PAQ. An accepted wage

structure is regressed on the 187 item scores to identify the best set of predictors (McCormick, Mecham, Jeanneret, 1977). The equation derived is then used to generate predicted salaries from subsequent PAQ ratings. Thus the components, like point system factors, are used to capture or represent policy or market determinants of pay differentials.[2]

In summary, traditional job evaluation systems serve two functions. Evaluation plan criteria and processes are instrumental to the achievement of perceived equity in pay structures through rationalizing differentials. In addition, evaluation procedures are designed to link the internal job structure to the external market. They do not attempt to measure the job in any absolute sense, but merely facilitate the pay differential decision making process.

Pay Equations

Job evaluation processes are not the sole determinant of pay differentials. Administrative considerations also play a major role in the selection of relevant market values and in the manner in which evaluation results are clustered and equated to dollar values. These wage setting practices also have significant implications for comparable worth job evaluation.

The preceding discussion of job evaluation treated market values of benchmark jobs as a given, whereas earlier discussion of market model imperfections indicated that employers can choose from a range of values. Rates for a particular benchmark job will vary by product market (industry) and

locality as well as by candidate availability. Wage and salary surveys frequently identify rate differentials on common jobs as high as 50%. The more inclusive the survey, the greater the range. Thus managements can exercise considerable latitude in establishing their definition of the market.

Different markets are often associated with separate segments of organizational pay structures. Wages do not always move simultaneously or in equal increments across a broad job population. Secretarial position rates, for example, are primarily affected by local labor market forces whereas the design engineer market is national in scope. Further, the criteria perceived as relevant to pay differentials differ by occupational cluster. Physical effort and work conditions may be considered important to factory operatives, but responsibilities and expertise demands are more likely critical to professional/managerial personnel. Therefore, most larger organizations employ different evaluation plans for clerical, professional, sales and blue collar job groups. Different benchmarks, different labor markets, and different compensable factors are utilized in the design of these systems.

While multiple wage-setting systems may enhance employee acceptance of differentials within a broad functional group, they also affect the magnitude of job differentials across systems. Employers can adopt different pricing strategies for the separate systems. Wage level policy may be to be

competitive with the average local market rate for clerical personnel, with the 75th percentile for skilled trades, and with the top 10% of the national market for upper managers. Legitimate reasons may exist for the pursuit of such a policy. Nevertheless, the net result can be an increase in between system pay differentials.

Similar policy decisions affect within system differentials. For example, market values in the upper reaches of the range may be equated to higher level benchmark jobs and below average rates assigned to the lower positions. Definitions of market could even vary at each anchor point in the structure, depending upon management pay objectives.[3]

Differences within a system are also affected by interpretation and treatment of job evaluation results. The basic issue is the determination of the number of distinct job levels represented by the evaluation "scores." It was noted that classification system design begins with a judgment regarding the appropriate number of job levels. Quantitative evaluation systems require the same decision. Figure 4 illustrates some of the considerations.

One answer to the job level question is to allow every point score on the X axis a unique pay range. Jobs A, B, and C, in Figure 4 represent this approach. A one level change on any single compensable factor rating will change the pay range of the job. The conversion equation (point to dollar ratio) remains very similar for jobs in the same general area of the curve, varying only with the slope of the pay line

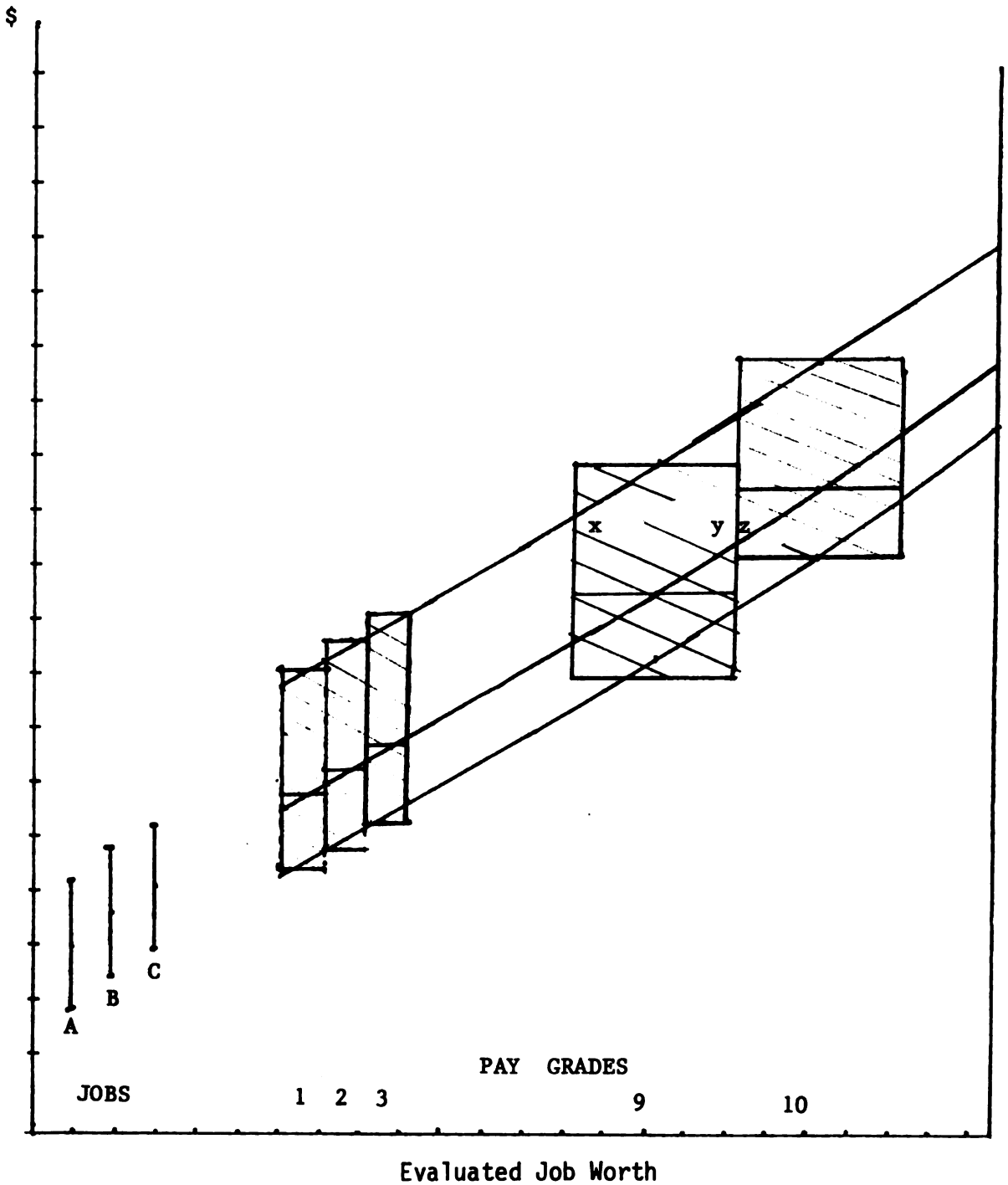


Figure 4. Pay Grade Alternatives

(which is typically non-linear due to the common practice of maintaining percentage differentials). This solution treats evaluation scores as interval scale data suggesting a degree of evaluation plan sensitivity to job content variance which would be difficult to justify in the face of challenges to minor evaluation differences. Conversion equations of this type thus place great pressure on the evaluation system in the form of requests for reevaluation.

Labor grades 1-3 and 9-10 represent two versions of the traditional pay grade structure approach to establishing pay differentials. The more inclusive grades 9 and 10 provide larger pay differentials between grades but are much less sensitive to evaluated differences than are grades 1-3. Note that the job worth difference between jobs X and Y is greater than between Y and Z, yet the former pair is graded equally and a significant grade differential exists between the latter. Grades 1-3, on the other hand, probably reflect insignificant differences. Neither of these approaches is necessarily right or wrong; they simply reflect different philosophies regarding pay progression, promotional incentives, cost control and other administrative concerns. The essential point here is that pay differentials are not only a function of evaluation judgments and market decisions, but that the number of job levels recognized by the organization also affects job worth differentials.

The purpose of this brief review of traditional wage determination practice was to establish the frame of reference

for contrasting comparable worth job evaluation proposals with established practice. In the language of the preceding chapter, it was noted that the basic framework of job worth determination is based on exchange value as supplemented by procedures distinguishing among jobs in use value to the organization. A key contention in the following discussion of the comparable worth movement is that advocated changes in wage setting are not simply job evaluation procedural reforms, they are attempts to increase the role of socially determined norms of proportionality in income distribution (use value) at the expense of market determination of worth

COMPARABLE WORTH JOB EVALUATION

The belief that male-female earnings differentials reflect unjust discrimination in wage determination is neither new nor limited to the United States. Equal pay historically has been one of the basic goals of the women's rights movement although in the first two decades of this century efforts were concentrated primarily on voting rights. Support for the equal pay principle grew slowly through the 30's while the Western world suffered through a prolonged depression, but during World War II the number of proponents sharply increased as large numbers of women successfully performed in non-traditional roles. In fact, wartime conditions led to the first recognition of the equal pay principle in U.S. policy with the 1942 issuance of General Order Number 16

authorizing equal pay adjustments for females without approval of the Board (NWLB Report No. 32, 1945).

From the immediate post war period to the present, the equal pay principle has gathered support to the point where it is now almost universally recognized in some form in Western industrialized nations. The West German parliament in 1949 passed a protocol referring to the equal pay principal and in 1955 the Federal Labour Court abolished the "women's deduction" in rate structures. In 1951 the International Labor Organization adopted Convention No. 100, the "principal of equal remuneration for men and women workers for work of equal value" (ILO, 1951). Article 119 of the Treaty of Rome, which established the European Economic Community, initially called for equal pay for equal work and was subsequently revised to agree with the ILO "equal value" phrasing (Sullerot, 1975). Consequently all of the original (France, W. Germany, Italy, Luxembourg, Belgium, Netherlands) EEC members had some form of legal instrument in place by the late 60's guaranteeing women workers equal rights and the newer members (Britain, Denmark, Ireland) followed suit in the mid-70's. By 1979, at least 17 European countries, in addition to the United States, Canada, and New Zealand, had implemented equal pay policy through legislation, court rulings, constitutional provision and/or national bargaining agreements (Ratner, 1979).

Definitions of equal work vary significantly among countries. In some cases, only jobs actually performed by both

men and women concurrently are covered; others, including the U.S., apply a "substantially equal" standard allowing minor variability in job content. The UK Acts of 1970 and 1974 employ both a "like work" and an "equivalent work" standard, with the latter defined as work of equal value under a job evaluation scheme (Glucklich, Hall, Povall, Snell, 1977). The 1978 Canadian Human Rights Act, the most recent legislation in this area, specifies only a broad equal value standard. In general, legislation and rulings over the past thirty years reflect a trend toward less restrictive definitions of equal work, thus potentially broadening the scope of protection.

Apparent diversity from country to country in equal work standards diminishes considerably when enforcement practices and implementation provisions are reviewed. In the majority of countries, the effective standard is a narrow "equal work" concept. In many cases wage structures can only be revised through collective agreements, and reluctance by one or both parties has effectively forestalled changes other than those necessary to rectify the most blatant discriminatory practices. Many countries provide recourse only through labor court systems, some of which may not consider individual complaints or require individual resources beyond those of most grievants. However, the Canadian situation is a clear exception. The Human Rights Act established a commission which handles complaints, checks wage-schemes against basic evaluation system standards, advises employers on wage system

compliance, and verifies that corrections are made. Employers retain the right to determine their criteria of job worth but the wage determination procedures implemented must meet bias-free guidelines. In short, Canada has adopted a version of comparable worth job evaluation as national wage determination policy.

Comparable Worth in the U.S.

Comparable worth or equal pay for equal value is not an accepted concept in U.S. law. Discrimination claims under the EPA must demonstrate substantial equality of work on each of the four so-called universal compensable factors--skill, effort, responsibility, and working conditions. Comparisons between jobs which are dissimilar, but comparable in value to the employer, do not come under the purview of the Act.

Until recently the same work standard was applied to actions brought under the Civil Rights Act since the Bennett Amendment to that Act (last sentence of Section 703H) had been interpreted as incorporating the EPA equal work standard into Title VII.[4] The recent Supreme Court decision in County of Washington v. Gunther, 80 US 429(1981) modified that position holding that Title VII is not co-extensive with the EPA. However, the court also emphasized that the claimants case was not based on the comparable worth theory, but on direct evidence of employer discriminatory action. Other new rules or principles may become acceptable for establishing a case of pay discrimination under Title VII, but the courts are unlikely to accept the comparable worth concept.[5]

A second Title VII based wage discrimination theory should be noted here because if it is adopted, comparable worth job evaluation will be a primary employer defense or remedy. The theory of job segregation and wage discrimination (Blumrosen, 1979; 1980) builds upon the labor market segmentation theories introduced in the previous chapter as well as on a large stream of social science research establishing that sex, per se, is a major determinant of the value-status-prestige associated with occupations. The basic thesis is that the value assigned predominantly female or minority jobs is depressed because of the views of society toward the jobholders and/or the cumulative actions by employers exploiting the bargaining advantages accruing to them from labor market segmentation. Therefore, wage discrimination can be presumed to exist in any situation involving predominantly female or minority jobs and low wages. Thus, a demonstration of segregation and low wages is sufficient to establish a prima facie case of discrimination under 703(a)(2) of the CRA which forbids segregation of employees in any way that could adversely affect the status of any covered individual.

Under this theory employer rebuttals to job segregation claims must demonstrate either that wages are not low or that wage determination procedures are not biased. The wage level defense would be dependent upon employer ability to equate internal rates with similar or lower rates for the same jobs in other non-segregated establishments. This defense is

possible only in situations where the segregation pattern is unique to the organization or industry (and the rates are, in fact, at least equivalent). More commonly segregation patterns are similar across establishments. Consequently, the primary defense would be the demonstration of sex blind wage-setting procedures, i.e., bias free job evaluation.

The job segregation theory has not been directly tested in the courts, but the argument of disparate impact on female rates due to job segregation and biased evaluation systems is currently before the district court in IUE v. Westinghouse, 19 FEP Cases (D.N.J. 1979).

Strategies and efforts of comparable worth advocates in the U.S. have not been limited to EPA and Title VII litigation. Activities are also targeted at federal agencies, state and local governments, unions, professional societies, and the general population (Grune, 1980). Of particular interest here are the role of the Equal Employment Opportunity Commission (EEOC) and the impact of state and local government commissioned comparable worth job evaluation research projects.

EEOC--Jurisdiction over the EPA was lodged with the Department of Labor, Wage and Hour Division until July 1, 1979 when it was moved to the EEOC to facilitate harmonious interpretation of Title VII and the EPA. Since its inception the EEOC has held that Title VII is not co-extensive with the EPA regarding sex based pay discrimination. However, the priority placed on job opportunity issues, DOL responsibility

for EPA enforcement, and EEOC litigation failures in the equal pay area (Leach, 1978) combined to forestall any significant EEOC initiatives until 1978. The Commission entered into a two year contract with the National Academy of Sciences in 1978 to ". . . determine whether appropriate job measurement procedures exist or can be developed to assess the worth of jobs" (Trieman, 1979:1). The interim report of the NAS committee (Trieman, 1979) served to legitimize many of the previous complaints of women's rights advocates (Remick, 1979) and intensified the already widespread debate over the feasibility and consequences of major governmental intervention into wage determination practices (Livernash, 1980).

Criticisms of traditional job evaluation practice by the NAS committee center on four issues:

1. The use of multiple plans within organizations--facilitates differential treatment of predominantly female occupations and frustrates equity comparisons across occupational groups.
2. Selection of compensable factors--is subject to bias in that factors and measures are oriented to predominantly male jobs thus potentially undervaluing predominantly female jobs.
3. Factor weighting schemes--capture the status quo in occupational wage rates which are negatively correlated with the proportion of female workers in the occupation.

4. Evaluation process--involves a high degree of subjectivity at each stage (job study, data summary, job rating).

While the interim report of the NAS committee formally reserved judgment regarding the desirability of changing the problematic aspects of wage determination, the committee staff drafted a set of guidelines for the development of bias-free evaluation procedures. The proposed guidelines, which were unofficially but widely circulated,[6] address each of the major issues:

1. Each enterprise should use a single job evaluation system for all its employees.
2. The criteria of job worth (compensable factors) should be made explicit, and
3. Factor measures should adequately represent the criteria of worth by:
    a. accounting for all the compensable factors,
    b. being non-reflective of job sex composition,
    c. being demonstrable valid measures of job worth criteria, and justified by business necessity when evaluation results adversely impact female or minority jobs on the average. (emphasis supplied)
4. Point system scales should represent the full range of factor variability, and scale levels (anchors) should be accurate.

5. Factor weights should be assigned in a bias-free manner.
6. Design and implementation procedures should ensure bias-free operation through broad employee representation in system design and through documentation, employee disclosure, administrator training, operation audits, and appeals procedures.
7. A final overall validity guideline requires employers to demonstrate that the evaluation of specific jobs is valid.

Whether this particular set of guidelines will be formally proposed by the EEOC is doubtful. The NAS committee final report took the position that male-female pay differentials are rooted in the institutions and traditions of the labor market and that it is naive to believe that they can be eliminated simply through modification of an administrative process (Miller, 1981). Further, civil rights/ discrimination concerns are low on the priority list of the current federal administration. Thus, the commissioners are unlikely to sustain the equal pay initiatives of the preceding administration. Nevertheless, the guidelines are of continuing importance for three reasons. First, they represent the most explicit statement of the operational meaning of comparable worth job evaluation. Second, deviations from one or more of these standards may expose organizations to pay discrimination actions under the Title VII precedent established by Washington v. Gunther. Third, an

EEOC retreat from the comparable worth debate does not signal the demise of the comparable worth movement. Equal pay advocate resources will simply be shifted to other targets, one of the most promising of which is state and local government pay structures.

Comparable Worth Studies--Women's rights advocates recognize that the federal congress is seldom the leader in civil rights legislation. Twenty-nine states had passed equal pay laws and twenty-six civil rights acts were adopted before Congress passed similar legislation. Thus, state legislatures most likely will be the major forum for comparable worth debates. Comparable worth studies are a primary tactic employed to place the issue before the legislators.

Comparable worth analyses of state and local wage and job structures (public sector) are designed to document sex discrimination in pay rates. Discrimination is defined as unequal pay for work of comparable intrinsic worth to the organization. Job worth is typically measured by one of the standard evaluation systems, sometimes modified to more closely conform to the bias-free model. Sex discrimination is then a documented "fact" to the degree individual or collective pay rates for women are below those of males on comparably evaluated jobs. Regression analyses are normally used to estimate the probability and extent of the differentials.

Formal state government sanctioned comparable worth studies have been conducted or are in process in at least

six states (Wash., Minn., Mich., Conn., NY, Nebraska). General studies of the state classification and compensation systems are underway in at least three other states stimulated by concerns over male-female disparities (Grune, 1980). In addition, numerous smaller scale studies of local public sector and regulated utility pay practices have been initiated in response to union and employee association pressures. The writer is not aware of a single instance in which a comparable worth study failed to "document" the existence of sex discrimination in pay.

Comparable worth studies and results serve a number of functions. First, and most fundamentally, they stimulate awareness of the inconsistency between internal norms of job worth and existing pay patterns and thus generate commitment to the equal pay cause. Second, study data can serve as the basis for litigation, although the consistent failures in comparable worth based actions have probably reduced support for this strategy, at least temporarily.[7] Third, comparable worth studies are bargaining tools providing an additional means for generating union member and citizen support for work stoppages and other actions seeking equity adjustments. This bargaining strategy at the local government level has yielded the only concrete results in terms of classification changes and/or salary budget reallocations. The American Federation of State, County, and Municipal Employees has been particularly active in the use of comparable worth studies in collective bargaining (Grune, 1980; Matthews,

1981). Finally, the studies appear to be a necessary prerequisite to galvanizing support for new legislation, although no state has yet enacted legislation mandating pay according to job worth. Countervailing forces in the form of dollar and political costs, traditional job relationships, and reluctance to challenge fundamental assumptions of market fairness make the legislative route difficult (Taber and Remick, 1977).

In summary, the comparable worth movement in the United States is not a passing phenomenon. It is rooted in a women's rights movement with a demonstrated capacity to persist and prevail. Comparable worth as a legal standard for establishing sex based pay discrimination is not likely to be accepted by the courts, Congress, or state legislatures in the near future. However, state laws specifying comparable worth as a primary principle governing state employee compensation programs are a probable development in the next 2-3 years; local jurisdictions have already adopted such policies. Further, union bargaining and arbitration efforts in female intensive occupations (communication, electrical equipment, health care, clerical) will continue to stress inequity adjustments based on comparable worth comparisons, and female dominated professional associations (American Nurses Association, Women Library Workers, etc.) can be expected to be more militant in support of the comparable worth principle. The allure of an apparently more objective (and favorable) pay determination principle to low paid

female workers can be expected to generate increasing support for the comparable worth concept in the years ahead.

CRITICAL ISSUES

Opposition to comparable worth job evaluation stems from concerns about both feasibility and potential adverse consequences. In the latter category most of the attention has been on the cost impacts (organization and societal levels) and on disruption of internal and external human resource allocation processes due to realignment of occupational differentials. While these concerns are interesting and arguable, they are beyond the scope of this paper. Discussion here is limited to the question of whether objective job evaluation is possible.

The basic premises of comparable worth job evaluation are two: (1) a generally accepted definition of job worth can be determined within an organization (the "true" worth of jobs to the organization); (2) Job worth can be objectively measured through bias-free evaluation procedures. Both of these assumptions are problematic.

As noted above, job evaluation practice has been predicated on the notion that the relevant aspects of jobs for job worth assessments vary across broad functional groups. For example, physical effort and working conditions may be important on the factory floor, but responsibilities and expertise demands are the critical aspects of professional and managerial positions. A similar belief underlies comparable worth advocate arguments for participative evaluation

plan design processes (guideline no.6). Differing concepts of job worth among males, females, and minorities are thought to require open processes to enable the identification of "all" the relevant factors. Thus, both job evaluation practitioners and critics agree with the perspectives reviewed earlier; job worth criteria are particular to situations and individuals. Consequently, a generally accepted specification of the conceptual boundaries and relevant dimensions of job worth is improbable in all but very small and homogeneous establishments. The definition of job worth is not an empirical question but a question of whose values will prevail. The "true" worth of a job is defined in specific situations by the criteria determined through the decision processes of the organization. The validity of such criteria seldom will go unchallenged by some members of the organization.

The reality of differing value systems within organizations and the desire for a single system of job evaluation create a dilemma for comparable worth advocates (CWA's). If they insist upon adoption of their criteria of job worth, the male-female equity issue takes on power distribution dimensions, i.e., a direct challenge to management rights is issued. Such a position could be fatal to CWA's by providing a powerful counter-argument to the equity plea. If, however, CWA's recognize the right of employers to determine job worth criteria, the evaluation plan implemented may be disadvantageous to women. The NAS committee draft guide

lines took the latter position but considerable disagreement among CWA's exists on this issue.[8]

The fact that job worth must be situationally defined also creates the possibility that Mahoney (1981) has labelled the paradox of aggregation. Judgments of the relative worth of jobs in individual establishments may be based on criteria generally acceptable to that social unit but may aggregate into measures of earnings are another level of analysis which conflict with broad social norms. "It is just possible that a male-female earnings gap for the American labor force is consistent with accepted judgments of comparable worth within employer and work group settings" (Mahoney, 1981).

Bias-free Job Evaluation

Talk of bias-free judgments of job worth is a contradiction in terms; value judgments are inherently subjective. However, other wage determination mechanisms are similarly flawed, although perhaps less directly or obviously. Thus the relevant issue is not whether objective job evaluation is possible but whether a job worth judgment process can be devised such that job structure decisions accurately and consistently reflect the concept of worth defined by the affected parties. In short, given a subjective job worth construct, can it be validly measured via job evaluation procedures?

Three basic questions must be addressed in assessment of evaluation validity. First, does the evaluation system

as designed truly mirror the specified job worth concept? Second, are appropriate design and implementation practices and procedures followed to ensure maximum objectivity of ratings? Third, are the obtained evaluation results contaminated by error due to evaluation system characteristics or evaluator judgments? Answers to questions one and two require detailed analyses of evaluation system design features and implementation processes--the focus of the NAS committee proposed guidelines. Question three raises the issue of "criteria for criteria," i.e., by what standards can the validity of scores be assessed? Measurement validity cannot be demonstrated empirically since job worth is defined in conceptual rather than market terms as previously. Thus, assessment of job evaluation measurement adequacy requires a construct validation process; validity of measurement can only be inferred from examining a variety of indirect indices. Job worth measurement validity considerations pertaining to these design, implementation and criteria assessment questions are summarized here. (For purposes of discussion point-factor evaluation is assumed since this is the method reflected in the comparable worth job evaluation guidelines).

Design Considerations--Job worth measures are generated by raters employing verbal scales to assess the degree of various compensable factors in a particular set of jobs. The possible score on each factor varies according to the relative importance (weight) of the factor. Thus, the major

design considerations relating to job evaluation validity focus on factor selection, scales, and weights.

Assuming a job worth construct is clearly specified in a particular situation, conceptual agreement on its dimensions does not automatically translate into concurrence on factor measures. Physical effort, for example, can be measured in terms of maximum periodic force applied (in lifting, pushing, etc.) or in terms of cumulative impact on the employee (fatigue). Using measures of the latter type, traditional "light-work" occupations (often female dominated) would likely receive greater credit than under the former. Thus conceptual definition problems apply to compensable factor as well as job worth constructs. Each major factor may have sub-dimensions for which definitions must be specified and appropriate measures identified. Validity in job evaluation requires a "complete" set of measures; each conceptual dimension must have a relevant measure.

The generation of a set of relevant job worth measures can be expected to increase in difficulty with organization size/complexity. Differing perspectives create pressures to include factors and subfactors pertinent to specific functions.[9] This phenomenon can negatively impact validity in two ways. First, the typical factor selection/definition procedure is a committee activity, and therefore a political process. Objectivity is seldom the only or even the primary concern. Second, the necessity to reconcile different views

typically leads to inclusion rather than exclusion of factors. This proliferation of compensable factors is incompatible with the measurement validity requirement for factor independence (discriminate validity). Where a high level of factor interrelatedness exists, the effective factor structure (reflected in the actual measures) may differ substantially from the designed structure.

A set of comprehensive, independent measures does not, of course, ensure validity. The extensive literature on rating scale development in other areas of human resource management strongly attests to that fact (Kane and Lawler, 1979). Technical issues of scale anchors, intervals, and the discriminability of scales (design and operational) also must be addressed, considerations regarding which little research data in the job evaluation context currently exists.

Factor weighting schemes are critical to the validity of comparable worth job evaluation systems, but ensuring that the operational weights correspond to the relative importance of factors intended in the system design is extremely difficult. Where weights are obtained by means of regression techniques, design-operation discrepancies are inevitable. Even if the criterion structure is bias-free, the probability that statistically derived weights will correspond with judgments of relative factor importance is nil. If factor weights are judgmentally assigned, obtained measures will still not necessarily be in accordance with the weighting scheme since true (operative) weights are a

function of actual factor score variability. Variance in factor ratings, in turn, is partially a function of the context in which the rating system is applied. Therefore, while it may be theoretically possible to obtain correspondence between design and operational weights in the development stage, distortions are probable when the system is implemented in differing contexts by a variety of evaluators.

A final design consideration pertains to pay equation decisions. As noted earlier, job clustering or classification processes often allow a certain amount of latitude in job worth decision making, particularly regarding job scores at the extremes of the ranges. Thus the job worth score/dollar value ratio may vary within and between pay grades. Such flexibility is inconsistent with the concept of measurement validity.

Evaluation Process Considerations--The job evaluation process can be segmented into three phases--job analysis, data summary, job rating--each of which provides multiple opportunities for error in human judgments. Figure 5 summarizes the primary sources within each phase.

The validity of job evaluation is totally dependent upon the quality of job-related information available to raters. However, as McCormick has noted, job analysis efforts". . . have tended to be rather unsystematic and to be more subjective than objective, and have been shrouded in verbiage that does not lend itself to systematic analysis" (McCormick, 1976:652). Thus the data base upon which job

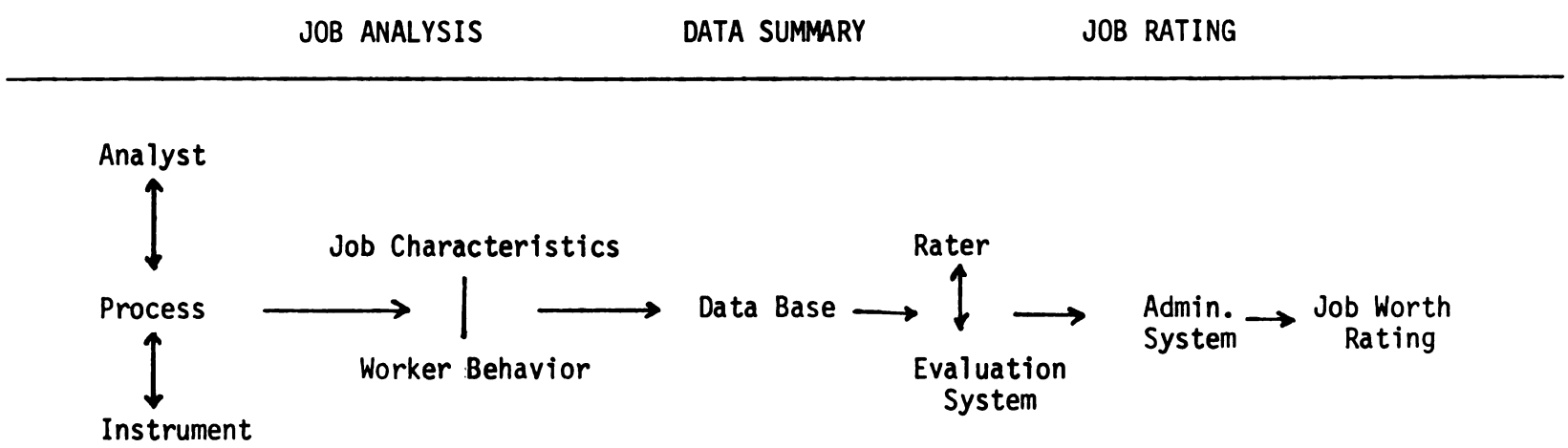


Figure 5. Error Sources in the Job Evaluation Process

worth judgments are made may be inadequate due to rater, instrument, and/or analysis process deficiencies.

Data summaries or job descriptions are bypassed in some job evaluation processes but the more common practice is to prepare them as supporting documents. Job descriptions also are used to "validate" analysts' perceptions, i.e., a review and approval by interested parties is obtained prior to actual job rating. Thus job descriptions can contribute to rating inadequacies in at least three ways: (1) Misinterpretation of verbal descriptions; (2) Data deficiency due to summary format; (3) Inaccuracies attributable to distortions generated in the review process.

Opportunities for inconsistency abound in the actual job rating process. Differences in rater characteristics --demographic, organizational status, training/ability, personality, motivation--could affect outcomes. Administrative characteristics can also affect results, particularly where they increase susceptibility to bargaining such as in jointly administered systems or those including appeals provisions.

In summary, the design and implementation considerations pertaining to valid job evaluation measures actually go beyond the issues reflected in the proposed guidelines, particularly with respect to rating scale construction, weights, and the job analysis process. It is at least possible, if not probable, that an employer could rigorously adhere to the proposed guidelines in developing and

implementing a system, yet obtain invalid results--outcomes reflecting deficiencies or contamination in job worth measurement. Therefore,job worth measurement results must also be analyzed for evidence of validity.

Validity Criteria--In traditional job evaluation the ultimate criteria of job worth are those forces represented by the market or traditional key job wage structures. Thus the appropriate validation strategy is to determine the degree of correspondence between evaluation outcomes and the criterion structure (Fitzpatrick, 1949; Fox, 1962). Construct validation, however, is a more difficult and indirect process.

Construct validity here may be thought of as the degree of correspondence between the job worth measures and the job worth concept. This hypothetical correlation between evaluation scores and true job worth will be less than 1.0 to the degree the measure does not reflect true job worth variance (deficiency) or it reflects variance attributable to extraneous sources (contamination). The former condition quite obviously cannot be detected through analysis of measures from a single instrument, but can be inferred through comparisons with other measures of the same construct. If multiple independent measures of a construct yield corresponding results, measurement deficiency is less likely to exist and an inference of validity is supported. This confirmation of findings by independent measurement procedures is typically referred to as convergent validity.

Contamination, on the other hand, can be partially estimated through analysis of a set of obtained measures because the method itself is a possible contaminant. For example, job evaluation disagreements between raters can signal evaluation scale ambiguity or systematic rater error such as bias toward certain jobs, differing interpretation of evaluation dimensions and scales, or inadequate use of the evaluation plan discriminability. Similarly agreement between raters on different compensable factors across a set of jobs can provide evidence for the absence of contamination due to factor overlap or ambiguity, i.e., factor discriminate validity.

A number of statistical approaches to estimating rater consistency, measure discriminability, measurement dimension independence, and potential range of measurement error have been developed and tested (Schwab, 1980). The problem in the job evaluation context is not in generating evidence of validity/invalidity but in interpretation of the indices obtained. Measurement property norms appropriate to other areas such as psychological testing or performance appraisal may not be relevant. For example, a reliability coefficient of .90 would be considered acceptable in many testing situations but its interpretation in terms of consistency in rater pay grade assignments is unknown. Therefore, if statistical analyses of job worth measures are to be useful, criterion values specific to job evaluation ratings need to be developed. Examination of this issue of evaluation

validity criteria is a primary thrust of the research effort reported in chapters 4 and 5.

Research Issues

The preceding summary of problems/considerations involved in valid job worth measurement raised a number of potential research questions pertaining to the design, implementation, and evaluation of job evaluation systems. The basic purpose of the current research effort is to assess the feasibility of comparable worth job evaluation by testing different systems for measurement error and method variance. The general proposition under investigation is that job evaluation system differences in capacity to reflect job content variability, in the relative importance attached to various job characteristics, and in susceptibility to measurement error, ultimately translate into inter-system variance in relative job worth assessments. Thus the research focus here is on generating information regarding two basic and interrelated issues: (1) To what degree is measurement error a potential problem in job evaluation? Do evaluation plans differ significantly in their measurement properties? (2) Is method variance a significant concern in job worth measurement? Does the hierarchy of jobs obtained depend upon the method used or do the results of different evaluation plans tend to converge? If evaluation outcomes are significantly impacted by measurement error and/or are method dependent, the feasibility of comparable worth job evaluation is seriously called into question.

The following research questions were formulated to address these issues:

1. What level of agreement regarding job worth (inter-rater reliability) is obtained by different analysts evaluating the same jobs employing a comparable worth job evaluation method?
2. Are there systematic differences in the level of specific job ratings (bias) or overall rating levels (leniency/severity) between analysts?
3. Is comparable worth job evaluation more or less susceptible to inter-rater differences than traditional methods?
4. To what degree are individual factor measures independent?
5. Do evaluation results reflect the intended factor weighting scheme?
6. What is the range of potential error for job worth scores?
7. Do summary job evaluation ratings produced by different methods converge in terms of job position in the hierarchy or pay grade assignment?
8. Are inter-rater reliability coefficients and inter-method correlations appropriate indices for assessing the validity of job evaluation measures?

Some limited evidence bearing on these questions can be found in the research literature on job evaluation and the application of rating techniques in other areas. If these

findings are applicable to the comparable worth evaluation context, they are not particularly encouraging regarding the probable adequacy of evaluation measures. It is to a review of this literature we now turn.

CHAPTER NOTES

1. Findings in this study reported in Chapter 5 support this observation

2. To the writer's knowledge the only exception to this procedure is one phase of the study currently being completed for the Office of Women and Work, Michigan Department of Labor, in which PAQ items were assigned weights on an a-priori basis.

3. This type of variability in policy was recently found to violate the Civil Rights Act in a case where the classifications were also sex segregated (County of Washington v. Gunther, 80 US 429 (1981).

4. The Bennett Amendment states that sex based pay differentials are not illegal under Title VII if ". . . such differentiation is authorized by the provisions of Section 6 (d) of the Fair Labor Standards Act of 1938, as amended."

5. Judge Winner, in Lemons v City and County of Denver 17 FEP Cases 907 (D. Colo., 1978) saw in comparable worth the potential to disrupt the entire economic system of the United States.

6. ACA News, June-July 1979; D.J. Thomsen, "Proposed Guidelines for the use of Job Evaluation in a Non-Discriminatory Manner," Compensation Institute, April, 1979; Bureau of National Affairs, Daily Labor Report, June 6, 1980.

7. See for example, Christensen v. State of Iowa, 16 FEP Cases 232 (8th Cir. 1977) or Lemons v. Denver (note 5).

8. The deference to employer definitions of value is one of the major criticisms of the Canadian Act by women's organizations (Ratner, 1979).

9. This phenomenon was observed in both the instant research project and a subsequent project in a different organization.

CHAPTER 3

Job Worth Decision Research

The outstanding characteristic of job evaluation research is its paucity in comparison to research in other personnel management decision areas such as selection and performance appraisal. A number of studies were reported in the late 40's and early 50's, a natural consequence of the widespread adoption of formal evaluation systems during that era. For the past 25 years, however, job evaluation research reports have been infrequent in spite of continually increasing utilization of evaluation systems. Only in the past 5 - 7 years is a slight upswing in job worth decision research noticeable, attributable primarily to increasing interest in quantitative approaches to the development of job and task taxonomies. Job evaluation applications of these instruments are typically by-products; the primary concern is usually with hiring, promotion and training decisions.

The most probable explanation for the low level of research interest in job evaluation rests in the fundamental nature of evaluation processes. Selection and performance appraisal decisions present complex theoretical and technical measurement challenges, but traditional evaluation entails only routine prediction. There is no "criterion problem," the market or policy structure fulfills this need. Compensable factor measures are not independent assessments

of worth, rather they are predictors of a desired outcome. Consistency, efficiency, and acceptability of prediction are the primary critieria of evaluation results; validity of measurement in the complete sense of the concept is unimportant. Thus traditional job evaluation research generally has been confined to simple empirical issues pertaining to reliability and the explanatory power of various evaluation models, questions which do not sustain interest over a long period of time.

A general caveat is appropriate at this point. The fact that researchers have heretofore considered job evaluation as primarily a rationalization process, whereas in the comparable worth context job evaluation is a measurement process, could limit the relevance of past findings. For example, the level of rater agreement regarding the evaluated worth of a set of jobs might logically be expected to be higher when the rating scheme is designed to predict a known benchmark structure than when the rating, per se, is the determinant of worth. Similarly, greater agreement between evaluation methods should be expected when each of the methods has been designed and "validated" relative to a common criterion than when they have been independently designed to measure an abstract construct. Thus the findings summarized here must be carefully considered in terms of the particular process employed in the study; considerable caution in interpretation of their relevance to comparable worth evaluation is warranted in all cases.

Research studies addressing issues of method variance and/or measurement error in job evaluation are the central interest of this chapter. The review begins with a focus on methods, first in terms of the measurement properties of individual evaluation plans and then through method comparisons in terms of a variety of criteria. These sections are followed by investigations of error determinants within the job evaluation process categorized somewhat arbitrarily according to major error sources (Figure 5). No attempt is made to review research pertaining to the content of evaluation plans. Evidence of criterion relevance variability is mixed (Belcher and Atkinson, 1970; Selznick, 1969; Carrell and Dittrich, 1978). The operating assumption is that a generally accepted concept of worth can be obtained, i.e., the emphasis here is on the process aspects of "nonbiased" evaluation. Also excluded is a thorough review of the comparable worth research projects referenced in chapter 2 which explore pay distributions and correlates and only incidentally reveal qualities of the measurement approaches utilized.

Evaluation Methods: Properties

Two questions dominate the research characterizing job evaluation plan measurement properties. First, researchers are concerned with measurement error as reflected in reliability estimates of evaluation ratings. In judgmental processes rater differences are the most probable source of error, thus studies typically utilize inter-rater reliability indices to estimate measurement error. Quite obviously,

measurement reliability is a necessary condition for both traditional and comparable worth job evaluation. Second, the dimensionality of job rating systems is explored, motivated initially by a desire to determine whether abbreviated scales offer efficiencies in the evaluation process (Lawshe and Satter, 1944), and more recently by the necessity to reduce data from lengthy task questionnaires to interpretable dimensions (McCormick, Jeanneret, and Mecham, 1972). As noted earlier, non-biased evaluation requires dimensions which are complete, yet independent, reflections of the job worth construct.

Dimensionality--Early and recent approaches to dimensionality analysis - both employing factor analysis procedures - appear on the surface to lead to conflicting to conclusions. Lawshe and associates (1944, 1946, 1948) conducted a series of studies on commonly employed 10-12 factor point evaluation plans. Their first analysis of an eleven factor plan suggested a two factor solution, skill demands and job context/physical demands. In the two subsequent studies, three and five factor interpretations were adopted. A similar recent study applying a fifteen factor system to state government jobs yielded a four factor solution featuring knowledge/skills, working conditions, supervisory responsibilities, and work pace (Michigan Department of Labor, 1981). Thus the widely held conclusion that relatively few factors are necessary to adequately differentiate the worth of jobs (Belcher, 1975) has been recently corroborated.

The findings of McCormick et al., (1972, 1977) on the other hand, indicate a need for multiple dimensions to adequately characterize jobs. The 187 PAQ items reduce to only 32 dimensions when the six divisions are individually factor analyzed. For example, three logically distinct job context dimensions are identified - potential hazard, physical demands, and stressful/unpleasant environment. When all 187 items are submitted, a more global 13 factor solution emerges, still considerably in excess of the 2 - 4 factors resulting from analysis of traditional evaluation plans.

Discrepancies between the two sets of studies are likely attributable to differences in the ultimate objectives of the PAQ and evaluation plans. The PAQ was designed to reflect all worker behaviors on all jobs regardless of whether they are important to worth determination. Evaluation plans typically incorporate only those dimensions thought to influence wage determination in a given context. Factor analyses of evaluation plans suggest that even these limited sets of dimensions are redundant; they do not address the issue of whether all relevant dimensions are considered. Thus PAQ factor analysis results could be more indicative of the "true" dimensionality of jobs, i.e., the differences between PAQ dimensions and reduced evaluation factors might represent potential criterion deficiency.

Reliability -- Inconsistencies in job evaluation scores could be the result of employing different raters or due to factors affecting the ratings of evaluators on a particular

occasion. Only one small sample (N=38) study addressed the latter issue. Richardson (1971) tested the stability of incumbent and supervisor ratings of jobs on four common evaluation factors after a two month interval. Moderate to high (.62 - .91) correlations for the supervisors were found but incumbent coefficients were below .25 on 3 of the 4 factors. The discrepancy between supervisor and employee rating stability would seem to rule out true change in the jobs as the source of variance. Thus temporal consistency may be a greater concern in evaluation than has heretofore been acknowledged.

Inter-rater reliability coefficients for the total scores of point evaluation systems tend to be above .75 but mask considerable factor level variation. Lawshe and Wilson (1947) reported overall reliability for their long and short plans as .77 and .89 respectively, but individual factor coefficients ranged from .37 to .86. Similarly Chesler (1948) found an average overall score inter-rater correlation average of .97 but indicated that point variation was considerable (no indices reported). The Michigan study (1981) reported an overall score coefficient of .80 with individual factor correlations ranging from .66 to .97. However, these coefficients apparently can be improved somewhat by pooling judgments of multiple raters (Lawshe and Wilson, 1947; Satter, 1949; Christal, Madden and Harding, 1960).

Inter-rater reliability estimates for other approaches to job evaluation are apparently comparable to slightly lower

than for point plans. Satter's (1949) analysis of evaluation factor rankings by multiple judges (paired comparison rankings) yielded average pair ranking reliabilities from .62 to .81. The procedure employed was quite similar to that required by factor comparison systems. Whole job ranking was thoroughly analyzed during the criterion determination phase of the federal government Factor Evaluation System development program. Anderson and Corts (1973) obtained coefficients of agreement among multiple rankings (Kendall's concordance) on eight sets of 21 jobs each, ranging from .61 to .75. Finally, Ash found inter-rater correlations ranging from .81 to .94 utilizing a procedure similar to classification approaches but his results are probably inflated due to the fact that the judges (Labor Department analysts) were ranking standard occupational, rather than job, descriptions.

PAQ reliability estimates have been reported on three of the four possible bases (item scores, division dimension scores, overall dimension scores, predicted salaries). McCormick et al. (1972) obtained job component inter-rater reliabilities (187 items) averaging .79 across multiple analysts and jobs and report a similar study (1977) where the average coefficient was .68. Hakel and Smith's (1979) findings were even lower with rater pair correlations ranging from .49 to .63. It should be noted at this point that PAQ item reliabilities might logically be lower than traditional factor evaluations due to the numerous and wide-ranging

judgments required. However, reliability coefficients apparently do not increase significantly when aggregated PAQ items scores are employed. The median PAQ division dimension reliability coefficient among four studies was .64 with the range extending from .15 to .95 (McCormick et al., 1977), and overall dimension scores in the Michigan Study (1981) ranged from .41 to .94 with a mean of .80. Item, division dimension, and overall dimension reliability estimates may not be directly comparable to other job evaluation coefficients due to differences in objectives and approach of the instruments, but a reliability analysis of PAQ based salary predictions would provide an appropriate referent. No such studies have been reported. However, Gomez-Mejia, Page, and Tornow (1979b) obtained correlations of .56 and .61 in their estimates of inter-rater agreement regarding salary grade prediction when utilizing questionnaire and regression methodologies similar to the PAQ.

In summary, it should be emphasized that while the reliability estimates in the majority of studies were interpreted as acceptable by the investigator(s), the basis for the conclusion is not always apparent. In none of the reports is logical or empirical support of the contention offered. Taken in totality the studies are cause for pessimism rather than optimism; significant differences in hierarchical assignments of jobs among raters is probable within a particular evaluation procedure.

Evaluation Methods: Comparisons

Comparative analyses of job evaluation methods typically have been feasibility studies attempting to assess some new or modified approach relative to existing evaluation techniques. The usual objective of the modification has been administrative efficiency, sought via reducing subjectivity and/or the number of criteria. Analytical strategies focus either on explaining variance (regression) or assessing degree of correspondence, the categories utilized here to facilitate presentation.

Regression Analyses--Regression based evaluation procedure comparisons test the degree to which a new or modified plan X captures variance measured by plan Y in a given context. A high degree of explained variability (R^2) is interpreted as evidence of the interchangeability of X and Y. Three studies of this type have been reported.

Lawshe and Maleski (1946) tested short (3 factor) and long (11 factor) versions of a common point plan and found the abbreviated version accounted for 96% of long plan variance with one factor, skill demands, accounting for over 90% of explained variability. More recently Robinson, Wahlstrom, and Mecham (1974) obtained an R^2 of .90 utilizing a stepwise procedure regressing PAQ dimension scores on the evaluation points assigned to positions in a major utility. These results are not surprising in view of the similarity between certain PAQ and point evaluation scales. However, Foster (1977) also developed high levels of agreement between the

point values of management jobs in four companies (at least two different evaluation systems) and predictions based on objective job measures such as reporting level, number of people supervised, age, and unit or function. In each case two or three of the "hard" factors accounted for at least 80% of the evaluation point variance.

The regression studies are important in that they demonstrate traditional models of the underlying determinants of job worth may be replaced/improved, but they do not address the issue at hand here - the extent to which hierarchical assignment varies by method.

Method Correspondence--Simple correlations among two or more sets of evaluation scores have been used by most of the studies testing degree of method agreement. The scores of short and long versions of the same plan correlate quite highly (Lawshe, 1945; Davis and Tiffin, 1950). The situation with respect to different types of plans is not so clear. Chesler's report of intermethod correlations ranging upward from .89 among 7 different plans must be discounted on the basis of insufficient information. Bellows and Estep (1950) found a correlation of only .74 between a checklist and a crude two factor system and Atchinson and French (1967) obtained coefficients ranging from .54 to .82 among classification, maturity curve, and time span measures. The NAS review (Treiman, 1979) included an unpublished study reporting a .94 correlation between a guide chart approach and the federal classification system, however the heavy reliance on

benchmark comparisons in both systems would render surprising a divergence of any greater extent.

The Robinson et al. (1974) findings also must be cautiously interpreted. Nineteen benchmark jobs in a city classification system, ranging from meter maid to the chief financial officer of the city, were evaluated by five methods (market, ranking, point, factor comparison, PAQ standard equation), with method intercorrelations ranging from .82 to .95 (median = .89). Limiting the evaluations to benchmarks, pooled judgments, and factor comparison system dependence on market data all operate to inflate method correspondence. However, investigation of hierarchical assignments under the five systems reveals major rank differences. For example, one job is ranked number 1, 2, 4, 6, and 9 by the five systems, a fact which underscores the major weakness in studies of this type. Correlation coefficients alone are inadequate for comparing evaluation system outputs because major rank and interval discrepancies can be masked. The appropriate analysis is an assessment of the degree to which methods assign jobs to the same classification or pay grade, i.e., method convergence or agreement rates.

Only one study has directly compared evaluation methods in terms of convergence rates. In Lawshe's first study (1945) 62% of the short and long plan pay grade assignments were found to be the same and 99% of the cases were within one level. Gomez-Mejia et al. (1979b) provide indirect evidence that agreement rates may vary significantly among

method pairs. Prediction hit rates were employed to assess the accuracy (current grade structure = criterion) of seven statistical and traditional evaluation methods. Significant differences in the rates of the approaches were found. The low magnitude of the hit rates - errors of greater than one grade level ranging from 27%-51% of the predictions - also reinforces concerns about measurement validity since the scores from all methods were generated from the same questionnaire data base, i.e., job analysis variance across methods was controlled.

Convergence rates, like correlation coefficients, probably decrease with evaluation method divergence. However, the relationship between correlation indices and salary grade assignments has not been directly explored. The Robinson et al. (1974) rankings and Lawshe's (1945) findings (r = .90 and convergence = 62%) suggest significant discrepancies can be expected. Contrary to the conventional wisdom in job evaluation, the judgment method employed probably has and will affect grade assignments determined through traditional and/or comparable worth procedures.

Rater Effect

Measurement contamination in the job evaluation process can take the form of individual or interactive effects attributable to characteristics of the judge or rater, evaluation instruments, and the job or object of the rating. None of these error source categories could be described as thoroughly researched. Rater bias has received the most

attention but the demographic, psychological, and environmental variables explored in other realms of human judgment research generally have been ignored in the job evaluation context;[1] effects due to the rater's role and/or job familiarity have been the phenomena of primary interest.

While there is no evidence that inter-rater reliability coefficients vary among analysts, incumbents, and supervisors (Satter, 1949; Smith and Hakel, 1979), the effect of the rater's frame of reference on evaluation outcomes is still open to question. Wiley and Jenkins (1963) found bias to be present and estimable, but the other Air Force Studies reported in the early 60's (Madden, 1960, 1962; Harding and Naurath, 1960) found general consistency of ratings among judges (officers) varying in their personal knowledge of jobs. Only minor differences on certain managerial factors were indicated. Similar consistency was found between worker and supervisors (Hazel, Madden, and Christal, 1964). These findings were partially supported in the PAQ ratings obtained by Smith and Hakel (1979) in which supervisor-incumbent ratings were consistent, but analysts' ratings were considerably lower than the other two groups. In contrast, PAQ ratings of analysts in the Michigan study (1981) were higher than those of incumbents. The PAQ frame of reference also apparently can vary by the rater's sex. Arvey, Passino and Lounsbury (1977) found females tended to rate jobs lower than their male counterparts. Finally, outside of the job evaluation context, Dubin, Porter, Stone, and Champoux (1974) reported a

tendency of incumbents to rate their jobs higher than did supervisors on a variety of job characteristics, and the recent literature in the job design - job satisfaction area indicates worker frame of reference significantly impacts perceptions of job characteristics (James and Jones, 1980; O'Reilly, Parlette, and Bloom, 1980).

The contradictory results regarding rater differences are not easily reconciled. It is the writer's opinion that the military studies probably understate rater effects due to clearer specification of organizational stratification and homogeneity of judges than is typically the situation in civilian organizations.

Interestingly, in spite of the common practice of assigning new employees to job analyst positions, and in spite of the importance attached to rater training by job evaluation consultants, no studies considering analyst training and/or experience were found. This void is even more surprising in view of obvious individual differences in observation and judgment capabilities, skills of critical importance to the job evaluation process.

Instrument Characteristics

A discussion of the effects of instrument characteristics on job evaluation outcomes theoretically is pertinent to both point and classification systems, the methods in which fixed evaluation standards are provided. However, since classification descriptions in practice are often classic

examples of purposeful ambiguity and the process more akin to ranking than rating, concern about instrument characteristics is limited to point type rating systems.

Instrument characteristics here are defined to include both content and scale format variables. First, the ratability of evaluation factors or sub-factors could vary due to their inclusiveness or complexity. Some systems require summary judgments on multidimensional criteria; in others relatively simple criteria are employed. Similarly, scale content could be differentiated in terms of quantitative v. qualitative aspects, although no studies considering this issue were identified. Second, scale formats vary in terms of number and types of degrees, use of grids or graphics, etc., variables which could also result in differing psychometric properties.

The research literature regarding instrument characteristics is extremely lean. Mosel, Fine, and Boling (1960) found that ratability varies by factor, a conclusion that is supported by variation in factor level inter-rater reliability estimates. Lawshe and Wilson (1947) observed that coefficients were highest for skills type factors and Harding, Madden, and Colson (1960), found similar evidence favoring concrete over abstract criteria. These findings appear logical in that greater specificity of criteria should result in higher reliability. Decomposing global criteria into their components would seem to be a way to improve rater objectivity. In fact, quantitative job analysis questionnaires embody

such a strategy. However, whether ratability improves with a large number of more specific items is debatable. The inter-rater reliability estimates from PAQ scores are in the same general range as those of traditional approaches, and in a recent assessment of physical effort ratings, (Hogan, Ogden, Gebhardt and Fleischman, 1980), inter-rater reliability coefficients of only .59 and .60 were estimated from ratings (N = 50) of 25 specific (and common) physical tasks. Finally, Cornelius and Lyness (1980) found overall judgments of jobs to be the equal of task rating combinations (same criteria) in terms of rating stability and inter-rater reliability. Thus, at this point, no conclusions regarding scale content can be drawn.

Analyses of job evaluation scale formats have been limited to the Air Force investigations of scale interval definitions (Madden 1964) and layout (Horizontal v. vertical, graphics) or scale numbering differences (Madden and Bourdon, 1964). Thoroughly defined scales yielded more reliable ratings than those defined only at the extremes, and layout had no affect on reliability but some modest impact on rating level.

A curious research omission in the scale format area is that regarding scale intervals since some of the manuals take strong positions on the relative merits of arithmetic v. geometric progressions. More importantly to the comparable worth context, nothing in the area of the sensitivity or discriminability of factor scales has been reported. It is

conceivable that true job variability in traditional factors is only fractionally being evaluated.

Object Effects

The quality of job evaluation outcomes is not only a function of the rater and method, but is also potentially affected by the object of assessment, the job. Analysts' inputs may come from direct exposure to the job and/or from descriptive documents. In instances where the job-person distinction is tenuous, the jobholder is also a determinant. Thus, jobs, job descriptions, and incumbent characteristics are potential object effects. Again in this category, studies are few in number.

While the basic nature of jobs in terms of their amenability to analysis seems to be an obvious consideration in job evaluation system assessment, only two studies exploring this area were located. In both cases the effect of job level on reliability was investigated - with surprising conclusions. Smith and Hakel (1979) found a slight tendency for PAQ rating reliability to increase with job level, and the Michigan study (1981) revealed a similar tendency in PAQ ratings. A strong main effect for job level in the reliability of traditional factor ratings was also found in the latter study; cell means indicated a positive relationship. Considering object effects alone, the logical expectation would be for rater inconsistency to increase with the greater complexity typical of higher level jobs. Thus these findings

may be indicative of some form of rater by job interaction, an unexplored area of job evaluation research.

Job effects on rater agreement also could be operationalized in terms of broad occupational categories (blue collar, clerical, professional/managerial), but the standard practice of utilizing different evaluation plans for each of these groups likely accounts for the absence of such studies. This issue could be important to comparable worth evaluation which requires a single evaluation system within administrative units.

Job descriptions vary in terms of amount and objectivity of information, presentation format and orientation (tasks, behaviors, responsibilities or outputs). With one exception however, job description characteristics have been overlooked. Madden and Giorgia (1965) found merit rankings based on brief narratives differed from those derived from factor score profiles, perhaps indicative of an information quantity effect.

Object effects attributable to stereotyping processes have been widely examined in other personnel decision making areas (Rosen and Jerdee, 1974a; Milula, 1974; Cohen and Bunker, 1975; McIntyre, Moberg, and Posner, 1979) but only two studies have examined the issue in the job rating context. Arvey et al. (1977) examined incumbent effect on job ratings and found that PAQ scores were unaffected by sex of incumbent. However, Mahoney and Blake (1979) attempted to investigate the more likely phenomenon of occupational

stereotypes (Shinar, 1975) and found evidence of sex effect in student valuations of standard occupations. Whether such stereotypes affect job worth decisions and the methodology by which their impact can be assessed are, of course, central issues in the comparable worth debate.

One final study deserves mention here, not because of its findings (weak) or methodology (inappropriate), but because of the implied interaction hypothesis. Prien and Saleh (1963) attempted to investigate the impact of method (interview) and incumbent characteristics (tenure and performance) on job evaluation ratings. Rating scores were hypothesized to be influenced by analysts impressions of the interviewee. Given the obviously complex observation and judgmental processes involved in job analysis/evaluation, it is somewhat surprising that propositions of this type have not been investigated more frequently.

Job Analysis

The job analysis process, per se, can also be a major source of variance in evaluation outcomes with error sources paralleling those previously mentioned. This fact already has been evidenced by the numerous preceding references to outcomes and determinants of job component questionnaires, job analysis methods in which the job observation and judgment processes are seemingly merged into a single phase. However, the distinction between the two processes is more clearly seen in traditional methodology in which job analysis information is typically reduced to a verbal summary which,

in turn, provides the basis for evaluation. The critical role of data collection and summary (job analysis-description) is readily apparent; job evaluation validity is wholly dependent on job analysis validity.

Traditional job analysis has not been critically assessed for the obvious reason that verbal job descriptions are not readily analyzed. Thus, estimates of job analysis reliability/validity are based on scores obtained via more recently developed quantitative approaches. PAQ and PDQ (Gomez-Mejia et al., 1979b) reliabilities were noted above, both of which utilize structured questionnaire approaches to data collection. Jenkins, Nadler, Lawler, and Cammann (1975) assessed analyst consistency when utilizing structured observation as the data collection method and found poor agreement regarding job effort and pressures, and moderate observer correspondence on a wide variety of other characteristics. More importantly, intercorrelations between structured interview and observation data only ranged form .16 to .48 on the characteristics. In short, the limited available data suggest substantial variation can be expected in job analysis results, contingent upon rater and/or data collection method.

Summary

Job analysis and job evaluation research findings are sketchy and often the product of flawed methodology, but even overlooking these shortcomings, they offer little encouragement regarding the feasibility of comparable worth evalu ation. Factor level reliability estimates - those most

germane to comparable worth evaluation - indicate a substantial amount of measurement error is probable. While direct evidence of bias in job evaluation is particularly scarce, research in other personnel decision areas suggests a significant level of systematic contamination must be anticipated. Factor analytic studies point to probable discriminant validity contamination if traditional job worth criteria are utilized. And finally, at the most general level, method comparisons outcomes seem to directly reflect method divergence. Put succinctly, little support for the concept of bias free evaluation can be found in the existing literature.

However, the caveat noted earlier bears repeating; measurement validity has not been a primary objective of job evaluation procedures. Thus it is possible that the psychometric properties of comparable worth evaluation systems will differ from those of traditional approaches, a possibility which provides part of the basis for the current study.

Finally, the existing literature does not address the criteria evaluation issue in job evaluation. The practical impact of various inter-rater reliability coefficients on wage assignment needs to be defined. Methodologies and indices for estimating bias and discriminate validity need to be tested and the notion of criterion deficiency in job evaluation deserves some initial exploration. These criteria

evaluation issues comprise a substantial portion of the agenda for the research reported here.

CHAPTER NOTES

1. For an overview of some of the variables see T. Connally, "Information Processing and Decision Making in Organizations" in B. M. Staw and G. R. Salancik, eds., New Directions in Organizational Behavior (Chicago: St. Clair Press, 1977) pp. 205-234, and Landy, F. J. and J. L. Farr, "Performance Rating," Psychological Bulletin, Vol. 87 (1980) pp. 72-107.

CHAPTER 4

Research Method

One conceptual approach to analysis of comparable worth job evaluation is to consider the possible sources of measurement variation in analysis of variance terms. In the typical job evaluation situation one or more raters (analysts) each evaluates an array of jobs on a number of dimensions (compensable factors). Thus raters, jobs and factors are all possible sources of variation (main effects) which could comprise a three-way ANOVA with one case (factor score) per cell. Furthermore, individual raters may respond in differing ways to specific jobs or dimensions, or may interpret the rating guides differently, thus generating interactions among the variance sources. Since measurement validity is defined in terms of the degree to which evaluated job worth reflects true job worth, relevant evidence regarding measure contamination is generated by determining whether job evaluation measures are sensitive to legitimate sources of variance (jobs, factors) and nonreactive to other sources (rater; rater interactions).

Construct validation also requires analysis of the degree to which measures from different methods converge because of the possibility of criterion deficiency noted earlier. Thus a fourth error source, measurement method must be

considered. If ratings of each job using each method are made by more than one analyst, the resulting model is a completely crossed four-way classification, the data collection design employed in this project. Four analysts each rated the same set of twenty jobs using three different instruments.

The four-way model does not accurately describe the analytical strategy pursued here, however. First, evidence of rating contamination is sought within each method by analyzing factor scores for various forms of rater inconsistencies. Three separate sets of intra-plan analyses are conducted. Second, factor scores are collapsed into summary job worth scores in the convergence (inter-plan) analysis and method differences in the resultant job rankings are examined.

The convergence analysis bears a strong similarity to repeated measures designs often used in experimental situations to increase precision (Dayton, 1970). Drawbacks to such designs include the possibility of carry-over effects from one treatment to another and the passage of time between treatments which could significantly change treatment conditions. The parallel concerns in the present non-experimental design might be labelled the familiarity effect from repeated use of the same instrument, the sequence effect which could occur if one instrument always precedes another, and interval effects which result from a significant time interval between application of the three methods.

Repeated use of an instrument can lead to subtle recalibration of scales which could be experienced differentially within methods thus affecting convergence assessment. Familiarity effect is assumed to be insignificant in his study since the scales were employed only twenty times distributed over a six week period and analysts were previously trained in scale meaning and interpretation. Undesired sequence effects were controlled through counterbalancing the evaluation schedule and interval effects were prevented by preparing all three evaluations of each job at the same session. Intervals between job analysis and evaluation did vary between jobs but such time lapse differences likely affected methods in a similar fashion.

JOB SAMPLE

The sample of positions studied and evaluated was drawn from the field staff and central administrative and support positions in a public sector information and education services organization. Field offices are located throughout Michigan serving the residents of each county with a wide variety of programs administratively grouped into four program delivery areas. The field office staffs range in size from 2 to over 100 people with the total field staff of 500 comprised of approximately 40% professionals, 40% paraprofessionals, and 20% management and clerical. The central office office staff includes administrative, program planning and development, top executive and clerical support personnel.

Since the focus of the research is on job evaluation measurement properties, the initial sample was selected to maximize diversity in job content and context. Location was hypothesized to be a source of diversity due to the need of county offices to respond to widely varying local conditions and clientele. Variability among counties in the type of services, program delivery modes, and political/social demands were anticipated. Therefore jobs in an urban, a rural, and a mixed urban-rural county were selected as sites in addition to a central office job sample. The expected area related differences were, of course, in addition to normal activity, function, and organization level differences.

A preliminary sampling scheme was developed based on the writer's estimates of the number of different jobs in the organization (25) and the average number of position analyses required to generate the requisite level of job understanding (3). The size and makeup of the actual sample was determined through continuing discussions within the evaluation team regarding (a) whether the jobs were distinct from others previously rated, and (b) whether additional position analyses were required to sharpen understanding of a particular job. The result of this process was a reduction in the final position sample from the estimated 75 position analyses and 25 job evaluations to 51 analyses and 20 evaluations. Determination of the positions which constituted separate jobs for evaluation purposes was made independently of formal titles. Thus a number of the evaluated jobs actually have

the same formal title but descriptive labels were used to distinguish among them (Appendix A).

DATA COLLECTION METHODOLOGY

Three of the four analysts employed in the project were staff support people in the organization already familiar with a large number of the positions. Two were male personnel department administrators responsible for the employment and compensation functions. The third, a female, was a member of the internal education institute serving the organization and personally familiar with many of the people and positions at all levels and in all locations. All three were volunteers sharing a common desire to increase their familiarity with actual field conditions and personnel. The fourth analyst, the writer, has conducted numerous job analyses in different organizations, but was essentially unfamiliar with this organization at the outset of the project.

Job Analysis

Job information was collected through incumbent and supervisory interviews, the method most frequently utilized in professional, administrative, and clerical pay administration (Belcher, 1974). An interview guide based on the information requirements of the evaluation systems was developed by the writer and tested by the analyst team in practice interviews during training sessions preceding data collection (Appendix B). Since one of the evaluation instruments was the Position Analysis Questionnaire (PAQ), the job components

analysis tool mentioned earlier, the interview content covered a broader range in greater detail than is typical of job analyses for pay purposes which usually focus only on the limited number of criteria in the evaluation plan.

The interview method was semi-structured, i.e., the interviewee was encouraged to talk about his/her job and the guide was used by the analysts to ensure that all areas were covered. Interview duration averaged approximately one and one-half hours and ranged from one to three hours. The normal interview sequence at a location was top down in order to obtain supervisors' views of subordinate positions as a part of their job analysis. The interview setting usually involved the four analysts plus the position incumbent. In three instances, in which a number of people occupied apparently similar positions (lower level jobs; same title; same supervisor), the analysts split into two teams to interview different people and subsequently discussed their respective findings. In all cases, extensive discussions about the jobs followed the interviews to increase the likelihood that the analysts developed a common job understanding.

Job Worth Measures

Three different job evaluation instruments were used to rate each of the twenty jobs: (1) the PAQ; (2) a commercially available plan widely used to evaluate management and technical jobs in both the private and public sectors (Standard Plan); (3) an in-house plan developed as phase one of the project (Custom Plan).

PAQ--the Position Analysis Questionnaire is a 187 item instrument focusing on worker oriented job variables which either measure directly or imply worker behavior. The developers argue that because the PAQ uses human behavior as the common denominator of work, it is therefore applicable to virtually all types of jobs. It is designed to encompass the entire repertoire of human behavior as it might be expressed on the job (McCormick, 1979).

The PAQ is probably the most advanced attempt to develop a taxonomy of universal work components and requirements, an approach to job analysis/evaluation which some feel is the most promising avenue to operationalizing the comparable worth concept (Jeannerette, 1980; Milkovich, 1981). The recommended implementation procedure involves analyst observation and/or interviews of job incumbents and rating of each position on each item utilizing one of the six PAQ scales (Figure 6).

PAQ ratings are converted to job worth measures through a "general" (and proprietary) regression equation developed originally by using job dimensions (defined empirically through principal components analysis of PAQ item scores as predictors of current earnings on a sample of 340 jobs from 45 organizations across the nation. (McCormick, Mecham, and Jeanneret, 1977).[1] The predicted worth of a job in this approach is thus a market derived value which, it is argued, is non-biased due to the representativeness of the sample. Even though market derived, the PAQ evaluation is intended to be a

I. Extent of Use

N Does not apply
1 Nominal/very infrequent
2 Occasional
3 Moderate
4 Considerable
5 Substantial

II. Importance

N Does not apply
1 Very minor
2 Low
3 Average
4 High
5 Extreme

III. Amount of time

N Does not apply
1 Under 1/10 of time
2 Between 1/10 and 1/3
3 Between 1/3 and 2/3
4 Over 2/3
5 Almost continually

IV. Possibility of Occurrence

N Almost none
1 Very limited
2 Limited
3 Moderate
4 Fairly high
5 High

V. Applicability

N Does not apply
1 Does apply

VI. Special

(19 different single item scales)

Figure 6. PAQ Scales

measure of relative, rather than actual, worth of jobs. Patterns of job relationships are considered to be quite stable over time and across geographic areas.

Standard Plan--The Standard Plan is modification of the evaluation plan utilized in the 1975 State of Idaho study reported in Treiman (1979). Three primary factors--knowledge requirements, problem solving requirements and accountability--are assessed by this instrument. Each of these factors has two or three dimensions or sub-factors. The plan provides an extreme amount of flexibility in rating jobs by combining the sub-factors into two and three dimension guide charts (point matrices) providing overlapping point value progressions for each possible combination of sub-factors. Figure 7 presents the Idaho study "knowledge" guide chart showing the manner in which the degrees of technical knowledge, managerial know-how and human relations skills are combined in a scoring matrix. Three distinct judgments are required to select the point score for the knowledge requirement of a job. Similar charts are used for the other factors.

The Idaho plan was modified in three ways for this study. First, the points within the knowledge and accountability matrices were reduced in equal proportion to facilitate summary score conversion to a common metric as required by the convergence analysis. Second, a fourth factor, working conditions, was eliminated. Working conditions are a constant within the organization studied and the relative

APPENDIX 15: The Hay System Guide Charts

STATE OF IDAHO
APRIL 1976

GUIDE CHART
KNOW-HOW
© HAY ASSOCIATES 1976

DEFINITION: Know-How is the sum total of every kind of skill, however acquired, needed for acceptable job performance. Know-How has 3 dimensions - - - the requirements for:

- • Practical procedures, specialized techniques, and scientific disciplines.
- •• Know-How of integrating and harmonizing the diversified functions involved in managerial situations (operating, supporting and administrative). This Know-How may be exercised consultatively as well as executively and involves in some combination the areas of organizing, planning, executing, controlling and evaluating.
- ••• Active, practicing, person-to-person skills in the area of human relationships.

MEASURING KNOW-HOW: Know-How has both scope (variety) and depth (thoroughness). Thus, a job may require some knowledge about a lot of things, or a lot of knowledge about a few things. The total Know-How is the combination of scope and depth. This concept makes practical the comparison and weighing of the total Know-How content of different jobs in terms of: "HOW MUCH KNOWLEDGE ABOUT HOW MANY THINGS."

•• MANAGERIAL KNOW-HOW

- I. NONE OR MINIMAL: Performance or supervision of a single activity, with appropriate concern for related activities.
- II. INTERMEDIATE: Primarily within single field or toward single objective with some integration of, or external integration with, other fields.
- III. BROAD: Integration and coordination of diversified activities in an important management area or consulting field.
- IV. COMPREHENSIVE: Comprehensive integration and coordination of diversified activities and functions in a major management area.
- V. MAJOR: Management at the level of policy making which affects the over-all operation of the state or which directs its most complex and far-reaching components.
- VI. TOTAL

••• Human Relations Skills →

• SCIENTIFIC DISCIPLINES / SPECIALIZED TECHNIQUES / PRACTICAL PROCEDURES

•	I. 1.	I. 2.	I. 3.	II. 1.	II. 2.	II. 3.	III. 1.	III. 2.	III. 3.	IV. 1.	IV. 2.	IV. 3.	V. 1.	V. 2.	V. 3.	VI. 1.	VI. 2.	VI. 3.	
A. PRIMARY Elementary plus some secondary (or equivalent) education; plus work indoctrination.	50	57	66	66	76	87	87	100	115	115	132	152	152	175	200	200	230	264	
	57	66	76	76	87	100	100	115	132	132	152	175	175	200	230	230	264	304	A
	66	76	87	87	100	115	115	132	152	152	175	200	200	230	264	264	304	350	
B. ELEMENTARY VOCATIONAL Familiarization in uninvolved, standardized work routines and/or use of simple equipment and machines.	66	76	87	87	100	115	115	132	152	152	175	200	200	230	264	264	304	350	
	76	87	100	100	115	132	132	152	175	175	200	230	230	264	304	304	350	400	B
	87	100	115	115	132	152	152	175	200	200	230	264	264	304	350	350	400	460	
C. VOCATIONAL Procedural or systematic proficiency, which may involve a facility in the use of specialized equipment.	87	100	115	115	132	152	152	175	200	200	230	264	264	304	350	350	400	460	
	100	115	132	132	152	175	175	200	230	230	264	304	304	350	400	400	460	528	C
	115	132	152	152	175	200	200	230	264	264	304	350	350	400	460	460	528	608	
D. ADVANCED VOCATIONAL Some specialized (generally nontechnical) skill(s), acquired on or off the job giving additional breadth or depth to a generally single function.	115	132	152	152	175	200	200	230	264	264	304	350	350	400	460	460	528	608	
	132	152	175	175	200	230	230	264	304	304	350	400	400	460	528	528	608	700	D
	152	175	200	200	230	264	264	304	350	350	400	460	460	528	608	608	700	800	
E. BASIC SPECIALIZED Sufficiency in a technique which requires a grasp either of involved practices and precedents; or of scientific theory and principles; or both.	152	175	200	200	230	264	264	304	350	350	400	460	460	528	608	608	700	800	
	175	200	230	230	264	304	304	350	400	400	460	528	528	608	700	700	800	920	E
	200	230	264	264	304	350	350	400	460	460	528	608	608	700	800	800	920	1056	
F. SEASONED SPECIALIZED Proficiency, gained through wide exposure or experiences in a specialized or technical field, in a technique which combines a broad grasp either of involved practices and precedents; or of scientific theory and principles; or both.	200	230	264	264	304	350	350	400	460	460	528	608	608	700	800	800	920	1056	
	230	264	304	304	350	400	400	460	528	528	608	700	700	800	920	920	1056	1216	F
	264	304	350	350	400	460	460	528	608	608	700	800	800	920	1056	1056	1216	1400	
G. SPECIALIZED MASTERY Determinative mastery of techniques, practices and theories gained through wide seasoning and/or special development.	264	304	350	350	400	460	460	528	608	608	700	800	800	920	1056	1056	1216	1400	
	304	350	400	400	460	528	528	608	700	700	800	920	920	1056	1216	1216	1400	1600	G
	350	400	460	460	528	608	608	700	800	800	920	1056	1056	1216	1400	1400	1600	1840	
H. PROFESSIONAL MASTERY Exceptional competence and unique mastery in economic, judicial, educational and/or political affairs.	350	400	460	460	528	608	608	700	800	800	920	1056	1056	1216	1400	1400	1600	1840	
	400	460	528	528	608	700	700	800	920	920	1056	1216	1216	1400	1600	1600	1840	2112	H
	460	528	608	608	700	800	800	920	1056	1056	1216	1400	1400	1600	1840	1840	2112	2432	

••• HUMAN RELATIONS SKILLS

1. BASIC	2. IMPORTANT	3. CRITICAL
Ordinary courtesy and effectiveness in dealing with others.	Understanding, influencing, and/or serving people are important, but not critical considerations.	Alternative or combined skills in understanding, selecting, developing and motivating people are important in the highest degree.

Reprinted by permission of Hay Associates, Washington, D.C.

Figure 7. Idaho Study Knowledge Chart

weight of the working conditions factor is negligible. Third, the accountability guide chart was modified by redefining (lowering) the "magnitude of accountability" category limits making them more relevant to the situation. In addition, the "impact" dimension of accountability was reduced to either "primary" or "contributory" to facilitate guidechart understanding with little or no loss in operational discriminability since more than 50% of the cell values overlap with adjacent cells in the matrix as designed. Appendix C contains a copy of the Standard Plan.

The standard plan was initially a three factor plan designed for high level position evaluation (Hay and Purves, 1951), but has been gradually extended to cover the full range of administrative and technical jobs. The guide charts are constructed to weight knowledge/skill requirements most heavily at the lowest job levels (over 50%). Accountability and problem solving assume increasing weight at higher levels until all three are weighted approximately equal for positions rated highly on all three factors.

The process by which point values for jobs are established using the standard plan involves a comparison of a set of jobs on each of the point matrices and selection of the values which best represent the judged relationships among them. The total of the points on the three factors is the point "worth" of the jobs.

Custom Plan--The decision to develop a point-factor custom plan was based on three considerations. First,

point-factor job evaluation is the most common method employed, a fact of potential importance to any subsequent findings. Second, comparable worth advocates typically utilize point-factor methodology to describe the development of an "unbiased" evaluation system. Third, point factor evaluation plans represent a different methodology than the PAQ and Standard plans, thus providing a three-way methods comparison.

Development of a point-factor evaluation plan is a lengthy process. Step one involves the definition of job worth in terms of the characteristics/demands of most importance to organization. The rating scales to measure the amount or degree of the factors in each job must then be developed. Subsequent steps involve a determination of the relative importance of each of the factors, assignment of points to the factors to reflect these differences (factor weights), and distribution of individual factor points along their respective scales. Finally, the plan must be tested for interpretability and consistency of understanding. (The traditional testing of results relative to existing internal or external key rates is, of course, omitted here). These design/development tasks were all performed in accordance with recommended procedures for the development of non-biased evaluation systems.

A five person evaluation plan development committee was formed consisting of people in each of the major program areas having both field and central office experience. Two of

the member were females, two minorities were represented (Black and Hispanic), and all members were middle level personnel with continuing exposure to all organizational levels. Thus a wide range of views regarding relevant compensable factors was represented on the committee. In addition to these five people, the writer served as group facilitator/consultant throughout their deliberations.

A modified nominal group process was followed in the initial factor selection stage in an attempt to provide adequate opportunity for diversity to surface. Individual lists of potential factors were generated and defined by each committee member. The factor lists were reviewed and debated by the entire committee over a series of sessions culminating in the identification of ten compensable factors (Appendix C). The factors identified and defined represent a committee consensus.

It should be noted at this point that the committee members viewed their activities primarily as contributions to a research project with only a low probability of application to the organization sometime in the future. Thus, their ownership or identification with the results of their effort was not as strong as would be the case if the outcome was destined to become operational. This fact could have had a bearing on the number and composition of the final factor set; committee members may have more readily yielded or modified their positions regarding relevant factors.

Factor scales were constructed through an iterative process of committee review and critique of proposed scales developed by the writer. After three such cycles, tests for consistency of understanding of the scales were then conducted by comparing individual committee members' appraisals of various benchmark positions. This, in turn, led to additions, deletions, and revisions of scale degree definitions.

As noted earlier, the statistical derivation of factor weights is potentially discriminatory through capturing existing bias in the criterion pay structure. Therefore weights were assigned on a judgmental basis, again through a consensus determination process. Individual committee members were asked to distribute 100 points across the ten factors on two separate occasions spanning a ten day interval. All ratings were then provided to the committee as a whole in order that they might consider their own inconsistencies (which were minor) as well as others' weighting schemes as they arrived at a final set of weights. The factor weights were then applied to the arbitrarily established plan maximum of 1000 points to establish the maximum factor point value. Points were distributed to the factor degree levels under the assumption of equal intervals between scale levels and using an arithmetic progression in accordance with typical job evaluation practice (Belcher, 1974).

Job Evaluation Procedure

The first phase in the evaluation process was analyst training in the use of the evaluation systems. The primary

training objective was to achieve common understanding of the rating scales among the four analysts. This was accomplished by a half-day scale familiarity session followed by two practice sessions involving comparisons and discussions of actual job ratings using the three systems. It should also be noted that the two personnel department analysts had previously attended a one-day workshop on the use of the PAQ.

Since the job analysis field work extended over a six week period, the job evaluations were prepared in four batches based on analyst team determination that the necessary level of understanding regarding particular jobs was achieved. As noted above, each analyst evaluated the 20 jobs using all three methods. In each case the PAQ was completed first under the rationale that its extensiveness and more descriptive (as opposed to evaluative) nature would serve as a stimulus to sharpen recall of the various job components, i.e., PAQ carryover effect would be desirable. The Standard and Custom method sequence was varied by rater and batch. Raters submitted PAQ response forms and summary evaluation sheets containing factor and job scores at the conclusion of each rating session.

The PAQ rating procedure varied from that recommended by its developers in two ways. First, the job analysis interview was not guided solely by PAQ concepts and items but included questions designed to generate information for the other two instruments as well. This departure should have been inconsequential, however, because the PAQ components

were considered in the interviews. Second, and potentially more important, in many cases the PAQ questionnaire was not immediately completed after the interviews but was filled out at a later date when the evaluation batch was submitted. In a 187 item instrument variability of rater recall could be a problem. However, the time lapse is not considered an issue in the present study because (1) only about 40% of the PAQ items were applicable to most of the jobs; (2) the time lag was typically days, not weeks, and (3) three of the four raters were very familiar with the jobs from their daily activities. Thus the PAQ results should be representative of any results obtained by strictly following recommended procedures.

DATA ANALYSIS PROCEDURES

The present research asks two general questions about analysts' ratings of job components and overall job worth. First, to what degree do the judgments under the three evaluation systems possess the measurement qualities critical to validity of the evaluations. Second, do the evaluation systems tap the same "job worth" construct, i.e., do they tend to converge? As noted earlier, these questions are central to the construct validation process. Thus the analytical techniques employed follow procedures for operationally assessing construct validity first demonstrated by Campbell and Fiske (1959) and later enhanced by others (Kavanaugh, MacKinnney and Wolins (1971), Kalleberg and Kluegel (1975) and Kane and Lawler (1979).[2]

The analytical procedures address both issues of deficiency and contamination while emphasizing the latter. Since a low level of measurement variance within each method is prerequisite to claims of convergent validity, the analysis first focuses on the measurement properties of the separate plans, and then moves to issues of relative qualities and method convergence.

Intraplan Analysis

Multitrait-Multimethod Analysis--Campbell and Fiske (1959) suggest four criteria which can be employed for determining construct validity when multiple psychological traits are measured by more than one method, and all the intercorrelations are arrayed in a multitrait-multimethod (MTMM) matrix. Evidence for convergent validity exists when the correlations of trait scores across methods (validity diagonal values) are significantly different from zero. Discriminate validity can be assessed in three ways. First, the pattern of trait interrelationships should be the same across all blocks. Second, the validity diagonal correlations should be higher than others in that block where neither trait nor method is common. Third, the validity diagonal value should be higher than measures of different traits employing the same method.

In the present job evaluation application of the MTMM scheme of analysis, the matrix is generated from the intercorrelations of the four raters' compensable factor scores (dimension scores) across the 20 jobs yielding a

multidimension-multirater matrix. Separate matrices are generated for the Custom and Standard plans. (The PAQ is not amenable to this type of analysis due to the size of a 187 dimension matrix). Since multiple raters are substituted for independent measurements in the matrix, the interpretation of the correlation values differs from the model presented by Campbell and Fiske (1959). The validity diagonal values in this case are intercorrelations of individual dimension scores between raters which are indicators of rater convergence (interrater reliability), not methods convergence.[3] Thus, these values are hereafter referred to as reliability diagonal values. However, the discriminate validity criteria are applicable. Intercorrelations on a particular dimension should be higher than those between different dimensions, whether by the same or different raters. Where the reverse is true, dimension independence is questionable.

While inspection of the multidimension-multirater matrix provides evidence of inter-rater reliability and dimension discriminability, it is an essentially subjective mode of analysis which does not provide a basis for statements regarding the probable types of measurement error reflected in the factor scores. Further, comparison of effects between plans is impossible. For these reasons Stanley (1961) and others (Kavanaugh et al., 1971, Kane and Lawler, 1979) have suggested going beyond MTMM inspection to empirical estimates of variance sources in the unreplicated three-way classification.

Analysis of Variance Technique--For many years the assessment of error variance in construct measurement has been guided by reliability theory which conceptually defines measurement in terms of inconsistencies reflected in two efforts to measure the same construct through maximally similar methods. Reliability coefficients are thus indicators of the level of systematic variance (which may include both "true" construct variance and systematic contamination). Reliability estimating procedures have focused heavily on three possible sources of inconsistency, the time interval between measures, the items utilized in the measurement scale, and the effect of judges or raters when used in measurement procedures. Of these three, only the last is directly relevant to this study.

Generalizability theory (Cronbach, Gleser, Nada, Rajartnam, 1972) utilizes the ANOVA framework for considering measurement error. The measurement conditions--raters, jobs, dimensions--are considered factors in an ANOVA design and standard test procedures can then be used to test for their effect on factor scores. In addition, the magnitude of variance due to each source can be estimated on the basis of the expected mean squares for the design. These estimates of variance components enable comparisons of the amount of variance due to each source to be made, and thus provide evidence regarding inter-rater reliability and/or causes of decrements in rater convergence (bias toward particular jobs, rater calibration differences, and factor structure shortcomings).

Indices reflecting the ratio of particular variance components to total variance (Katerberg, Smith, and Hoy, 1977) or to error variance (Kavanaugh et al., 1971) provide a means of comparing across studies, a capability which allows the issue of relative adequacy of evaluation methods to be directly addressed.

The ANOVA design most applicable to the typical job evaluation situation is one in which raters and jobs are considered random and evaluation dimensions are fixed. The normal ANOVA model is hypothesized to describe the data, i.e., an analyst's rating of a job on specified dimensions is the sum of the components--the overall mean, the individual main and interaction effects, and the unexplained (error) variance. The effects of interest here and their respective variance component and index formulas are presented in Table 1.

Evidence for inter-rater reliability (consistency of ratings between analysts) is found in the main effect for jobs which indicates agreement on job value across all raters and dimensions. This is, of course, subject to the usual caveat in ANOVA interpretation that the relevant interaction effects do not approach a comparable level of significance. If the rater by job interaction is of the same or greater magnitude, the evidence against would outweigh the support for the existence of inter-rater reliability. A significant rater by job interaction evidences one form of bias, a tendency on the part of one or more analysts to rate certain

Table 1. ANOVA Variance Component and Index Formulas

ANOVA EFFECT	Variance Component (V)	Intraclass Correlation
Rater	$\frac{MS_R - MS_{RxJ}}{dj}$	$\frac{V_R}{V_R + V_E}$
Job	$\frac{MS_J - MS_{RxJ}}{dr}$	$\frac{V_J}{V_J + V_E}$
Rater X Job	$\frac{MS_{RxJ} - MS_{RxJxD}}{d}$	$\frac{V_{RxJ}}{V_{RxJ} + V_E}$
Rater X Dimension	$\frac{MS_{RxD} - MS_{RxJxD}}{j}$	$\frac{V_{RxD}}{V_{RxD} + V_E}$
Job X Dimension	$\frac{MS_{JxD} - MS_{RxJxD}}{r}$	$\frac{V_{JxD}}{V_{JxD} + V_E}$
Error	MS_{RxJxD} *	--

Notes: 1. d, j, and r = the number of evaluation plan dimensions, number of jobs and number of raters respectively.

2. Formulas adapted from Kane and Lawler (1979; p459).

* The three-way interaction term is substituted here since no error estimate is possible in a non-replicated design, and it is reasonable to assume that this interaction is completely comprised of error.

jobs higher or lower. The intraclass correlation for job effect is also an indirect indicator of the extent to which this interaction may be significant since job effect is reduced by the amount of the interaction in its calculation.

In addition to bias toward certain jobs, rater evaluations may be inconsistent on specific job dimensions or across all dimensions. Control or elimination of these rater calibration problems is, of course, the primary objective of analyst training efforts. The ANOVA framework provides a means of evaluating the training in this respect through inspection of the rater by dimension interaction effects (dimension calibration) and the main effects for raters (leniency and severity).

Discriminant validity analysis in job evaluation focuses on the factor structure, addressing the question of whether sufficient differences between compensable factors exist to warrant considering them as distinct. If the factors measure different dimensions of worth, the job by dimension interaction should be significant. The practical significance of the interaction is reflected by the intraclass correlation index; if dimension distinctions are not greater than measurement error, the factor structure is clearly deficient.

In addition to the MTMM and ANOVA approaches to compensable factor structure analysis, principal components analysis and multiple regression procedures could have been utilized to assess the factor structure composition and relative importance of individual factors, respectively.

However, these procedures are subject to limitations present in this study. The small sample size effectively rules out principal components analysis of a ten variable plan such as the Custom plan; the 2:1 sample to item ratio is inadequate (Schwab, 1980). The three factor nature of the Standard plan renders principal components analysis pointless. With respect to regression procedures, factor interrelationships are the problematic element, a concern which will be clarified below in the discussion of results. Regression analyses are conducted, but only as a secondary and tentative approach to corroborating MTMM and ANOVA findings.

PAQ Reliability

Assessment of PAQ measurement properties focuses exclusively on indicators of the consistency of rater judgments. The derivation of PAQ components and job dimensions (e.e., the PAQ factor structure) have been thoroughly discussed elsewhere (McCormick, 1979; McCormick et al., 1972) and this standard factor structure is assumed acceptable here. The small sample size in the present PAQ application does not permit a meaningful analysis of the obtained PAQ dimensions.

Two separate reliability analyses of PAQ data are conducted. First, PAQ job analysis reliability is assessed by analyzing the 187 individual component ratings across the four raters to determine the nature and extent of any inconsistencies. Average component ratings are compared for evidence of rater calibration differences, and the percentage of similar ratings (within one scale point) between analysts on

each job is reviewed for fluctuations which could indicate rater bias or PAQ instrument inadequacies regarding particular jobs or job types. Component score reliability coefficients are also obtained for each pair of analysts on all 20 jobs--120 total correlations. The frequency distribution of the coefficients and the overall average provide evidence of PAQ job analysis reliability in a form directly comparable to PAQ reliability data reported previously (McCormick et al., 1977; 1972; Smith and Hakel, 1979).

Second, the reliability of PAQ jobs evaluations--the job worth values generated by the PAQ Inc. general regression equation--is estimated. These correlations of predicted values for the 20 jobs among the four analysts are directly comparable to the total score inter-rater correlations of the Custom and Standard Plans.

Interplan Comparisons

Three approaches to comparing the methods are utilized. First, as discussed earlier, the measurement properties of the Custom and Standard plans are compared by means of the variance component indices. Second, inter-method correlation and analysis of variance by ranks are used to estimate methods convergence and divergence respectively. Third, the outcomes of the three plans are converted to pay grades and assessed in terms of the similarity/differences of the resulting job hierarchies.

Interpretation of the differences in intraclass correlations (ICC) between the plans follows the same logic

discussed above regarding intraplan analysis. The ICC for jobs is considered an indicator of reliability and the indices for raters, rater by job, and job by dimension effects provide evidence regarding the relative likelihood of leniency, job bias, and discriminant validity in the two sets of measures.

The analysis of method variance is directed to the question of method convergence, i.e., to test the null hypothesis of no method difference in job worth measurements. Sums of Custom and Standard Plan factor scores, and PAQ predicted salaries provide the job worth measures. Since the three methods do not share a common metric, the first step in the analysis is a transformation of the Standard and PAQ job score distributions of each rater to correspond to the Custom Plan distribution in accordance with the procedure described by Nunnally (1967: 108). This linear transformation does not change the shape of the Standard and PAQ score distributions while facilitating direct comparisons.

The question of whether methods, per se, are a source of job worth score variance is similar to the common issue of the significance of differences in scores from k independent samples. If the analysis is limited to job worth scores of individual raters, the single source of job worth score variance is evaluation method. The three job worth score distributions can be tested to determine whether inter-method differences in scores by each rater signify genuine

differences (statistically significant) or whether they reflect chance variation.

A nonparametric one-way analysis of variance (Kruskal-Wallis) procedure is employed for two reasons. First, the job score transformation procedure equalizes means and dispersion of the three job score score distributions for each rater. Consequently the parametric F test is ruled out. Second, job evaluation is fundamentally ordinal measurement regardless of the apparent sophistication of some scaling techniques. This is particularly true of the PAQ which yields job worth scores indirectly derived from job analysis scales and clearly should not be treated as interval data. The Kruskal-Wallis technique converts the job scores to ranks across all three methods and determines whether the sums of ranks within methods are so disparate that method difference is probable. This technique is not as powerful as the parametric approach but if significant differences are found, the findings may be more readily generalized because questionable F test assumptions are avoided.

Classification Convergence--The use of total points as a job worth operationalization has a major shortcoming in that it does not reflect normal job evaluation practice which is to translate evaluation scores into specific classifications or job levels by means of a conversion table. This approach implicitly recognizes the essential subjectivity in job evaluation judgments and is thus the most defensible use of job evaluation scores. Therefore, methods convergence in terms

of the percent of jobs similarly classified is also calculated under differing classification conversion formulas.

The development of a job evaluation conversion table is an entirely subjective procedure involving two critical judgments. First, the number of classifications (pay grades) must be determined, a decision often based on a variety of administrative considerations in addition to concerns with equity. This decision also determines the width of the point ranges associated with particular classifications (inverse relationship). Second, the point values must be "pegged" to the classifications at one or more points and cutoffs established for each classification. Because of this subjectivity, both a nine and a twelve classification conversion table are used here to determine and compare intermethod "hit rates"--the percent of jobs identically classified.

The key question in convergence analysis is the degree to which correspondence of outcomes must be present to support a claim of measurement validity. Or, put negatively, at what point does divergence of results become unacceptable? A convergent validity criterion is needed, but the methods ANOVA provides only the statistical significance criterion. Classification convergence analysis provides possible operating criteria. One approach is to prespecify an acceptable degree of correspondence in the specific pay grade assignments generated by the systems under particular conversion equations. Another approach might utilize a "potential pay grade assignments" concept relating the job worth score range

of a measure plus and minus its standard error to the conversion table ranges. In either case the ultimate criterion is still a subjective assessment of the level of classification convergence which supports a conclusion of measurement validity. In the current research that level is arbitrarily assumed to be 80% under the simple logic that classification differences exceeding 20% of the cases provide overwhelming evidence of measurement deficiency or/and contamination.

In summary, the data analysis procedures of the study involve two phases. First, the measurement properties of the three methods are assessed primarily through interpretation of multi-rater correlation matrices and ANOVA findings. Second, the outcomes of the three evaluation methods are compared, both to make judgments regarding the relative quality of the obtained measures and to assess validity in terms of measurement convergence. Chapter 5 presents the findings of these analyses.

CHAPTER NOTES

1. The equation has recently been revised through an expansion of the sample size and updating of the salary information according to the developers, but specific information is not available at present.

2. The paper by Schwab (1980) is an excellent discussion of both the issues and procedures involved in construct validation.

3. The Kavanaugh et al. (1971) discussion of the convergent validity of performance appraisal ratings in misleading on this point. Rater agreement when utilizing the same performance assessment instrument is not evidence of convergent validity in the normal sense of the concept.

CHAPTER 5

FINDINGS AND DISCUSSION

Research results are presented here in accordance with the sequence of discussion in the preceding chapter. The initial focus is on measurement properties of the three evaluation plans, followed by presentation of interplan comparisons. The chapter concludes with a discussion of research limitations and issues raised by the findings.

MEASUREMENT PROPERTIES

Custom plan inter-rater factor and total score correlations exhibit the same pattern noted during the review of previous evaluation system reliability studies (Harding et al., 1960; Lawshe and Wilson, 1947). Job score correlations among raters are high, ranging from .93 to .97 as shown in Table 2, but rater intercorrelations on individual factors are often lower (Table 3). The average correlation is .75 on Custom plan factor 2 and .76 on factor 3 with pair correlations as low as .58.

Previous studies either have not addressed the issue of differing factor and total score reliabilities or simply suggested that the underlying cause was differences in the degree of abstraction among factors. Since rater inconsistencies indicate measurement error, factor score differences

Table 2. Custom Plan Job Worth Score Inter-rater Correlations

Rater	1	2	3	4
1	1.0			
2	.93	1.0		
3	.96	.94	1.0	
4	.98	.95	.97	1.0

Table 3. Custom Plan Factor Score Inter-rater Correlations (Reliability Diagonal Values)

	Rater Pairs						Avg.
Factor	1,2	1,3	1,4	2,3	2,4	3,4	
1	.90	.90	.87	1.0	.92	.92	.92
2	.79	.73	.78	.65	.74	.81	.75
3	.58	.93	.90	.61	.64	.88	.76
4	.86	.84	.89	.91	.81	.86	.86
5	.90	.89	.91	.86	.81	.89	.88
6	.75	.91	.80	.81	.71	.80	.80
7	.95	.95	.85	.91	.87	.88	.90
8	.76	.94	.87	.85	.80	.85	.85
9	.95	.93	.95	.93	1.0	.93	.95
10	.93	.87	.93	.81	.93	.90	.90

could be critical to comparable worth measurement where individual factor measures directly generate job worth values. Therefore, two questions need further investigation in the comparable worth job evaluation context. First, the primary determinants of variability in factor score inter-rater reliability need to be identified. This issue is partially addressed here through analyses of possible error sources in each method. Second, explanations are needed for the discrepancy between dimension (factor) score inconsistencies and summary score reliability coefficients. Possible answers to this issue are offered in the Standard plan comments below.

A review of analysis of variance values also suggests high Custom plan inter-rater reliability. Table 4 presents F ratios and estimates of variance components and intraclass correlations for relevant sources of rating variability. The strong main effect for jobs and negligible rater by job interaction attest to a high level of rater agreement on job and factor values and an absence of rater bias toward particular jobs. The low rater main effect and rater by dimension interactions provide additional evidence of consistency of rater judgments. These ANOVA results, when combined with the total score intercorrelation findings, are persuasive evidence of rater consistency in Custom Plan evaluations.

The situation with respect to instrument dimensionality is less positive. The suggested ANOVA test for factor independence, the job by dimension interaction (Kane and Lawler, 1969), indicates high discriminant validity. However,

Table 4. Analysis of Variance of Custom Plan Factor Scores

Source	df	MS	F	Variance Component	Intraclass Correlation
Rater (R)	3	144	0.8	.19	.00
Dimension (D)	9	51603	-		
Job (J)	19	13803	75.0*	340.6	.81
R x J	57	184	2.3*	10.4	.11
R x D	27	296	3.7*	10.8	.12
J x D	171	963	12.0*	220.7	.73
R x J x D	513	80	-	80.0	-

* .01 significance level

inspection of the MTMM matrix strongly argues to the contrary. Application of Campbell and Fiske's (1959) two primary discriminant validity criteria, the magnitude of reliability diagonal values relative to same rater-different factor correlations (Criterion 1) and different rater-different factor correlations (Criterion 2), leads to the conclusion that the ten factors probably collapse into 2-3 dimensions. Table 5 summarizes the comparisons in terms of the number of off diagonal values exceeding the average reliability diagonal values under both criteria for each factor pair. The maximum possible frequency is four (4 raters) under criterion one and 12 under criterion two (six rater pairs, 2 sets of ratings). It is clear from Table 5 that considerable commonality exists in factors 1 through 6 and distinctions among the financial responsibility factors (7, 8, 9) are minimal. In fact, the table may understate the degree of factor overlap due to the high reliability diagonal values. An inspection of the actual matrix (Appendix D) reveals that factor 1 (education) typically correlates highly with factors 2 through 6. Of 80 inter-correlations (same and different raters), only 6 are below .50 (five of the six involve one rater) and 33 exceed .75. Consequently, a strong argument can be made for the existence of one general knowledge type factor underlying the Custom plan factors which purport to assess requirements for knowledge, experience, human relations skills, teaching and communication skills, and the autonomy and impact of the job.

Table 5. Custom Plan Discriminant Validity Analysis Summary

Frequency of Factor Intercorrelation Values Exceeding a Mean Reliability Diagonal Value

Criterion 1: Same rater (maximum frequency = 4)

Factor	1	2	3	4	5	6	7	8	9	10
1	(92)									
2	1	(75)								
3			(76)							
4				(86)						
5	1	1	1	1	(88)					
6	1	2	2		2	(80)				
7							(90)			
8							3	(85)		
9								3	(95)	
10										(90)

(mean reliabilities on the diagonal)

Criterion 2: Different Raters (maximum frequency = 12)

1	(92)									
2	2	(75)								
3			(76)							
4				(86)						
5	5	3	3	1	(88)					
6	8	3	1	1	1	(80)				
7							(90)			
8						1	7	(85)		
9								5	(95)	
10										(90)

It is interesting to note that the issue which generated the most debate in the Custom plan development process, the distinction and definition of different dimensions of fiscal responsibility, proved of little consequence in the actual ratings. Table 6 presents the factor inter-correlations for each rater on factors 7-9. Quite obviously only one factor is reflected by these three scales. This overall fiscal responsibility factor, however, does appear to be somewhat distinct from factors 1 through 6. Rater inter-correlations between factors 1-6 and 7-9 are generally in the mid-range area of .25 to .60

Table 6. Interrater Correlations, Fiscal Responsibility Factors

	Rater 1		Rater 2		Rater 3		Rater 4	
Factor								
7								
8	87		93		92		85	
9	89	89	85	95	88	92	86	86

On a logical basis supervisory responsibility (Factor 10) could be expected to be closely aligned with fiscal responsibility but less strongly related to knowledge requirements. The average correlations of .66 and .50 between factor 10 and factors 7-9 and 1-6 respectively, support this expectation. The modest strength of these relationships may be an indication of a third minor factor operating to

distinguish job worth in addition to the general and fiscal responsibility factors.

Since MTMM matrix interpretation is essentially subjective, a stepwise procedure was used to regress summary job worth scores on factor scores in an attempt to cross-check the discriminant validity conclusions. In recognition of the fact that regression coefficients are unstable where a high degree of predictor collinearity is present, the purpose of this first regression analysis was limited to exploring whether separate analyses for each rater yielded 2-3 factor patterns similar to the conclusions of the MTMM analysis. If so, the concern about unstable regression results is significantly reduced and the factors identified can be considered proxies for the hypothesized underlying dimensions.

Table 7 lists the first four factors included and their incremental contribution to explained variability for the regression equations derived from each of the four sets of ratings. In each case factor one is the predominant predictor followed in 3 of the 4 equations by a fiscal responsibility factor. Since factor 10 also was included in three of 4 cases, these results were interpreted as corroborative of the subjective analysis.

The simplified three-factor structure was then assessed by means of a second set of regression analyses employing only factors 1, 7, and 10 as proxies for general knowledge, fiscal responsibility and supervisory responsibility dimensions. Consistency of findings in degree of variability

Table 7. Factors Selected Through Stepwise Regression of Custom Plan Summary Job Scores on Factor Scores.

	Order of Factor Inclusion	Incr. R^2		Order of Factor Inclusion	Incr. R^2
Rater 1	1	.59	Rater 3	1	.73
	7	.29		7	.13
	2	.01		3	.10
	3	.07		10	.002
		.96			.962
Rater 2	1	.70	Rater 4	1	.68
	3	.17		8	.21
	2	.01		4	.02
	10	.05		10	.02
		.93			.93

Table 8. Regressions of Summary Job Worth Scores on Simplified Factor Structure.

	Rater 1	Rater 2	Rater 3	Rater 4
Beta				
Factor 1	.55	.51	.60	.62
7	.36	.28	.26	.41
10	.29	.37	.28	.17
R^2 Change				
Factor 1	.59	.70	.73	.69
7	.30	.17	.13	.24
10	.04	.07	.04	.01
	.93	.94	.90	.94

explained (R^2) and relative importance of the factors (Beta weights) among the four sets of ratings provides a reasonable basis for inferences regarding an underlying (simplified) factor structure. Furthermore, the problem of collinearity is significantly reduced with this factor structure. The results of this analysis are presented in Table 8. In all four cases the percent of variance explained by the simplified factor structure is 90 or above with the knowledge factor the most important predictor. The regression weights for the fiscal and supervisory responsibility factors are significant in each of the equations but they are considerably lower than the knowledge factor and vary in relative importance across the four samples. Furthermore, the incremental contribution of the supervisory factor to explained variance is quite low. Thus the three factor interpretation of the Custom plan evaluation scores is only weakly supported by these results but a two factor view is strongly supported.

Interpretation of the Custom plan factor structure findings can take two basic directions. On the one hand the assumed multidimensionality of job worth constructs can remain unchallenged with the current findings attributed to measurement deficiencies. In this view the Custom plan is clearly an inadequate evaluation tool due to its discriminant validity shortcomings, and future efforts should focus on developing new and improved measures of the "other" dimensions. On the other hand, results here may be interpreted to indicate the necessity to redefine the

underlying job worth construct, perhaps as a more internalized and unitary job worth standard akin to the thinking of Jaques (1961). From this perspective the findings are more likely seen as contaminated than deficient, and the appropriate remedial strategy is one of reducing and refining measurements. For example, the simplified factor structure might be a sufficient measure of job worth. In short, interpretations of the findings can stress either the necessity for construct redefinition or for measurement improvement, and logical support can be generated for either position. At this early stage of job worth measurement research, additional effort in both directions is probably appropriate.

Standard Plan Measurement Properties

Analyses of the job worth measures generated by means of the Standard Plan raise significant questions regarding their adequacy. Inter-rater reliability coefficients suggest considerable inconsistency in rater assessments of total job worth (Table 9) and even greater differences regarding factor scores (Table 10), an assessment which is supported by the Standard Plan ANOVA results. The highly significant job main effect in (Table 11), which would normally indicate a strong degree of agreement on job worth among the raters, is severely qualified by the presence of both a significant rater by job interaction and rater main effect. The interaction reflects inconsistency between raters on a job or jobs; rater effect indicates leniency or severity error on

the part of at least one rater. Thus it is reasonable to conclude that rater differences in the application of the Standard Plan to the job sample were significant.

The factor structure of the Standard Plan is also called into question by the findings. Since the Plan involves only three factors, and two of these are similar in definition to

Table 9. Standard Plan Summary Job Worth Score Score Inter-rater Correlations

Rater	1	2	3	4
1	1.0			
2	.72	1.0		
3	.77	.80	1.0	
4	.88	.93	.90	1.0

the general and fiscal factors suggested by the Custom plan results, a finding of factor independence could be reasonably anticipated. However, the matrix presented in Table 12 indicates that such is not the case. Reliability diagonals (underlined) are often exceeded by correlations between that factor and the remaining two in both the within and between rater triangles. No discernible pattern of factor intercorrelations exists other than the fact that if rater one is excluded, the lowest factor intercorrelation among the other three raters is .71 and the average is .83.

The ANOVA results also point to a lack of discriminant validity. The job by dimension interaction is modestly

Table 10. Standard Plan Factor Score Inter-rater Correlations (Reliability Diagonal Values).

	Rater Pairs						
Factor	1,2	1,3	1,4	2,3	2,4	3,4	Avg.
1	76	66	78	74	82	89	78
2	53	56	70	83	90	87	73
3	65	90	77	71	87	80	78

Table 11. Analysis of Variance of Standard Plan Factor Scores.

Source	df	MS	F	Variance Component	Intraclass Correlation
Rater (R)	3	11253	5.6*	153	.23
Job (J)	19	30087	14.9*	2339	.82
R x J	57	2014	3.9*	501	.49
R x D	6	792	1.6	14	.03
J x D	38	2113	4.2	405	.44
R x J x D	114	510	-	510	-

*.01 significance level

Table 12. Standard Plan Rater by Factor Correlations

		Rater 1			Rater 2			Rater 3			Rater 4		
Factors		1	2	3	1	2	3	1	2	3	1	2	3
R_1	1	1.0											
	2	91											
	3	76	65										
R_2	1	76	69	61									
	2	56	53	53	93								
	3	64	70	65	76	88							
R_3	1	66	71	63	74	76	69						
	2	50	56	44	83	83	75	98					
	3	79	83	90	73	75	71	85	86				
R_4	1	78	83	65	82	86	84	89	83	79			
	2	62	70	50	87	80	93	90	87	79	97		
	3	75	77	77	75	79	87	78	75	80	78	82	1.0

significant but the variance component estimates suggest that measurement error exceeds dimension distinctions. Thus a particular factor measure could be reflecting any of the three (or other) constructs.

Explanations for the poor properties of the Standard Plan measures obtained in this study might center on the quality of the rating process in this particular application. Certainly rater leniency points to a training/learning deficiency, and commonality among the three factors could reflect inadequate "job profiling" in the rating process. Rater inconsistencies also could indicate that analysts were not sufficiently skilled in the nuances of the system's semantics, a necessity strongly emphasized by the vendor. However, it is the writer's opinion that the ultimate sources of the apparent Standard Plan measurement inadequacies are the method per se, the comparable worth application, and the nature of the organization.

As noted earlier, the subject organization's mission is information dissemination and training. Job differences revolve heavily around the type and depth of information services required by varying clients. Most of the jobs have knowledge dominated profiles and the knowledge requirements factor rating establishes the base for problem solving-decision making ratings which are expressed as a percentage of knowledge. Furthermore, financial accountability in the organization directly reflects the scope of educational program area activity, i.e., it corresponds to the breadth

and depth of knowledge required. Thus the lack of factor discriminant validity in the Standard Plan can be rationalized to some degree in terms of the type of organization studied.

As mentioned above, degree of abstraction or factor ambiguity is a plausible explanation for low factor reliabilities. The Standard Plan provides an excellent example of this possibility. Highly abstract factor and degree level definitions are necessary for a broadly applicable system, but the resulting ambiguity is a serious problem when the method is used in a comparable worth mode. Degree levels within Standard Plan factors can be interpreted consistently only through comparisons with other jobs, a fact which reflects the plan's origins as a factor comparison system. The degree definitions are too vague to function as absolute rating scale anchors and multiple point values within each degree compound the scoring problem. Benchmark job values for degree levels within factors are absolutely necessary to anchor the scales. However, benchmarks are inappropriate for comparable worth job evaluation for reasons previously noted. Thus benchmark jobs were not systematically identified for the Standard Plan guide chart levels in this study. The predictable consequence of this situation was rater confusion at the onset of the rating process until a framework of job levels evolved through experience and rater discussions. In short, over a period of time the Standard Plan tended to

reassume its job to job comparative nature reflecting personal and organizational norms of job worth.

Organizational and personal norms of job worth also may be a part of the explanation for differences between factor and summary score reliability coefficients in point factor plans. Such norms very likely operate in the rating process to reduce potential differences in summary job ratings. Raters adjust initial sets of factor scores to generate summary scores more closely conforming to their perceived acceptable job hierarchy. In comparable worth job evaluation this "adjustment" process is difficult initially since benchmark jobs are not used to help define scale values. However, after an initial array of jobs has been evaluated, raters can use the obtained job worth values as reference points in subsequent evaluations. New summary job worth scores are tested for their legitimacy against the values in the emerging job worth hierarchy. Where the results do not "make sense," factor scores are revised somewhat randomly to achieve the desired end result resulting in greater agreement between summary scores than individual factor scores. Through this process, point-factor systems, even the comparable worth variety, tend to devolve into whole job ranking or slotting procedures rationalized in terms of factor scores. This deviation from intended job rating procedures is not unusual among practitioners and was partially confirmed in post-rating discussions with the three raters from the participating organization. Because this was a research

project, the raters did not consistently strive for agreement between personal norms and evaluation results; the study had no immediate implications for pay distribution and curiosity about method comparisons tended to serve as a counteracting force.

PAQ Reliability

Two sets of PAQ reliability estimates are discussed here. Inter-rater reliability coefficients based on pairs of analyst ratings across all 187 job element provide a means of comparing the quality of this particular PAQ job analysis application to other reported studies. The reliability of predicted job worth values then provides the basis for the methods comparison and convergence aspects of the present study.

Table 13 displays the frequency distribution of reliability coefficients obtained from all pairs of analyst ratings (20 jobs, 6 analyst pairs). The distribution and the average coefficient are very similar to the results obtained in other PAQ studies (averages noted at bottom of Table 13). This suggests that the slight departures from recommended PAQ analysis procedures utilized in the present research did not materially affect the outcome. Whether the level of reliability is adequate is a separate question, the answer to which depends upon the ultimate application of the measure. Norms for job evaluation have not been developed, an issue which will be addressed below. Further, the global judgments involved in job worth assessments are likely to cancel out some of the apparent inconsistency of analysis. Thus,

Table 13. Frequency Distribution of Reliability Coefficients for PAQ Analyst Pairs.

Reliability Coefficient Interval	Number	Percent of Total
.90 - 1.0	-	
.80 - .89	6	5.0
.70 - .79	58	48.3
.60 - .69	46	38.3
.50 - .59	10	8.3
	120	100

Average Reliability Coefficient: .70

Previous Studies:

McCormick et al. (1972)	.74
Taylor (1977)	.68
Smith and Hakel (1979)	.63

Table 14. Summary of Rating Means by Analyst

	Analyst			
	1	2	3	4
Average Analyst Item Scores - all jobs	.92	1.21	1.29	1.22
Frequency of Avg. item score:				
..exceeding job avg.	1	15	17	11
..below job avg.	19	5	3	6
..at job avg.	0	0	0	3

whether the job analysis reliability coefficients reported here are adequate depends upon their relationship to evaluation reliabilities and upon a determination of the requisite level of job evaluation reliability.

Tables 14 and 15 provide some insight into the nature of rater disagreements in the PAQ analysis process. The average scores displayed in Table 14 suggest similarity in use of the PAQ scales by analysts 2, 3, and 4. However, analyst 1 shows evidence of calibration error (severity). Since the PAQ scales include a "non-applicable" option, the consistently lower item scores could also reflect lower analyst 1 motivation to rate or perceptual differences regarding relevant items.

Table 15 lists the percentage of similar ratings among analysts for each job calculated by relating the number of item ratings within one point of other analysts' ratings to the total number of rating comparisons. The primary purpose of this analysis was to determine whether the degree of rater agreement varies appreciably between jobs. Less rater agreement would logically be expected regarding the more complex jobs, and a finding of job related rater inconsistency raises questions about the quality of the analysis process and/or the general applicability of the PAQ. However, no discernible pattern emerged. Relatively simple jobs were among the highest (secretary) and lowest (supervisory aide) percentages of similar ratings and the complex position of an urban

Table 15. Percentage of Individual Analyst Job Component Ratings within one Scale Point of Other Analysts' Ratings

Job	Analyst				Average
	1	2	3	4	
1	81	84	75	83	81
2	85	83	84	87	85
3	83	83	79	83	82
4	80	81	75	79	79
5	84	84	83	84	84
6	83	80	82	86	83
7	84	82	82	83	83
8	80	76	76	82	78
9	83	84	80	87	83
10	83	82	77	82	81
11	80	83	72	80	79
12	81	76	72	82	78
13	83	81	82	84	83
14	82	81	85	83	83
15	79	81	76	81	79
16	74	74	73	78	75
17	82	83	82	83	83
18	80	78	75	79	78
19	87	88	87	86	87
20	87	86	86	87	87

director had the second highest percentage of similar ratings.

The primary message of Table 15 is actually the degree of general disagreement it reflects. Of the 187 PAQ items, 16 are dichotomous and approximately 35-45% of the remainder were not applicable to the jobs studies. Of the remaining items (Approximately 100), all measured on 5 point scales, rater pairs disagreed by 2 or more scale points about 20% of the time. This points strongly to the advisability of a multiple rater approach using averages or some form of inter-rater reconciliation process in any applications of PAQ job analysis.

As stated stated above, rater inconsistencies in analysis of jobs are not likely to be reflected to the same degree in job worth assessments. Some of the variance in job characteristics is of little or no consequence to determinations of relative worth, whether obtained directly through human judgment processes or indirectly through regression analysis. Table 16 evidences this fact. The interrater reliability coefficients of PAQ generated job values range from .93 up, roughly equivalent to those of the Custom plan summary job worth scores.

Table 16. Inter-rater Correlations of PAQ Generated Job Evaluation Scores.

Rater	1	2	3	4
1	1.0			
2	.95	1.0		
3	.94	.94	1.0	
4	.98	.93	.93	1.0

METHOD COMPARISONS

The findings reported in this section address the issue of evaluation method convergence. As noted earlier, convergence of independent measurement methods provides evidence of measurement validity. Evaluation plan outcomes can be expected to diverge to the degree they reflect different levels of measurement error and/or consider compensable factors dissimilar in content or relative importance. Thus, the comparisons reported here consider the relative measurement properties of the three methods as well as the degree to which the job worth scores of the methods converge.

Tables 17 and 18 summarize the previously discussed evidence relating to the relative measurement qualities of the methods. The lower reliability coefficients of the Standard Plan (Table 17) indicate the presence of a greater degree of error and the variance component indices presented in Table 18 suggest possible types of error. All three variance

Table 17. Comparison of Inter-rater Reliability Coefficients: Custom, Standard and PAQ Job Worth Scores.

Rater Pairs	Custom	Standard	PAQ
R_1R_2	.95	.72	.95
R_1R_3	.96	.77	.94
R_1R_4	.98	.88	.98
R_2R_3	.94	.80	.94
R_2R_4	.95	.93	.93
R_3R_4	.97	.90	.93

Table 18. Comparison between Custom and Standard Plan Measurement Properties.

Property	Variance Component	Intraclass Correlation	
		Custom	Standard
Job Bias (Halo)	R x J	.11	.49
Leniency/ Severity	R	.00	.23
Discriminant Validity	J x D	.73	.44

indices are consistent with the reliability coefficients indicating less Custom plan bias in the form of halo and leniency effects and greater discriminant validity (keeping in mind that both plans apparently suffer from serious discriminant validity shortcomings). While additional comparative data on the PAQ measure is not available it is reasonable to assume, based on reliability coefficients, that PAQ job evaluations are also superior in measurement characteristics to Standard Plan outcomes. Thus based on relative measurement properties only, divergence in results between the Standard Plan and both the Custom plan and PAQ methods should be expected, and any claims of convergent validity are correspondingly diminished.

Method Convergence

Assessment of method convergence was conducted on three levels. First, statistical tests for convergence (correlation) and divergence (ANOVA) were utilized. Second, raw job scores were ordered and graphically inspected for convergence. Third, raw scores were translated into pay classification levels and classification convergence was assessed.

Table 19 presents the inter-method correlations for all rater pairs. The total range of correlations spans from .69 to .95. Standard-PAQ r's of .68 to .89 dominate the low end of the range. Custom-Standard r's range from .81 to .95, and Custom-PAQ values cluster at the higher levels (.87 - .93). The latter two sets of values are similar to the findings reported by Robinson et al. (1974) of .82 to .89 and the .89

Table 19. Job Worth Score Multimethod-Multirater Correlations.

Rater		1			2			3			4		
Method		C	S	P	C	S	P	C	S	P	C	S	P
1	C												
	S	84											
	P	91	78										
2	C	-	81	88									
	S	87	-	71	92								
	P	91	76	-	92	79							
3	C	-	89	90	-	84	91						
	S	84	-	87	85	-	76	91					
	P	92	81	-	91	77	-	95	81				
4	C	-	87	93	-	84	92	-	87	92			
	S	93	-	86	95	-	89	93	-	89	93		
	P	90	73	-	87	68	-	90	77	-	92	82	

Table 20. Kruskal-Wallis Analysis of Job Evaluation Method Variance by Rater.

Rater	Method	Sums of Ranks	H
1	1	616.0	0.14
	2	587.0	
	3	627.0	
2	1	617.0	0.11
	2	589.5	
	3	623.5	
3	1	623.0	0.39
	2	571.0	
	3	636.0	
4	1	613.0	0.14
	2	588.0	
	3	629.0	

to .97 range reported by Chesler (1948) which were interpreted as indicative of high evaluation score convergence. However, major discrepancies in job worth determinations can still exist. Additional evidence is necessary before convergence can be claimed or rejected.

Where correlations reflect strength of method convergence, analysis of variance provides a means to test for divergence. Individual rater evaluations of the 20 jobs were tested to determine whether method differences are statistically significant. The statistic (H) in the Kruskal-Wallis non-parametric one-way analysis of variance by ranks is distributed as chi square with k-1 degrees of freedom (Siegel, 1956). An H value of 5.99 or greater is necessary to reject the null hypothesis of no method differences at the .05 level of significance. However, Table 20 indicates that the largest obtained H value of the four sets of evaluations is 0.39. Thus no statistically significant difference was found in job ranks generated by the three methods.

Neither high inter-method correlations nor the negative analysis of variance results address the issue of the practical significance of method differences. Evaluation discrepancies of potential importance to the parties involved are still possible. In fact, the frequency and degree of such differences provides the most concrete basis for judgments regarding method convergence. Thus graphic comparisons of the evaluations were developed and the potential impact of job worth score differences on pay classifications was analyzed.

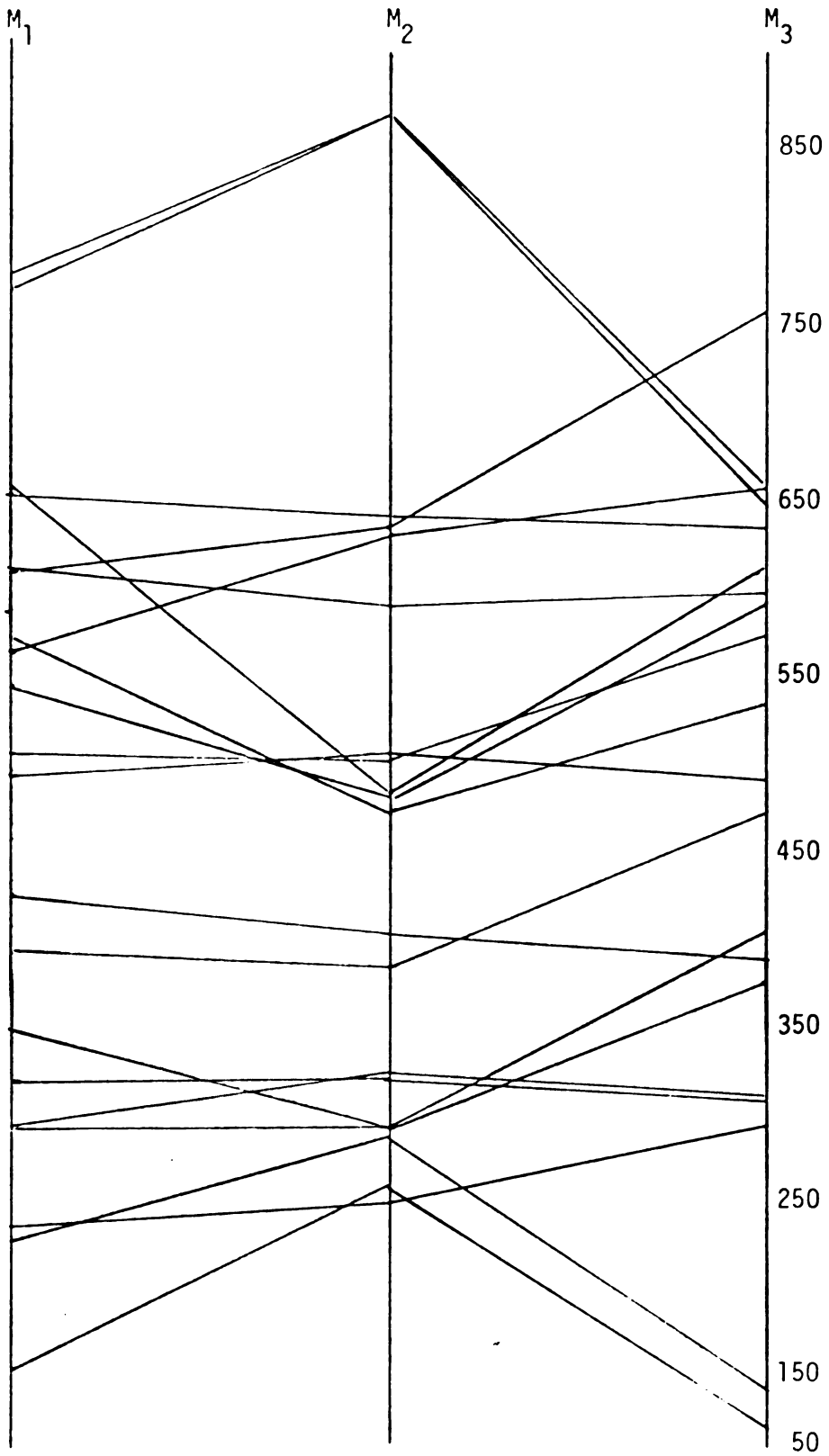


Figure 8. Relative Rank of Jobs Using Three Methods - Same Analyst.

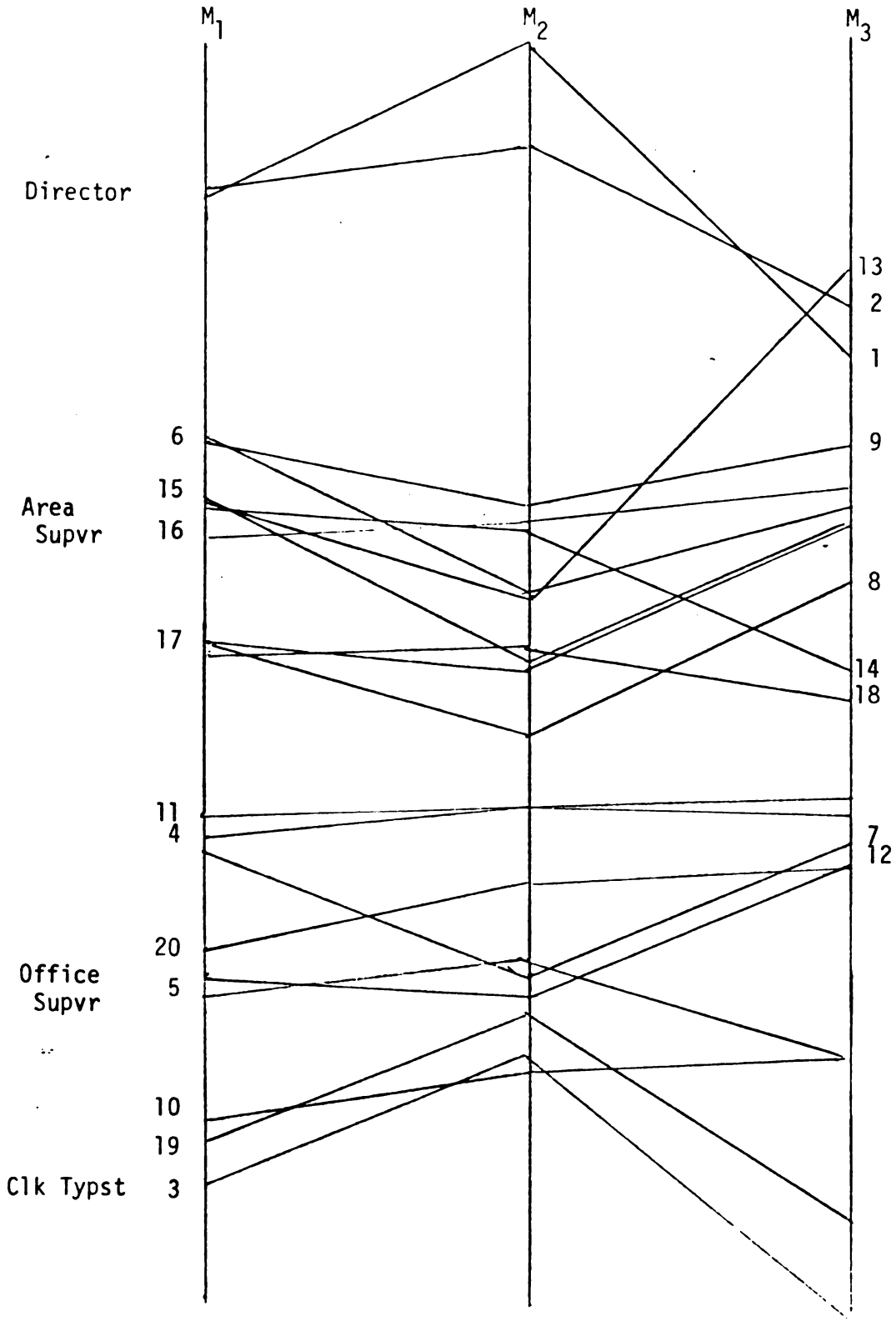


Figure 9. Relative Rank of Twenty Jobs Using Three Evaluation Methods: Rater Mean Scores.

As noted earlier, the job score transformation procedure facilitates direct comparison of the three distributions. Figure 8 provides an example of such a comparison for the ratings of analyst number four. Each line connects the job worth values under the three methods. Although the low H value for this set of ratings indirectly suggests a high level of method convergence, and the intercorrelation values of .93 (Custom-Standard) and .92 (Custom-PAQ) also support a judgment of very high convergence, Figure 9 leads to a different conclusion. Complete convergence would yield twenty parallel lines, but the diagram is characterized by intersections, each indicating a rank difference between methods. When the ratings of all four analysts are averaged (Appendix F), a procedure often recommended to avoid idiosyncratic judgments (Schwab, 1980), the diagram reflects greater parallelism (Figure 9) but numerous rank differences remain. Job number 13, for example, is ranked 6, 7, and 1 on the Custom, Standard, and PAQ systems respectively. Table 21 presents the correlation coefficients for these averaged ratings.

Table 21. Inter-method Correlations of Averaged Job Scores

Custom - Standard	.95
Custom - PAQ	.95
Standard - PAQ	.85

Differences in job worth score rankings do not automatically convert to pay differences with the rare exception of those organizations which directly translate points to dollars. In most cases an interim step takes place in which jobs are grouped into pay grades or classifications. Thus a key question is whether raw score rank differences are moderated when pay classification systems are utilized.

Classification Convergence -- Classification convergence results cannot be discussed without first considering the concept of standard error of measurement (SE_M) as it might apply to job evaluation. Job evaluation (observed) scores reflect both job differences and error of measurement, thus any single observed score is an estimate of the true worth of the job, and repeated evaluations around the true score. The standard deviation of the observed scores around the true score is called the standard error of measurement. Since the true job worth score is unknown, an exact value of the standard error cannot be computed but an estimate can be calculated from the standard deviation of the job score distribution and the reliability coefficient. The formula is simple:

$$SE_M = SD\sqrt{1-r_{tt}}$$

where SD is the standard deviation of the obtained scores and r is the reliability coefficient.

As the NAS committee correctly noted, a crucial question when considering the measurement qualities of job evaluation

is the probable variability in points assigned to any given job (Treiman, 1979:41). The example provided by the committee--a calculation of standard error in terms of labor grades--is erroneous since the degree to which measurement error can produce different grade assignments depends upon the range of job worth scores per grade (which in turn is a function of the number of pay grades). Point score variability should first be estimated and then related to grade or classification assignment variability under differing conversion tables.

Table 22 presents estimates of the Custom Plan standard error and the potential range of measurement variability using a 95% confidence limit for reliability coefficients of .90 and .95. If we assume for the moment that the 1000 point Custom plan is equated to a ten classification system (100 points per grade), the table suggests that grade assignment variability could be as high as 4 grades when r = .90 and 3 grades given r = .95. Table 23 uses the average job scores under the three methods and the point variability assumption of ± the average SE_M(42 points) to generate potential pay grade assignments using a 12 level classification scheme (see Table 25). Classification convergence is possible when one of the potential pay grade assignments is common to all three evaluation methods. In this case a 50% convergence rate is obtained.

Potential classification convergence can be increased by employing more conservative (broader) interval estimates of

Table 22. Standard Error of Measurement and Range of .95 Confidence Interval by Rater for .90 and .95 Reliability Coefficient Values.*

	SE_M r = .90	Conf. Intv. Range	SE_M r = .95	Conf. Intv. Range
Rater 1	58.1	228	41.1	161
Rater 2	57.2	224	40.4	159
Rater 3	63.9	250	45.2	177
Rater 4	58.5	229	41.3	162
Average	59.4		42.0	

* $SE_M = SD\sqrt{1 - r}$; Standard deviation (SD) value used is that of the Custom Plan Ratings, the base distribution for the common metric. Interval range = $2(1.96 \times SE_M)$.

Table 23. Potential Pay Grade Assignment Convergence of Three Evaluation Methods on Twenty Jobs.*

Job	Potential Pay Grade Assignments (1-12)			
	Custom	Standard	PAQ	Possible Convergence
1	10-12	12	9-10	no
2	11-12	11-12	10-11	yes
3	1-2	3-4	1	no
4	4-6	5-6	5-6	yes
5	3-4	3-5	3-4	yes
6	9-10	7-8	7-9	no
7	4-5	3-4	4-6	no
8	6-7	6-7	7-8	yes
9	8-10	7-9	8-10	no
10	2-3	3-4	3-4	yes
11	5-6	5-6	5-6	yes
12	3-5	3-4	4-5	yes
13	8-9	7-8	10-11	no
14	7-9	7-9	6-7	no
15	8-9	6-7	7-9	no
16	7-9	7-9	8-9	yes
17	6-7	6-7	7-9	no
18	6-7	6-8	6-7	yes
19	2-3	3-4	1-2	no
20	4-5	4-5	4-5	yes

*Potential pay grade assignments are the grades associated with point values within the range of scores represented by the mean method score $\pm$ average SE_M ($r = .95$).

job worth. However, the utility of the evaluation system for making job distinctions is correspondingly reduced and the necessity for subjective assignment to one of the potential classifications associated with the interval is increased. For example, the 84 point intervals illustrated in Table 23, a range of approximately 10% of the maximum point score, equate to two classifications in most instances. Consequently evaluation measures actually distinguish only six classification levels; final grade level distinctions would be subjectively determined.

Classification convergence is also, of course, a function of the number of pay levels. Tables 24 and 25 present pay grade assignments under 9 and 12 level conversion tables using mean job scores (point estimates) for the three methods. The degree of convergence is summarized in Table 26. The three methods generate the same classification in only three of the twenty jobs when 12 levels are employed and in six cases under the 9 level table. Method pair convergence does not exceed 40% and 60% in the 12 and 9 level schemes respectively, considerably below the 80% criterion suggested above.

In summary, the method comparison findings do not foster optimism regarding job evaluation system results as objective measures of job worth. First, the adequacy of evaluation plans in terms of measurement properties can apparently vary significantly. Whether sufficient improvement in the different approaches to measuring job worth can be developed is un-

Table 24. Pay Grade Assignments of Twenty Jobs under Three Evaluation Methods - Twelve Level Conversion Table.

		Job Assignments (Job No.)[1]		
Grade	Points	Custom	Standard	PAQ
12	820-900		1,2	
11	750-819	1,2		13
10	680-749			1,2(13)[2]
9	610-679	6,9,(13),15		9,16
8	560-609	13,14,16	(6)9,14,16	6,(8)15,17
7	490-559	8,17,18	6,13,15,17,18	8,14(18)
6	420-489	(11)	8,11,(17)	11,18
5	350-419	4,7,11	4,(11) 20	4,7,(11)12,20
4	280-349	5,12,20	5,7,12,19	
3	210-279	10(19)	3,10,(19)	5,10
2	140-209	3,19		19
1	70-139			3

[1] Mean Scores, four raters

[2] Marginal cases (within 7 points of adjacent grade) indicated by double listing - secondary classification in parentheses.

Table 25. Pay Grade Assignments of 20 Jobs Under Three Evaluation Methods[1]--9 Level Conversion Table[2].

		Job Assignments (Job No.)		
Grade	Points	Custom	Standard	PAQ
9	850-950		2 (2)	
8	750-849	1,2,	2	13
7	650-749	6,(9)		1,2,(9),(13)
6	550-649	(6),9,13,14,15,16	6,9,14,16,(13)	6,8,9,15,16,17
5	450-549	8,17,18	8,13,15,17,18(6)	(8)14,18
4	350-449	4,7,11	4,11,20	4,7,11,12,20
3	250-349	5,12,20	3,5,7,10,12,19	5,10
2	150-249	3,10,19	(3),(10	19
1	50-149			3

1 Mean Scores, four raters.

2 Marginal cases (within 10 pts of adjacent grade) indicated by double listing with the potential classification in parentheses.

Table 26. Pay Grade Assignment Convergency Among Three Methods Under 9 and 12 Level Conversion Tables.

	Number of Job Assignments		
	Same Grade	± One Grade	± Two or More Grades
A. Nine Levels			
Methods:			
Custom-Standard	11	9	-
Custom-PAQ	9	10	1
Standard-PAQ	9	8	3
Custom-Std-PAQ	6	-	-
B. Twelve Levels			
Methods:			
Custom-Std.	8	10	2
Custom-PAQ	7	12	1
Standard-PAQ	4	11	5
Custom-Std-PAQ	3	-	-

known at present, but based on experience in other areas of human resource management, pessimism is appropriate. Second, the findings here indicate that even if two evaluation systems (Custom-PAQ) individually exhibit inter-rater reliabilities of .90 and above and their summary job worth measures correlate at .95, the rate of pay grade assignment correspondence can still be less than 50%. The differing concepts of job worth reflected in factor structures and weights, and administrative considerations in pay structure design and use inevitably will result in some divergence of results between methods.

DISCUSSION

The basic research strategy employed in this project was to develop evidence for the presence or absence of job evaluation measurement error by assessing actual outcomes from three different methods. This approach directly explores one aspect (measurement validity) of the feasibility of establishing relative job worth via job evaluation, but it suffers from two fundamental limitations. First, findings are wholly dependent upon the quality of standards applied to the evaluation measures. Shortcomings in these criteria for criteria for criteria correspondingly diminish the strength of any conclusions regarding construct validity. Two areas of concern in this regard are discussed here. Second, the diagnostic value of results is severely limited; interpretation of findings in terms of causal influences is an essentially speculative exercise. This limitation is emphasized below by

briefly reviewing process and design variable potentially impacting results but not explicitly considered in the design.

Validity Criteria

Treiman (1979) first questioned the meaning of reliability coefficients in the job evaluation context by illustrating that pay grade assignments can vary considerably even when a reliability coefficient is high (.90) by traditional measurement standards. The potential pay grade assignment analysis presented in Tables 22 and 23 supports Treiman's concern; coefficients as high as .95 result in significant potential pay grade variation based on a standard error estimate. Comparison of pay grade assignments under each method based on summary evaluation scores of each job (point estimates rather than interval estimates of worth) also reveals a high level of pay grade variation between raters. Tables 27-29 present pay grade classifications by rater for each of the three methods, and Table 30 summarizes rater pair and method reliability data by indicating the similarity in pay grade assignments associated with summary score (Pearson) and rank order (Spearman) correlations. Custom plan reliability coefficients generally range from .95 up, yet jobs are assigned to the same grade in only half of the cases and in 10% of the classification decisions the discrepancy is two or more grades. Pay grade assignment correspondence is even less under the Standard and PAQ plans as would be anticipated from their lower reliability values.

Classification agreement frequencies in Table 30 may be artificially depressed since an arbitrary (and typical)

Table 27. Comparison of Rater Pay Grade Classifications (1-12) Based on Custom Plan Job Worth Scores.

Job	Pay Grade Classification			
	R_1	R_2	R_3	R_4
1	11	10	12	11
2	10	12	11	11
3	2	2	1	2
4	5	5	5	5
5	4	4	3	4
6	10	10	8	9
7	5	5	5	5
8	7	7	7	7
9	9	9	8	9
10	3	2	3	3
11	6	4	5	6
12	4	4	5	4
13	8	8	8	9
14	8	8	9	8
15	8	7	9	9
16	8	8	7	8
17	7	7	7	7
18	6	7	7	7
19	2	3	1	3
20	4	5	4	4

Table 28. Comparison of Rater Pay Grade Classifications (1-12) Based on PAQ Job Evaluation Values.

	Pay Grade Classification			
Job	R_1	R_2	R_3	R_4
1	9	10	11	9
2	10	11	10	9
3	1	1	1	1
4	6	5	4	6
5	4	4	2	4
6	8	8	8	9
7	5	5	5	5
8	8	7	8	8
9	9	9	9	9
10	3	3	3	3
11	6	6	5	5
12	4	5	5	5
13	11	11	10	11
14	6	7	7	7
15	9	7	8	9
16	9	7	9	9
17	8	8	8	8
18	6	6	7	7
19	3	3	1	2
20	4	5	6	4

Table 29. Comparison of Rater Pay Grade Classifications (1-12) Based on Standard Plan Job Worth Scores.

	Pay Grade Classification			
Job	R_1	R_2	R_3	R_4
1	12	11	12	12
2	11	12	11	12
3	3	4	3	3
4	7	5	4	5
5	4	4	4	4
6	6	9	7	8
7	3	4	4	4
8	6	6	6	7
9	7	10	7	9
10	3	3	3	3
11	7	5	4	5
12	3	4	4	4
13	6	7	8	9
14	7	7	10	6
15	7	5	8	7
16	7	9	7	9
17	9	5	7	7
18	8	7	7	7
19	3	4	3	4
20	5	5	5	4

Table 30. Custom, Standard, and PAQ Inter-rater Reliability Summary.

	Reliab. Coefficients*		Similarity of Pay Grade Assignments **		
Custom	Job Scores	Grades	Same	± 1	± 2-4
R_1R_2	.95	.96	12	6	2
R_1R_3	.96	.95	8	10	2
R_1R_4	.98	.98	14	5	1
R_2R_3	.94	.92	6	10	4
R_2R_4	.95	.95	12	5	3
R_3R_4	.97	.97	9	9	2
			61	45	14
Standard					
R_1R_2	.72	,77	5	8	7
R_1R_3	.77	.78	10	5	5
R_1R_4	.81	.79	5	8	7
R_2R_3	.80	.85	7	7	6
R_2R_4	.93	.94	10	7	3
R_3R_4	.90	.87	8	9	3
			45	44	31
PAQ					
R_1R_2	.95	.95	11	7	2
R_1R_3	.94	.94	7	8	5
R_1R_4	.98	.97	12	8	0
R_2R_3	.94	.95	8	9	3
R_2R_4	.93	.94	8	8	4
R_3R_4	.93	.94	10	6	4
			56	46	18

* Job Score Correlation = Pearson; Grade Correlation = Spearman.
** Twelve level classification scheme.

approach to establishing classification boundaries was used rather than a more sophisticated clustering analysis. However, this possibility does not alter the basic lesson illustrated here. Inter-rater correlations are an inadequate index of job evaluation reliability. Rater consistency in evaluation should be assessed in terms of the personnel decision involved--classification assignments.

In retrospect, the logic of assessing comparable worth job evaluation in terms of convergence of multiple methods is also probably faulty. Traditionally job evaluation method comparisons were made because the efficiency with which various plans capture the predefined true of benchmark structure was one consideration in evaluation plan adoption. In contrast, a desire to test for measurement validity motivates method comparisons in the comparable worth context. The convergent validity strategy presumes the existence of a single or true construct which is the object of independent measurement methods. Thus, convergent validity as a criterion assessment strategy is technically appropriate to comparable worth job evaluation only in the unlikely situation where multiple approaches to the assessment of a single, specifically defined job worth construct are developed. Since the present study did not constitute such a situation--indeed, a basic premise of the study was that method divergence is probable--the convergent validity strategy was inappropriate.

Diagnostic Limitations

While the present research was limited to determining the presence and degree of measurement error in job evalu-

ation processes, a brief discussion of some of the specific error sources discussed in Chapter 3 is appropriate for three reasons. First, interpretation of the findings is enhanced by considering the results in terms of possible determinants of measurement variance. Second, judgments regarding the degree to which the findings are generalizable to other organizations and/or analysts must take these error sources into account. Third, observations and opinions drawn from this study suggest a number of hypotheses for future job evaluation research.

Job Characteristics -- Differences in the basic nature of jobs directly impacts job evaluation validity, particularly in terms of analysts' ability to collect accurate and comprehensive data. However, job characteristics are not a likely source of measurement variance in the data reported here. The analysts expressed no difficulties in understanding job objectives, responsibilities and activities, and the commonality in basic job functions across organizational units generated considerable redundance in data collection, thus further enhancing job understanding. In fact, the job characteristics in this particular job sample may have served to inflate the measurement validity indices. Inter-rater reliabilities would probably have been lower had the research site been a more complex organization with greater differentiation of goals, activities, and/or blurred hierarchical relationships.

Rater -- Relatively few insights regarding rater errors were stimulated during the study. The presence of a female and an outside analyst created potential for sex or familiarity affect but no hint of such variance was found in the ratings or observed in rater interactions. Similarly, although one of the program areas involved only traditionally female occupations (home economists), bias toward these jobs was not revealed in any way. Personal biases toward particular incumbents were often expressed by individual raters but job scores did not reflect them.

The rater variable of most consequence in this study may have been rater motivation. Repetitive evaluations, particularly those involving a lengthy document such as the PAQ, become an onerous task after the first few applications. In a research context, with no perceived substantive consequences for the raters, the quality of effort is likely to deteriorate. Low rater motivation often translates into ratings reflecting leniency or central tendency. The PAQ ratings of analyst number 1 are a probable example of this shortcoming. Whether a similar motivation problem would exist in an operational application of a comparable worth evaluation system is, of course, unknown.

Job Analysis -- Since the job analysis was conducted immediately prior to the evaluation judgments by the same four analysts, the only remaining questions are whether the method employed generated comprehensive and commonly understood job information. No significant reason exists for a negative

response to either question. Multiple positions were analyzed for most jobs; job content was relatively simple and three of the raters were very familiar with the jobs prior to the formal analysis; the PAQ questionnaire forced consideration of detail, and finally, extensive discussions among the analysts regarding each job took place. A job description "validity check" was not employed due to the analysts' belief that job description documents, when prepared and "validated" for compensation purposes, are as likely to distort as to clarify job information.

Evaluation Plan Characteristics -- Rating scale ambiguity appears to be the major determinant of rater inconsistency. The highest reliabilities were obtained with the Custom plan, in which most scales are quantitatively anchored or reflect gradations defined in terms meaningful to the raters in that particular organizational context. The lowest reliability coefficients among the ten Custom plan scales were associated with the two scales which were least descriptive of job content requiring raters to subjectively estimate necessary personal characteristics (seniority, human relations skills). Standard Plan ambiguity, as discussed above, also attests to the importance of specificity in scale definitions and anchors.

The basic issue of job worth definition was not directly addressed in this study. Both the interim (Treiman, 1979) and final (Treiman and Hartman, 1981) National Academy of Sciences reports stressed the probable impact of factor

selection and weighting procedures on variance in ultimate job rankings. Regression approaches can be expected to yield different job worth criteria and weights depending upon the composition of the benchmark jobs and the salary criterion structure employed. The likely extent of such differences in a given situation is currently unknown but sample size limitations precluded investigation here. Similarly the factors and weights determined via a judgmental approach are a function of the designing group's values, but the necessity to maintain maximum confidentiality in this project did not allow any follow-up to determine whether or to what degree the Custom plan dimensions were in consonance with the values expressed by others within the organization.

Differences between regression and judgmental approaches to definition of job worth are inevitable. Job content based factors and weights defined a priority will not correspond completely with regression derived criteria which capture or reflect the composite of personal, economic, social, and institutional forces determining the criterion pay structure. Custom plan--PAQ outcomes partially reflect this regression-judgmental difference, but since the PAQ criterion structure is not unique to the organization, differences in results cannot solely be attributed to the approach.

In summary, eight questions were initially posed to guide this inquiry into the feasibility of employing job evaluation procedures for measuring the relative worth of jobs. These questions focused on psychometric properties of

method ratings, individual and comparative, as reflected in various construct validity indices. No criterion of "measurement adequacy" was proposed; rather, questions regarding interpretation of validity standards were raised. In addition, method comparisons in terms of ratings and job grade assignments were conducted.

The research findings tend to support the basic proposition that rater and method sourced variability in comparable worth job evaluation can be significant; they also reveal that traditional criteria of measurement adequacy may be inappropriate to job evaluation measures. However, the findings, per se, are an inadequate basis for making a comparable worth measurement feasibility determination. While they raise some serious concerns regarding measurement validity, they also indicate areas of potential improvement. Chapter six reviews these and other considerations in arriving at tentative conclusions regarding comparable worth job evaluation feasibility.

CHAPTER 6

Summary and Conclusions

The preceding chapters set out to explore the feasibility of designing and implementing comparable worth (bias f48ree) job evaluation procedures. The study was motivated by mounting criticism of current wage-setting practices and proposals for national policy intervention to eliminate perceived sex based pay differentials in wage-setting practices. To achieve the dissertation objectives, two central and interrelated concepts were analyzed. First, the concept of job worth or value was examined from philosophical and theoretical perspectives as well as via as a review of contemporary job evaluation procedures. Second, the concept of measurement validity as it applies to the job evaluation context was discussed in terms of requisite design and operational properties, and estimates of selected properties were generated from evaluation scores of three different methods. The primary observations and findings regarding both of these concepts and their implications for comparable worth job evaluation feasibility are summarized here.

Job Worth

By far the most critical element in comparable worth job evaluation proposals is the call for implementation of a

single system of evaluation within an establishment (or employer). This recommendation presupposes an ability to identify commonly accepted criteria of worth, an assumption which gives rise to two sets of criterion relevance issues. The first of these pertains to the type of criteria.

Job worth criteria can be grouped into four categories or models of value determination. Model one is based upon the concept of intrinsic worth of jobs and emphasizes job content. The final National Academy of Sciences report reflects this view in its definition of discrimination: "Wage discrimination exists when individuals of one social category are paid less than individuals of another social category for reasons that have little or nothing to do with the work they do"(Treiman and Hartman, 1981:9; emphasis supplied).

Models two and three are supply and demand side versions of the concept of exchange value, incorporating market forces as legitimate job value criteria. Relative value is attributed to shortages/surpluses of particular skills in model two, i.e., the critical criterion of worth is relative bargaining power. Model three is an employer utility or use value model in which job worth or value is linked to financial return. Operational measures of utility may include job characteristics, but personal traits and skills often outweigh content dimensions in pay grade assignments, particularly in sales and professional positions.

The fourth category of job worth criteria embraces societal or organizational customs and norms. Workplace beliefs

about relative value are often based on job content dimensions but may also reflect forces or factors irrelevant to contemporary operations by any rational assessment. Since attempts to modify or eliminate these criteria often result in severe workforce disruptions, they are considered in wage determination.

The conflicts between sets of job worth criteria are clearly delineated in the reviews of historical and contemporary approaches to job worth determination contained in chapters one and two. Comparable worth advocates have not adequately addressed the need to reconcile these content, market, and traditional criteria. Increased emphasis on rational assessment of job content may reduce sex related wage differentials, but it is not at all clear how market forces can be divorced from value determination in a decentralized economic system. At a minimum, some form of "business necessity" criteria (in addition to job content standards) would appear necessary to implement a comparable worth evaluation system.

It should be noted here that the comparable worth movement is not a call for a fundamental shift in values. The challenge is addressed to proximate rather than ultimate criteria of pay distribution. As in traditional wage setting, comparable worth advocates stress productivity as the ultimate determinate of value. In essence they are arguing that direct assessment of contribution to organizational results

based on job content is superior to the mixed content-market models characteristic of contemporary wage determination.

Even if agreement can be reached that the overriding determinant of worth should be job content, a second set of criterion relevance issues must be resolved. The basic question is whether generally accepted units of work content applicable to all jobs within an establishment can be identified. This question has both technical and value judgment facets.

From the technical perspective the issue is one of dimensionality, i.e., determining the relevant work units or job worth dimensions. On a purely logical basis there does not appear to be any compelling reason why a standard set of criteria cannot be determined. Job evaluation practitioners have long referred to the four so-called universal factors of job worth - skill, effort, responsibility, and working conditions. But development of widely applicable operational definitions of these factors has proven quite difficult; each of the universal factors is a multidimensional construct and sub-factor relevance varies considerable across jobs, occupations, and contexts. At this juncture the issue of the appropriate level of analysis has not even been resolved. In the present study, for example, PAQ scores were derived from task analysis whereas the Custom and Standard plans employed more global job level criteria. Furthermore, there is currently no empirical basis for selecting one analytical approach or set of criteria over others, although the result

reported here weakly support a conclusion of measurement superiority in locally defined criteria. In short, the technology of job worth criteria development is at a primitive stage, but the identification of common criteria probably is technically feasible.

More important and difficult to resolve than the dimensionality issue is that of differential importance (relative weights) of job content criteria. Value judgments in establishing factor weights are unavoidable in both of the approaches proposed by comparable worth advocates. Assignment of weights in accordance with an organizational consensus is the method most consistent with the comparable worth principle. However, difficulty in achieving consensus when factor weights are judgmentally determined has led to suggestions for employing multiple regression approaches. But weighting of factors via regressing a "nonbiased" pay structure (male only pay structures or hierarchies statistically adjusted for sex effect) on job content scores merely shifts value judgments from individual factors to potential criterion structures. Interested parties must first come to agreement regarding the "rightness" or objectivity of a job hierarchy. Whether general agreement is more readily obtained regarding perceptions of bias free criterion structures than factor weights is unknown, but the writer sees no basis for optimism on this score.

Regression approaches have an additional drawback in that they dilute and/or distort job content models of worth

because regression equations include only those factors necessary to maximally predict the criterion. For example, the original PAQ Inc. regression equation contained only nine of the 27 PAQ dimension scores, i.e., only a portion of the job content model of worth was actually utilized in wage determination. Job content dimensions simply captured or modeled variation in exchange rates. Therefore, statistically derived factors and weights inevitably depart from job content models of worth. Again, whether such weighting schemes can achieve general acceptance is unknown, but they would appear to potentially suffer from face validity shortcomings given the "black box" nature of their derivation.

This brief review of criterion relevance issues underscores the fallacy of referring to any evaluation system as non-biased. Selection of job worth models and operational criteria, and assignment of factor weights all involve value judgments. Thus evaluation systems inevitably reflect the biases of their designers. The issue is to whether the system is bias free but whether illegal bias is eliminated.

In summary, the answer to whether it is feasible to define a set of generally accepted job worth criteria within an organization is a hesitant yes - under three conditions. First, "generally accepted" is to be interpreted as a majority rather than consensus acceptance of the evaluation system. Unanimity of agreement is improbable in the vast majority of situations. To the degree a particular organization is

characterized by employee heterogeneity and/or organizational complexity, majority may be difficult to achieve.

Second, and following from number one, the criterion dimensions (conceptual and operational) must be participatively developed, and procedures implemented to ensure continued perceived relevance of criteria and consistency of application in job worth judgments. Unlike contemporary job evaluation, outcomes of a comparable worth evaluation system may force a realignment of traditional job relationships; acceptance of these results will be heavily dependent upon acceptance of the content scales and weights.

Third, the system must make provision for reconciling conflicts between job content and market concepts of worth. Assessment of relative job content ultimately must be translated into monetary exchange values. Accommodation of these rates to market forces is, of course, necessary to avoid competitive disadvantage in labor and/or product markets.

Measurement Validity

While the concept of value or worth serves to focus normative questions of criterion content in job evaluation, measurement validity issues center more on assessment processes or methods. It may be possible to define a generally acceptable set of evaluation criteria, but whether the evaluation process accurately and consistently reflects true variance among jobs on the defined job worth dimensions is another question. The concept of measurement validity is concerned with the degree to which random and systematic error

enter into the job rating process. On these questions the findings here suggest greater cause for pessimism than optimism.

Four general measurement validity conclusions are warranted. First, the findings here as well as those of Lawshe and associates cited earlier, indicate there is general poor correspondence between conceptual dimensions and operational measures of job worth. The number of dimensions, variability within dimensions, and the concept of male and female occupational dimensions have all figured prominently in the comparable worth debate. Yet analysis of evaluation results typically reveals only two to three operational factors. Either the hypothesized dimensions of job worth are interrelated to such a degree that it is effectively a global construct, or evaluation processes are seriously inadequate. Which of these conclusions is most valid cannot be stated on the basis of research to date; job worth dimensionality issues should be the highest priority item in future job evaluation research.

Second, the concern of comparable worth advocates with systematic bias against females due to evaluation criteria deficiencies and/or job bias may be misplaced. The greater problem could be measurement error. Certainly job and evaluation dimension bias are potential problems, but their seriousness should be significantly diminished by adoption of common internal criteria. However, perceived inconsistency in evaluation outcomes is virtually certain to result in

routine adjustments to scores as salary administrators strive to maintain consistency and face validity in their pay structures. Such "Flexibility" could easily translate into devaluation of traditional female occupations. Alternatively stated, evaluation system slack due to measurement error may adversely impact female incumbents and/or traditional female jobs within organizations.

Third, traditional reliability indices are inadequate for assessment of inter-rater reliability in job evaluation. Rater consistency is critical, whether the job worth outcomes serve as direct measures of worth or as components in a statistical prediction. However, findings here suggest rater disagreement on pay grade decisions is highly probable, even when evaluation score reliability coefficients exceed .90. Therefore, estimates of job evaluation reliability should utilize a classification convergence index, and practitioners would be well advised to routinely employ multiple rater strategies to enhance evaluation reliability.

Fourth, it is probable that differing levels of measurement adequacy among current evaluation plans, when combined with divergence in underlying concepts of worth, yields results that are method dependent. Pay structures emerging from internally developed evaluation plans will not correspond to those generated by instruments of outside agents, be they consultants or regulators. Consequently, if the comparable worth principle is adopted, determinations of discrimination will likely turn on employer fidelity to their

evaluation system. Furthermore, the fact of method variance provides an opportunity for employers to select, from commercial or internally generated alternatives, that system which is most consistent with a desired pay structure. Therefore, comparable worth job evaluation will not necessarily result in major realignment of organizational pay structures.

A fifth, and somewhat contradictory observation regarding measurement validity must also be offered for consideration here. A reasonable case can be made that measurement properties are of as little consequence to comparable worth evaluation as they were in the past. In chapter two it was noted that all four traditional methods essentially serve to rationalize a complete job structure around a given key job hierarchy. Similarly, rater tendency to assign individual Standard Plan factor and overall worth scores in accordance with their beliefs about the true hierarchy was discussed in chapter five. And the suggestion that regression techniques be employed to develop job content models of a non-biased pay hierarchy was noted above. Common to all of these observations is the notion of rationalizing subjective norms in job content terms, and the literature on human judgments has clearly identified a propensity for elaborately rationalizing normative judgments. Therefore, a key premise of this research may be faulty; rather than a measurement process, comparable worth job evaluation ultimately may be simply a new improved version of the traditional process of surfacing and rationalizing legitimate norms of relative worth. If so,

validity in the technical sense is a minor concern; employee acceptance remains the ultimate test of job evaluation systems.

In closing it should be remembered that while this study emphasized the measurement deficiencies of job evaluation, inadequate job evaluation measures may be perceived by some groups in society as preferable to existing practices. The comparable worth movement is a conscious strategy to effect changes in wage setting practices via political processes. As such it is supported by a constituency which appears to be growing. Therefore, feasibility issues may be fundamental in a technical sense but secondary politically if sufficient people decide that perceived favorable consequences override these concerns. The critical issues for future research may relate not to whether implementation of the comparable worth principle is feasible, but to means by which the process can be improved.

Appendices

Appendix A

Job Sample

Appendix A

Job Sample

Job No.	Label	Positions
01	Director, Ruran/Urban	1
02	Director, Urban	2
03	Clerk Typist	3
04	Program Asst., Commun.	1
05	Office Supervisor	2
06	Youth Agent	6
07	Asst. Youth Agent	6
08	Agricultural Agent	5
09	Home Economist	4
10	ENP Program Aide	3
11	Asst. Home Economist	1
12	Supervisory Aide	2
13	Regional Supervisor	3
14	Program Leader	3
15	Director, Rural	1
16	District Agent, A	1
17	District Agent, B	3
18	District Agent, C	1
19	Secretary	2
20	Accountant	1

Appendix B

Job Analysis Interview Guide

APPENDIX B

JOB ANALYSIS

INTERVIEWERS GUIDE

Location _____ Years on Job _____

Incumbent(s)_____ Years in CES _____

Job Title _____ Supervisor Name _____

Program Areas _____ Supervisor Title _____

- -

I. INTRODUCTION

a. Explain purpose of the study
b. Explain method of interview
c. Reassurances: -job focus, not person
-no possible negative impacts
d. Questions or concerns?

II. OVERVIEW OF JOB

a. Brief job description
- why job exists

- major functions

b. Position activities

- describe the activities/tasks involved in a typical day-week - month.

- most frequent or time consuming?

- important periodic or irregular tasks/activities

- describe your "job cycle"

III. MENTAL PROCESSES

a. Work Planning/Scheduling requirements. (self and/or others)

(consider planning aspects of each major function/activity)

- amount
- activity sequencing v. methods v. goals
- supervisor role
- planning horizons

b. Problem Solving Requirements

- types of problems (examples)
- most frequent
- most difficult
- approach to resolving; types of analysis required
- resources available; extent personal judgment is involved
- special skills required
- level of mathematics required

c. Decision Making Requirements

(consider type of decisions in each major job function/activity)

- guidance available (instructions, procedures, policies, precedents)
- consequences of decisions;

  on others

  on organizational goals

IV. RELATIONSHIPS WITH OTHERS

a. Primary contacts

- with whom
- frequency

b. Reasons

- information exchange
- coordinate activities
- persuade
- advise, counsel
- instruct
- negotiate; conflict handling
- entertain

c. Formal Communications

- writing
- speeches; other presentations
- other presentation skills

c. Relationship with Supervisor

- nature of direction
- frequency and type of review

d. Subordinate relationships

- # supervisory personnel directed
- # nonsupervisory personnel directed
- non-paid staff direction
- scope of subordinate activities

V. FINANCIAL RESPONSIBILITIES

a. Generating Funds: activities? local program? state programs?

b. Cash handling - bookkeeping

c. Budget accountability
- primary? contributory?
- reporting requirements

d. Budget magnitude (000)

VI. KNOWLEDGE - SKILL - ABILITY REQUIREMENTS

(Ask interviewee to disregard his/her own background in responding to these questions. What does the job require?

a. Level of facts, principles, concepts required to do the job? (educational equivalents)

b. Other job related skills

- equipment, machinery, tools?
- operating permits?
- licenses; certifications?

c. Experience required to learn job (length; type)

d. On the job training time required to reach minimally acceptable level

e. knowledge updating requirements? (estimated hours per month)

VII OTHER JOB REQUIREMENTS

a. Unpleasant physical conditions?

- type; frequency

b. Hazards?

c. Social obligations? (civic functions club meetings etc.)

d. Pressures?

Appendix C

Job Evaluation Instruments

Appendix C

Custom Plan

Factor Point Values

FACTOR	DEGREE LEVELS 1	2	3	4	5	6
Education/Knowledge	12	34	55	77	98	120
Experience	7	22	38	54	70	
Personal Skills						
Human Relations	15	52	91	130		
Teaching & Comm.	12	48	84	120		
Decision Making Respons.						
Autonomy/Initiative	12	39	66	93	120	
Scope/Impact	10	28	46	64	82	100
Fiscal Responsibility						
Fund Gener.	5	20	35	50		
Fiscal Mgt.	5	16	27	38	50	
Magnitude	6	16	27	38	49	60
Supervisory Respons.						

Supvn. of Paid Staff	Supervn. of Volunteers Min.	Mod.	Extensive
Level 1	18	45	72
Level 2	45	72	99
Level 3	72	99	126
Level 4	99	126	153
Level 5	126	153	180

1. Education/Knowledge Requirements

This factor evaluates the nature and extent of technical, administrative, or organizational information and/or skills which must be mastered to do acceptable work. The emphasis here is on knowledge and skills actually required by the job, however obtained, as distinguished from the incumbent's education or preferred applicant credentials.

Level 1 Requires sufficient knowledge and skills to understand and perform routine tasks or operations following standard procedures or oral instructions. Educational equivalent is some high school.

Level 2 Requires knowledge of administrative/technical procedures or operations which require some previous training or experience typically acquired in high school or post high school courses.

Level 3 Requires knowledge or skills in a particular field acquired through training or experience equivalent to high school plus two years of college or vocational training.

Level 4 Requires knowledge of basic principles, concepts and methods of a technical or administrative field equivalent to that obtained through a pertinent baccalaureate program.

Level 5 Requires advanced knowledge of technical or administrative concepts, principles, and practices gained through graduate study or experience such that assignments or positions of greater difficulty than those covered by level 4 can be effectively carried out.

Level 6 Requires technical/administrative knowledge of level 5 plus knowledge of the organizational system and its interrelationships with political, legal, and other environments.

2. Experience Requirements

This factor evaluates the length of time on related work (paid or volunteer) that is prerequisite to appointment to the job plus on the job training time required for a person with the necessary knowledge background to perform the job in a minimally acceptable manner.

Level 1 Up to three months

Level 2 Three to twelve months

Level 3 Over one year through two years

Level 4 Over two years through five years

Level 5 Over five years

3. Personal Skills Requirements

Jobs requiring equivalent technical or experiential background can differ significantly in terms of the personal skills required to effectively apply the knowledge or experience. The extent to which jobs vary in terms of these personal capability requirements is assessed by these factors.

Human Relations Skills: This factor evaluates the nature and extent of personal interactions and their importance to acceptable performance.

Level 1 Job requires ordinary courtesy and effectiveness in interactions with other people. The majority of personal interactions involve people in the same work unit; relationships with other CES people or clients are sporadic and generally well structured.

Level 2 Understanding and effectively interacting with others is an important but not critical aspect of the job. Interactions are predominently with other local or state CES personnel.

Level 3 Understanding, influencing and maintaining effective relationships with both clients and CES people are central to adequate performance of the job.

Level 4 Job requires a complete repertoire (communicating, influencing, negotiating, conflict handling) of polished interpersonal skills to effectively manage relationships with CES, client, and significant others in the community or state.

Teaching and Communications Skills: This factor evaluates the extent to which communication and teaching skills are necessary to acceptable job performance.

Level 1 Job requires the basic writing and speaking skills necessary to effectively communicate with others regarding day-to-day activities.

Level 2 Job requires sufficient skill in writing or speaking to effectively present information to individuals and/or groups in formal and informal settings.

Level 3 Job requires competence in writing, speaking and other communications media in order to both develop and present information in various settings.

Level 4 Requires a high level of proficiency in the use of a wide range of communications media to effectively develop and present information.

4. Decision Making Responsibilities

This factor evaluates the decision making requirements of the job in terms of the degree to which actions are prescribed by guidelines or supervisory directives, and in terms of the scope/impact of decisions.

Autonomy/Initiative: This factor evaluates the extent to which the job requires taking action without the benefit of written guidelines or supervisory direction.

Level 1 Job requires routine decisions covered by specific instructions. The incumbent works in close adherence to the guidelines or instructions; deviations must be reviewed with the supervisor.

Level 2 Job requires the exercise of some judgment and initiative in carrying out routine assignments independently. Activities are covered by written or supervisory guidelines but judgments are sometimes required to determine their application.

Level 3 Makes independent decisions regarding procedures or methods to achieve specified objectives; uses judgment in handling problems and deviations in assignments in accordance with policies, instructions, or accepted practice. Supervisor establishes priorities, deadlines, and resources.

Level 4 Consults with supervisor and other relevant individuals and groups in establishing objectives, priorities and deadlines; exercises initiative in planning and carrying out assignments; Interprets policy on own initiative in terms of established objectives. Informs supervisor of progress.

Level 5 Has primary responsibility for decisions in establishing objectives, designing, and carrying out programs, projects, or other work. Supervisor provides administrative support and direction in terms of broadly defined missions or functions.

Scope/Impact: This factor considers the extent of the potential impact of decisions on the organization and its clientele.

Level 1 Decisions have limited and minor impact beyond the individual involved.

Level 2 Decisions have moderate positive or negative impact on specific program area constituents or on effectiveness or efficiency of immediate work group.

Level 3 Decisions have major impact on the nature and quality of service in a single local program area or administrative support unit.

Level 4 Decision impact extends to total local area staff and clientele or to a specified range of activities within the total organization.

Level 5 Decisions affect the quality and nature of activities throughout the organization.

Level 6 Decisions often involve policy issues broadly impacting the direction, effectiveness, and efficiency of the organization.

5. Fiscal Responsibility

Financial responsibility within Cooperative Extension varies considerably among jobs, within and between counties, and at the state level. It includes responsibility for the generation of activity and program funds, cash receipts, disbursements and records, and operating budget planning and accountability. Included in this factor are monies appropriated from various units of goverment, and non-appropriated funds such as grants from a variety of sources,

donations from the private sector, and special activity money for a specific educational program and/or activity (i.e., collecting money for the Annual Dairy Banquet, Leaders Banquet, 4-H Exploration Days, College Week, Short Courses, etc. Fiscal responsibility can be divided into the three sub-factors of fund generation responsibilities, fiscal management responsibilities, and magnitude of responsibility.

Fund Generation Responsibilities:

Level 1 No responsibility for generating funds.

Level 2 Job requires finding monetary resources for single events or activities like College Week Scholarships, Leader Banquet, etc.

Level 3 Job requires determining financial needs, identifying sources, and obtaining either appropriated and/or grant money for one program area.

Level 4 Responsible for generating the total CES county budget, or responsible for generating state and/or federal appropriated money affecting program area at the state level.

Fiscal Management Responsibilities:

Level 1 No direct responsibility.

Level 2 Job requires the handling of minor cash transactions relating to small item sales, enrollment fees, etc.

Level 3 Job requires minor cash management responsibilities (record keeping, deposits, balancing accounts, summarizing financial records) or budget monitoring and reporting responsibilities within an administrative unit.

Level 4 Job involves responsibility for budget preparation, expenditure control, and financial reporting for a specific program area.

Level 5 Job involves responsibility for overall budget preparation and accountability for expenditure of funds against that budget.

Magnitude of Responsibility:

Level 1	Up to $2,500	Level 4	$50,000 to $100,000
Level 2	$2,500 to $20,000	Level 5	$100,000 to $300,000

Level 3 $20,000 to $50,000 Level 6 $300,000 Up

6. Supervisory Responsibility

This factor evaluates the amount of responsibility for hiring/dismissing, evaluating, developing, and coordinating the efforts of other people to achieve the goals of the Cooperative Extension Service. The supervisory responsibility factor may also include union relationships involving such duties as day-to-day contract administration/interpretation, and employee grievance processing. Supervisory responsibilities can pertain to board appointed staff, secretaries, aides and program assistants, as well as volunteers. The extent of the supervisory responsibility, the number of people directly and indirectly supervised, and the extent to which the subordinates are involved in diverse activities are all considered.

Supervision of Paid Staff	Supervision of Volunteers*		
	I Minimal	II Moderate	III Extensive
1. Normally not required to supervise others.			
2. Supervises a small group (usually 2-10) engaged in similar day-to-day activities by instructing, assigning, and/or checking work. Little or no responsibility for other personnel processes and/or procedures.			
3. Job requires direct and complete supervisory responsibility (budgetary, hire/fire, direct and develop) for a small group (2-10) engaged in two or more distinct activity areas or for a larger group involved in like activities.			
4. Job entails supervisory responsibility for more than one program area through subordinates who exercise full supervision. Includes the directing, training, and inputs into evaluation of board appointed as well as paraprofessional staff. This level may also include working with situations involving union contract interpretations.			
5. Involves the direction and coordination of a major function or state level			

Supervision of Paid Staff	Supervision of Volunteers* I Minimal	II Moderate	III Extensive
program area. May include hiring, evaluation, training and dismissal of board appointed staff for a specific program area.			

---

* Minimal - Normally not required to recruit and direct volunteers.

Moderate - Job requires some recruitment and direction of volunteers for delivery of specific programs, but volunteers do not constitute the major means by which job objectives are achieved.

Extensive - The recruitment, training, directing, and motivating of volunteers is a primary means of achieving objectives in a general program area.

Standard Plan

Knowledge Requirements

Measuring Knowledge: Knowledge has both scope (variety) and depth. Jobs require, in varying combinations, some knowledge about a lot of things, or a lot of knowledge about a few things. Thus, the knowledge concept reflected in this evaluation chart enables the comparison and weighing of the total knowledge content of different jobs in terms of "how much knowledge about how many things."

Definition: Knowledge is defined here as the sum total of every kind of skill, however acquired, needed for acceptable job performance. Knowledge has the three major dimensions of:

1. Practical procedures, specialized techniques, and learned disciplines (job technology requirements)

2. Managerial knowledge--the knowledge involved in integrating and harmonizing diversified functions and activities; usually involves some combinations of the functions of organizing, planning, executing, controlling, and evaluating.

3. Active, practicing skills in the areas of human relationships

- -

Levels of Managerial Knowledge--Definitions

I. None or minimal: Performance or supervision of a single activity.

II. Intermediate: Primarily within single field or toward single objective with some integration of, or external integration with, other fields.

III. Broad: Integration and coordination of diversified activities in an important management area or consulting field.

IV. Comprehensive: Comprehensive integration and coordination of diversified activities and functions in a major management area.

V. Major: Management at the level of policy making which affects the overall operation of the organization.

VI. Total

Measuring Problem Solving-Decision Making: This factor measures the intensity of the mental process which employs knowledge in analyzing, evaluating, creating, reasoning, arriving at and making decisions. To the extent that thinking is circumscribed by standards, covered by precedents, or referred to others, problem solving-decision making is diminished, and the emphasis correspondingly is on knowledge required.

Definition: Problem Solving--Decision Making is the original, self starting thinking required by the job to identify, define, and resolve a problem. "You think with what you know" . . . this is true of even the most creative work. The raw material of any thinking is knowledge of facts, principles, and means. Ideas are put together from something already there. Therefore Problem Solving--Decision Making is treated as a percentage utilization of Knowledge requirements. It has two dimensions:

(1) The environment in which the thinking takes place. (see chart)

(2) The challenge presented by the thinking to be done.

1. Repetitive: Identical situations requiring solution by simple choice of learned things.

2. Patterned: Similar situations requiring solution by discriminating choice of learned things.

3. Interpolative: Differing situations requiring search for solutions within area of learned things.

4. Adaptive: Variable situations requiring analytical, interpretative, evaluative, and/or constructive thinking.

5. Uncharted: Novel or nonrecurring pathfinding situations requiring the development of new concepts and imaginative approaches.

Accountability

Definition--Accountability is the answerability for action and for consequences thereof. It is the measured effect of the job on end results. It has three dimensions in the following order of importance:

(1) Freedom to Act--the degree of personal or procedural control and guidance as defined in the left hand column on the chart.

(2) Job Impact on end results--whether the impact is indirect (contributory) or direct (primary).

(3) Magnitude--indicated by the general dollar size of the area(s) primarily affected by the job.

Evaluation Chart

Job Technology Requirements - Concepts, principles - Facts - Techniques - Procedures	MANAGERIAL KNOWLEDGE REQUIREMENTS														
	None or Minimal I.			Intermediate II.			Broad III.			Comprehensive IV.			Major - Policy Making V		
Hum. Rel. Skills→	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
A. PRIMARY Elementary plus some secondary (or equivalent) education	25	28	33	33	38	43	43	50	57	57	66	76	76	87	100
	28	33	38	38	43	50	50	57	66	66	76	87	87	100	115
	33	38	43	43	50	57	57	66	76	76	87	100	100	115	132
B. ELEMENTARY VOCATIONAL Familiarization in standardized work routines and/or use of simple equipment and machines.	33	38	43	43	50	57	57	66	76	76	87	100	100	115	132
	38	43	50	50	57	66	66	76	87	87	100	115	115	132	152
	43	50	57	57	66	76	76	87	100	100	115	132	132	152	175
C. VOCATIONAL Procedural or systematic proficiency which may involve a facility in the use of specialized equipment.	43	50	57	57	66	76	76	87	100	100	115	132	132	152	175
	50	57	66	66	76	87	87	100	115	115	132	152	152	175	200
	57	66	76	76	87	100	100	115	132	132	152	175	175	200	230
D. ADVANCED VOCATIONAL Some specialized skill(s), giving additional breadth or depth to a generally single function.	57	66	76	76	87	100	100	115	132	132	152	175	175	200	230
	66	76	87	87	100	115	115	132	152	152	175	200	200	230	264
	76	87	100	100	115	132	132	152	175	175	200	230	230	264	304
E. BASIC SPECIALIZED Sufficiency in a technique which requires a grasp of involved practices and/or scientific theory/principles.	76	87	100	100	115	132	132	152	175	175	200	230	230	264	304
	87	100	115	115	132	152	152	175	200	200	230	264	264	304	350
	100	115	132	132	152	175	175	200	230	230	264	304	304	350	400
F. SEASONED SPECIALIZED Proficiency, gained through wide exposure of experiences in a specialized or technical field.	100	115	132	132	152	175	175	200	230	230	264	304	304	350	400
	115	132	152	152	175	200	200	230	264	264	304	350	350	400	460
	132	152	175	175	200	230	230	264	304	304	350	400	400	460	526
G. SPECIALIZED MASTERY Determinative mastery of techniques, practices and theories gained through wide seasoning and/or development.	132	152	175	175	200	230	230	264	304	304	350	400	400	460	526
	152	175	200	200	230	264	264	304	350	350	400	460	460	526	608
	175	200	230	230	264	304	304	350	400	400	460	526	526	608	700

Human Relations Skills: 1 = Basic Ordinary courtesy/effectiveness required.
2 = Important
3 = Critical

Evaluation Chart

PROBLEM SOLVING - DECISION MAKING

ENVIRONMENT	CHALLENGE				
	1. Repetitive	2. Patterned	3. Interpolative	4. Adaptive	5. Uncharted
Simple rules and detailed instructions	10% 12%	14% 16%	19% 22%	25% 29%	33% 38%
Established routines and standing Instructions	12% 14%	16% 19%	22% 25%	29% 33%	38% 43%
Somewhat diversified procedures and precedents.	14% 16%	19% 22%	25% 29%	33% 38%	43% 50%
Substantially diversified procedures and specialized standards.	16% 19%	22% 25%	29% 33%	38% 43%	50% 57%
Clearly defined policies and principles.	19% 22%	25% 29%	33% 38%	43% 50%	57% 66%
Broad policies and specific objectives	22% 25%	29% 33%	38% 43%	50% 57%	66% 76%
General policies and ultimate goals	25% 29%	33% 38%	43% 50%	57% 66%	76% 87%

Evaluation Chart

ACCOUNTABILITY

FREEDOM TO ACT	MAGNITUDE									
	0 - $10,000		$10 - 30,000		$30 - 100,000		$100 - 500,000		over $500,000	
	C	P	C	P	C	P	C	P	C	P
A. Prescribed - These job are subject to close supervision; direct and detailed instructions	8	15	11	19	15	25	19	33	25	44
B. Controlled - These jobs are subject to close supervision; instructions and established work routines.	13	22	16	28	22	38	28	50	38	66
C. Standardized - Jobs subject to standardized practices and procedures; general instructions; supervision of progress and results.	19	33	25	44	33	57	44	76	57	100
D. Generally Regulated - jobs subject to practices and procedures covered by precedents or policy. Supervisory review.	28	50	38	66	50	87	66	115	87	152
E. Directed - jobs subject to broad practice and procedures covered by functional precedents and policies. Managerial direction.	44	76	57	100	76	132	100	175	132	230
F. Guidance - Jobs are inherently subject only to broad policy and general management guidance.	66	115	87	152	115	200	152	264	200	350
G. General Guidance - Jobs are subject only to guidance from the Director's office or from appointed boards.	100	175	132	230	175	304	230	400	304	528

C = Contributory. Providing information, advice, services, etc. for use by others in taking action.

P = Primary Job has a controlling impact on end results where shared accountability of others is subordinate.

Appendix D

Custom Plan Multirater--Multidimension Correlations

Appendix D

Table A1: Custom Plan Multirater--Multidimension Correlations

		Rater 1									
R1	1										
	2	58									
	3	50	44								
	4	75	45	54							
	5	88	64	73	81						
	7	29	40	52	40	47	58				
	8	30	37	42	29	50	61	87			
	9	38	42	51	36	55	61	87	89		
	10	36	48	64	34	51	64	72	60	68	1.0
R2	1	90	47	13	70	86	61	42	46	44	34
	2	64	79	38	54	69	50	52	55	53	46
	3	64	29	58	59	69	72	58	60	53	62
	4	56	12	37	86	68	32	55	53	40	29
	5	89	45	33	78	90	68	60	61	56	51
	6	85	43	34	48	71	75	62	61	61	57
	7	41	23	60	43	42	40	95	88	86	68
	8	46	17	42	28	39	36	79	76	82	54
	9	53	33	52	48	52	55	82	89	95	72
	10	50	51	53	42	51	59	68	72	70	93

Table A1 (continued)

		Rater 1									
R3	1	90	79	48	76	74	74	31	36	32	40
	2	64	72	40	57	67	49	38	37	36	32
	3	64	67	93	61	81	84	51	53	53	71
	4	56	49	55	84	71	48	34	33	27	47
	5	89	85	67	87	89	83	46	55	47	51
	6	85	79	71	62	75	91	65	68	63	66
	7	40	35	46	50	45	43	95	89	87	71
	8	45	38	33	45	42	50	87	94	84	53
	9	53	48	45	55	51	56	82	91	93	63
	10	49	45	59	39	45	56	62	65	61	87
R4	1	87	62	54	66	74	68	27	22	44	37
	2	63	78	48	35	72	68	38	46	53	46
	3	70	41	90	52	76	66	55	39	53	67
	4	67	29	63	89	74	43	34	25	40	40
	5	91	69	74	75	91	78	42	36	56	53
	6	81	63	67	48	79	80	56	53	61	61
	7	33	39	43	38	63	48	85	84	86	71
	8	36	52	29	34	61	50	74	87	82	54
	9	41	52	42	38	61	55	82	82	95	66
	10	45	35	47	29	56	49	57	62	69	93

Table 1 (continued)

		Rater 2									
R2	1										
	2	56									
	3	19	17								
	4	61	29	53							
	5	88	56	38	79						
	6	83	56	38	52	81					
	7	50	28	64	60	55	47				
	8	57	37	67	66	62	60	93			
	9	56	41	61	55	56	58	85	95		
	10	50	55	55	50	58	69	67	76	74	1.0
R3	1	100	95	57	73	82	89	45	54	50	50
	2	56	65	25	29	40	44	18	23	29	23
	3	19	28	62	39	43	30	62	58	64	62
	4	61	58	67	90	82	52	36	40	42	50
	5	89	87	65	78	86	81	42	49	48	54
	6	83	84	63	48	69	82	43	52	56	64
	7	50	48	50	63	60	47	91	86	82	70
	8	57	55	52	64	63	53	82	85	87	71
	9	56	51	47	56	55	53	81	87	93	63
	10	50	48	57	42	47	54	56	65	66	81

Table 1 (continued)

	Rater 2									
R4 1	92	75	61	52	81	81	40	33	56	48
2	48	74	32	02	40	65	18	25	41	45
3	25	17	64	27	51	49	69	57	60	55
4	64	32	68	80	74	52	49	32	54	49
5	89	70	72	65	81	74	44	28	56	57
6	79	60	67	27	65	71	49	28	58	61
7	45	52	52	47	77	58	87	80	85	70
8	49	42	53	41	75	61	87	80	95	72
9	44	49	43	30	67	62	87	86	100	67
10	43	35	47	26	55	50	62	56	74	93

	Rater 3									
R3 1										
2	94									
3	57	60								
4	72	69	66							
5	82	86	80	89						
6	89	87	77	65	83					
7	45	41	46	46	46	47				
8	54	50	46	53	52	59	92			
9	50	47	48	48	48	47	54	88	92	
10	51	44	66	42	50	58	66	62	61	1.0

Table A1 (continued)

Rater 3

R4	1	2	3	4	5	6	7	8	9	10
1	92	75	61	52	82	81	40	33	56	48
2	88	81	68	46	80	86	37	28	51	43
3	60	33	88	59	68	62	54	36	47	61
4	69	49	66	86	83	65	49	40	57	41
5	82	59	81	73	89	75	51	40	55	47
6	84	68	74	52	75	79	45	36	53	53
7	37	43	44	32	66	57	88	85	81	60
8	41	50	43	34	68	64	85	85	87	61
9	39	47	46	25	60	63	88	82	92	61
10	51	21	57	43	58	39	66	54	63	90

Rater 4

R4	1	2	3	4	5	6	7	8	9	10
1										
2	71									
3	73	47								
4	59	32	59							
5	83	63	72	63						
6	73	78	69	33	80					
7	29	30	43	30	62	50				
8	20	36	22	21	56	48	85			
9	44	49	44	31	67	61	86	86		
10	46	33	49	35	58	42	63	67	67	1.0

Appendix E

Table of Expected Mean Square Values

Appendix E

Table A2. Expected Mean Square Values: Two Random - One Fixed Variable Non-Replicated Design.

Source	Expected Mean Square	F Ratio
R	$ds^2_{RxJ} + jds^{2*}_{R}$	$\frac{MS_R}{MS_{RxJ}}$
J	$ds^2_{RxJ} + drs^2_{J}$	$\frac{MS_J}{MS_{RxJ}}$
D	$s^2_{RxJxD} + rs^2_{JxD} + js^2_{RxD} + jrs^2_{D}$	$\frac{MS_D}{MS_{RxD} + MS_{JxD}}$
RxJ	ds^2_{RxJ}	$^{**}\frac{MS_{RxJ}}{MS_{RxJxD}}$
RxD	$s^2_{RxJxD} + js^2_{RxD}$	$\frac{MS_{RxD}}{MS_{RxJxD}}$
JxD	$s^2_{RxJxD} + rs^2_{JxD}$	$\frac{MS_{JxD}}{MS_{RxJxD}}$
RxJxD	s^2_{RxJxD}	-

* s^2 substituted for conventional sigma squared notation.
** No error term; RxJxD substituted.

Appendix F

Mean Job Worth Scores of Four Raters on Three Evaluation Methods

Appendix F

Table A3. Mean Job Worth Scores of Four Raters on Three Evaluation Methods.

Job	Custom	Standard*	PAQ*
1	784	889	695
2	795	843	722
3	161	255	91
4	387	412	408
5	296	319	264
6	652	555	599
7	372	304	387
8	518	468	556
9	647	598	644
10	223	258	260
11	418	420	426
12	313	292	377
13	603	547	754
14	594	574	507
15	610	503	582
16	571	588	620
17	518	494	585
18	509	528	483
19	208	280	175
20	333	365	370

* Transformed to common metric

Bibliography

BIBLIOGRAPHY

Abrahamson, M. Functionalism. Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1978. 110 pp.

Acker, J., "Issues in the Sociological Study of Women's Work," in Stromberg and Harkess, eds., Women Working, Palo Alto: Mayfield Publishing (1979) pp. 134-161.

Adams, J. S., "Inequity in Social Exchange," in L. Berkowitz, ed., Advances in Experimental Social Psychology, Vol. 2, New York: Academic Press (1965) pp. 267-297.

Adams, J. S. and S. Freedman, "Equity Theory Revisited: Comments and Annotated Bibliography," in L. Berkowitz and E. Walster, eds., Advances in Experimental Social Psychology, Vol. 9, New York: Academic Press, 1976.

Alverez, R., K. Lutterman & Assocs. (eds). Discrimination in Organizations. San Francisco: Jossey-Bass, 1979.

Anderson, C. H. and D. B. Corts. Development of a Framework for a Factor Ranking Benchmark System of Job Evaluation. U.S. Civil Service Commission, Personnel Research and Development Center, TS-73-3, December, 1973.

Aristotle, "Nicomachean Ethics, Books I to V," in S. Commins and R. N. Linscott, eds., The Social Philosophers, New York: Random House, 1947.

Arvey, R. D., E. M. Passino and J. W. Lounsbury, "Job Analysis Results as influenced by Sex of Incumbent and Sex of Analyst," Journal of Applied Psychology, Vol. 62 (1977) pp. 411-416.

Ash, P., "The Reliability of Job Evaluation Rankings," Journal of Applied Psychology, Vol. 32 (1948) pp. 313-320.

Atchinson, T. and W. French, "Pay Systems for Scientists and Engineers," Industrial Relations, Vol. 7, No. 1 (1967) pp. 44-56.

Baldwin, R. W., Social Justice. London: Pergamon Press, 1966. 292 pp.

Belcher, D. W., "Wage and Salary Administration," in D. Yoder and H. G. Heneman Jr., eds., Motivation and Commitment, Washington, D. C.: The Bureau of National Affairs (1975) pp. 75-120.

Belcher, D. W. and T. J. Atchison, "Equity Theory and Compensation Policy," Personnel Administration (July-Aug. 1970) pp.22-33.

Bellows, R. and F. Estep, "Job Evaluation Simplified: The Utility of the Occupational Characteristics Check List," Journal of Applied Psychology, Vol. 32 (August, 1948) pp. 354-359.

Blau F. D. and C. L. Jusenius, "Economists' Approaches to Sex Segregation in the Labor Market: An Appraisal," Signs, Vol. I, No. 3 (Spring, 1976) pp. 181-199.

Blumrosen, R. G., "Wage Discrimination, Job Segregation, and Title VII of the Civil Rights Act of 1964," University of Michigan Journal of Law Reform, Vol. 12, No. 3 (Spring, 1979) pp. 397-502.

Blumrosen, R. G., "Wage Discrimination, Job Segregation, and Women Workers," Employee Relations Law Journal, Vol. 6, No. 1 (Summer, 1980) pp. 77-136.

Boulding, Kenneth E., "Social Justice in Social Dynamics," in R. B. Brandt, ed., Social Justice, Englewood Cliffs, N.J.: Prentice-Hall, 1962. 169 pp.

Brandt, Richard B., ed. Social Justice. Englewood Cliffs, N.J.: Prentice-Hall, 1962. 169 pp.

Briscoe, Dennis R. and Larry E. Mainstone, "The Influence of Sex-Race Stereotypes on Personnel Decisions: One More Time," Paper presented at the 1978 Annual Academy of Management Meetings, San Francisco.

Cadieux, R., "Equal Pay for Work of Equal Value," in 1979 Regional Conference Proceedings, Scottsdale, AZ: American Compensation Association, 1979.

Campbell, E. T. and D. W. Fiske, "Convergent and Discriminant Validation by the Multitrait - Multimethod Matrix, Psychological Bulletin, Vol. 56 (1959) pp. 81-105.

Carrell, M. R. and J. E. Dittrich, "Equity Theory: The Recent Literature, Methodological Considerations, and New Directions," Academy of Management Review, Vol. 3, No. 2 (1978) pp. 202-210.

Chesler, D. J., Reliability and Comparability of Different Job Evaluation Systems," Journal of Applied Psychology, Vol. 32 (1948) pp. 465-475. Christal R. E., J. M. Madden and F. D. Harding. Reliability of Job Evaluation Ratings as a Function of Number of Raters and Length of Job Description Used. Lackland AF Base, Texas: Personnel Laboratory, October, 1960, as cited in B. Livy.

Cohen, S. L. and K. A. Bunker, "Subtle Effects of Sex Race Stereotypes on Recruiters' Hiring Decisions," Journal of Applied Psychology, Vol. 60 (1975) pp. 566-572.

Connolly, T., "Information Processing and Decision Making in Organizations," in B. M. Staw and G. R. Salancik eds., New Directions in Organizational Behavior, Chicago: St. Clair Press, 1977.

Cook, R. L. and T. R. Stewart, "A Comparison of Seven Methods for Obtaining Subjective Descriptions of Judgmental Policy," Organizational Behavior and Human Performance, Vol. 13 (1975) pp. 31-45.

Cornelius, E. T. III and K. S. Lyness, "A Comparison of Holistic and Decomposed Judgment Strategies in Job Analyses by Job Incumbents," Journal of Applied Psychology, Vol. 65, No. 2 (1980) pp. 155-163.

Cronbach, L. J., G. D. Gleser, H. Nada and N. Rajaratnam. The Dependability of Behavioral Measurements: A Theory of Generalizeability for Scores and Profiles. New York: Wiley, 1972.

Davis, M. W. and J. Tiffin, "Cross Validation of An Abbreviated Point Job Evaluation Plan," Journal of Applied Psychology, Vol. 34 (1950) pp. 225-228.

Dayton, C. M., Design of Educational Experiments. New York: McGraw-Hill, 1970.

Delafield, G. L., "Social Comparisons and Pay," in G. M. Stephenson and C. J. Brotherton, eds., Industrial Relations: A Social Psychological Approach, New York: John Wiley & Sons (1979) pp. 131-151.

Dubin, R., L. W. Porter, E. F. Stone, and J. E. Champoux, "Implications of Differential Job Perceptions," Industrial Relations, Vol. 13, No. 3 (October, 1874) pp. 265-273.

England, P. and S. D. Mclaughlin, "Sex Segregation of Jobs and Male-Female Income Differentials," in R. Alverez, K. Lutterman and Associates, eds., Discrimination in Organizations, San Francisco: Jossey-Bass, 1979.

Finn, R. H. and S. M. Lee, "Salary Equity: Its Determination, Analysis and Correlates," Journal of Applied Psychology, Vol. 56 (1972) pp. 283-292.

Fitzpatrick, Bernard H., "An Objective Test of Job Evaluation Validity," Personnel Journal, Vol. 28 (1949) pp. 128-132.

Fogarty, M. The Just Wage. London: Geoffrey Chapman, 1961. 300 pp.

Fogel, W., "Occupational Earnings: Market and Institutional Forces," Industrial and Labor Relations Review, Vol 33, No. 1 (1979) pp. 24-35.

Foster, Kenneth E., "New Developments in Job Evaluation," in American Compensation Association, 1977 Regional Conference Proceedings.

Fox, William M., "Purpose and Validity in Job Evaluation," Personnel Journal, Vol. 41 (1962) pp. 432-437.

Frankena, W. K. Ethics. Englewood Cliffs, New Jersey: Prentice-Hall, 1963.

Gardner, J. W. Excellence, Can We Be Equal and Excellent Too? New York: Harper, 1961. 171 pp.

Gomez-Mejia, L. R., R. C. Page & W. W. Tornow, "Meeting the Challenges of Job Evaluation: The Development and Implementation of a Computerized Job Evaluation System," Personnel Admininstration, Vol. 24, No. 2 (1979) pp. 46-52(a).

Gomez-Mejia, L. R., R. C. Page and W. W. Tornow, "Traditional v. Statistical Job Evaluation: Development and Comparison of Different Job Evaluation Approaches, Using the Position Description Questionnaire," Unpublished paper, Control Data Corporation, Minneapolis, Minn., 1979(b).

Grune, J. A. (ed.). Manual on Pay Equity. Washington, D.C.: Conference on Alternative State and Local Policies, 1980.

Gunderson, M., "The Influence of the Status and Sex Composition of Occupations on the Male-Female Earnings Gap," Industrial and Labor Relations Review, Vol. 31, No. 2 (1978) pp. 217-226.

Halaby, C. N., "Job Specific Sex Differences in Organizational Reward Attainment: Wage Discrimination v. Race Segregation," Madison: Institute for Research on Poverty, Discussion Paper #469-77, December, 1977(a).

Halaby, C. N., "Sexual Inequality in the Workplace: An Employer-Specific Analysis of Pay Differences," Social Science Research, Vol. 8, No. 1 (March, 1979) pp. 79-104(b).

Harding, F. D. and D. A. Naurath, "Effects of Job Experience and Organization on the Rating of Tasks," Engineering and Industrial Psychology, Vol. 2 (1960) pp. 63-68.

Harding, F., J. Madden and K. Colson, "Analysis of a Job Evaluation System," Journal of Applied Psychology, Vol. 44 (1960) pp. 354-357.

Hazel, J. T., J. M. Madden and R. E. Christal, "Agreement Between Worker-Supervisor Descriptions of the Worker's Job," Journal of Industrial Psychology, Vol. 2 (September, 1964) pp. 71-79.

Heilbroner, R. L. The Worldly Philosophers. Fourth Edition, New York: Simon and Schuster, 1972. 347 pp.

Hills, F. S., "The Relevant Other in Pay Comparisons," Industrial Relations, Vol. 19, No. 3 (Fall, 1980) pp. 345-351.

Hellberger, C. Job Analysis as Means of Achieving Equal Pay for Working Women: The Case of the Federal Republic of Germany. Paper prepared for conference in equal pay and equal opportunity policy for women in Europe, Canada, and the United States. Wellesley College, May, 1978.

Hogan, J. C., G. D. Ogden, D. L. Gebhardt and E. A. Fleischman, "Reliability and Validity of Methods for Evaluating Perceived Physical Effot," Journal of Applied Psychology, Vol. 65 (1980) pp. 672-679.

Hunady, R. and J. Wahrman, "Influence of Sex-Race Sterotypes on Personnel Decisions," Thirty-seventh Annual Academy of Management Meetings, Orlando, 1977.

Hyman, H. H. and E. Singer. Readings in Reference Group Theory and Research. New York: Free Press, 1968. pp. 3-21.

International Labor Conference, 60th. General Survey of the Reports Relating to the Equal Remuneration Convention and Recommendation. Geneva: ILO, 1975.

International Labour Office. Measures to Facilitate the Application of the Principle of Equal Remuneration for Men and Women Workers for Work of Equal Value, (Meeting of Experts on Women's Work, Report IV). Geneva: December, 1951. 26 pp.

James, L. R. and A. P. Jones, "Perceived Job Characteristics and Job Satisfaction: An Examination of Reciprocal Causation," Personnel Psychology, Vol. 33, No. 1 (1980) pp. 97-135.

Jaques, E. Equitable Payment. New York: John Wiley and Sons, 1961 Jeanneret, P. R., "Equitable Job Evaluation and Classifica-tion with the Position Analysis Questionnaire," Compensation Review, Vol. 12, No.1 (First Quarter 1980) pp. 32-42.

Jenkins, G. D. Jr., D. A. Nadler, E. E. Lawler III, and C. Cammann, "Standardized Observations: An Approach to Measuring the Nature of Jobs," Journal of Applied Psychology, Vol. 60 (1975) pp. 171-81.

Kalleberg, A. L. and J. R. Klugel, "Analysis of the Multitrati-Multimethod Matrix: Some Limitations and an Alternative," Journal of Applied Psychology, Vol. 60 (1975) pp. 1-9.

Kane, J. and E. E. Lawler III, "Performance Appraisal Effectiveness: Its Assessment and Determinants," in B. M. Staw, ed., Research in Organizational Behavior, Greenwich: JAI Press, 1979.

Katerberg, R., F. J. Smith, and S. Hoy, "Language, Time, and Person Effects on Attitude Scale Translations," Journal of Applied Psychology, Vol. 62 (1977) pp. 385-391.

Kavanaugh, M. J., "The Content Issue in Performance Appraisal: A Review," Personnel Psychology, Vol. 24 (1971) pp. 653-668.

Kavanaugh, M. J., A. C. MacKinney, and L. Wolins, "Issues in Managerial Performance: Multitrait - Multimethod Analyses of Ratings, Psychological Bulletin, Vol. 75 (1970) pp. 34-49.

Lawshe, C. H. Jr., "Studies in Job Evaluation: II. The Adequacy of Abbreviated Point Ratings for Hourly Paid Jobs in Three Industrial Plants," Journal of Applied Psychology, Vol. 29 (1945) pp. 177-184.

Lawshe, C. H. Jr. and A. Maleski, "Studies in Job Evaluation: III. An Analysis of Point Ratings for Salary Paid Jobs in an Industrial Plant," Journal of Applied Psychology, Vol. 30 (1946) pp. 117-128.

Lawshe, D. J. Jr., and G. Satter, "Studies in Job Evaluation: I. Factor Analysis of Point Ratings for Hourly Paid Jobs in Three Industrial Plants." Journal of Applied Psychology, Vol. 28 (1944) pp. 189-198.

Lawshe, C. and P. Farbro., "Studies in Job Evaluation: VII. The Reliability of an Abbreviated Job Evaluation System," Journal of Applied Psychology, Vol. 33 (1949) pp. 158-166.

Lawshe, C. H. Jr. and R. G. Wilson, "Studies in Job Evaluation: VI. The Reliability of Two Point Rating Systems," Journal of Applied Psychology, Vol. 31 (1947) pp. 355-365.

Lawshe, C. H. Jr., E. E. Dudek and R. F. wilson, "Studies in Job Evaluation: VII. A Factor Analysis of Two Point-Rating methods of Job Evaluation," Journal of Applied Psychology, Vol. 32 (1948) pp. 118-129.

Livernash, E. R. (ed.). Comparable Worth: Issues and Alternatives. Washington, D.C.: Equal Employment Advisory Council, 1980. 260 pp.

Livernash, E. R., "The Internal Wage Structure," in G. W. Taylor and F. C. Pierson, eds., New Concepts in Wage Determination. New York: McGraw-Hill, Inc. (1957) pp. 140-172.

Livy, B. Job Evaluation: A Critical Review. New York: John Wiley & Sons, 1975. 192 pp.

Maddden, J. M., "A Comparison of Three Methods of Rating Scale Construction, Journal of Industrial Psychology. Vol. 2 (1964a) pp. 43-50.

Madden, J. M. Context Effects in Job Evaluation. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, 1960(a).

Madden, J. M. Familiarity Effects in Evaluative Judgments. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, 1960(b).

Madden, J. M., "Policy-Capturing Model for Analyzing Individual and Group Judgment in Job Evaluation," Journal of Industrial Psychology, Vol. 2, No. 2 (1964b) pp. 36-42.

Madden, J. M., "The Effect of Varying the Degree of Rater Familiarity in Job Evaluation, " Personnel Administration, Vol. 25, (Nov-Dec, 1962) pp. 42-45.

Madden, J. M. and M. M. Giorgia, "Identification of Job Requirement Factors by Use of Simulated Jobs," Personnel Psychology, Vol. 18 (1965) pp. 321-331.

Madden, J. M. and R. D. Bourdon, "Effects of Variations in Rating Scale Format on Judgment," Journal of Applied Psychology, Vol. 48 (1964) pp. 147-151.

Mahoney, Thomas A., "Analytical Issues in the Definition of Comparable Worth," Paper delivered at the 41st Annual Meeting of the Academy of Management, San Diego, August 2-5, 1981.

Mahoney, Thomas A. Compensation and Reward Perspectives. Homewood, III.: Richard D. Irwin, Inc., 1979. 381 pp.

Mahoney, Thomas A. and R. H. Blake, "Occupational Pay as a Function of Sex Stereotypes and Job Content," Paper presented to the National Academy of Management, August, 1979.

Matthews, Jay "On the Equal Pay Issue, the Future is Now for San Jose," The Washinton Post, July 14, 1981.

McCormick, Ernest J. Job Analysis: Methods and Applications. New York: AMACOM, 1979.

McCormick, Ernest J., "Job and Task Analysis," in M. D. Dunnette, ed., Handbook of Industrial and Organizational Psychology, Chicago: Rand-McNally (1976) pp. 651-696.

McCormick, E. J., A. S. Denisi and J. B. Shaw, "Uses of Position Analysis Questionnaires in Personnel Admininstration," Personnel Administrator, Vol. 23 (July, 1978) pp. 50-55.

McCormick, E. J., R. C. Mecham and P. R. Jeanneret. Position Analysis Questionnarie, Technical Manual (System II), Logan, Utah: PAQ Services, Inc., 1977.

McCormick, E. J., P. R. Jeanneret and R. C. Mecham, "A Study of Job Characteristics and Job Dimensions as Based on the Position Analysis Questionnaire," Journal of Applied Psychology, Vol. 56 (1972) pp. 347-368.

McIntyre, S., D. J. Moberg and B. Z. Posner, "Race and Sex Discrimination in the Responses to Write-In Applicants," Paper presented at the 1979 Annual Academy of Management Meetings, Atlanta.

Michigan Department of Labor, Office of Women and Work, A Comparable Worth Study of the State of Michigan Job Classifications, a report prepared under contract to the Michigan DOL by Arthur Young and Company, Detroit, Michigan, 1981.

Miles, M. C., "Studies in Job Evaluation: 9. Validity of a Check List for Evaluating Office Jobs," Journal of Applied Psychology, Vol. 36 (1952) pp. 97-101.

Miller, A. R. in Fair Employment Practices. Washington, D.C.: Bureau of National Affairs, September 10, 1981, p. 4.

Milula, G., "Nationality, Performance, and Sex as Determinants of Reward Allocation," Journal of Personality and Social Psychology, Vol. 29 (1974) p. 435.

Morris, M. B., "Inequalities in the Labor Force: Three Sociological Explanations," in Stromberg and Harkess, eds., Women Working, Palo Alto, California: Mayfield Publishing Co. (1978) 458 pp. Mosel, J. N., S. A. Fine and J. Boling, "The Scalability of Estimated Worker Requirements," Journal of Applied Psychology, Vol. 44 (1960) pp. 156-160.

National War Labor Board. National War Labor Borad Policy on Equal Pay for Equal Work for Women. Washington: 1945 (Research and Statistics Report No. 32).

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967. 640 pp.

Oaxaca, R. L., "Theory and Measurement in the Economics of Discrimination," in F. E. Block et al. eds., Equal Rights and Industrial Relations, Madison, Wisconsin: Industrial Relations Research Association, 1977.

Oettinger, M. P., :Nationwide Job Evaluation in the Netherlands," Industrial Relations, Vol. 4, No. 1 (October, 1964) pp. 45-59.

O'Reilly, C. A. III, G. N. Parlette, and J. R. Bloom, "Perceptual Measures of Task Characteristics: The Biasing Effects of Different Frames of Reference and Job Attitudes," Academy of Management Journal, Vol. 21, No. 1 (1980) pp. 118-131.

Prien, E. P, and S. D. Saleh, "A study of Bias in Job Analysis," Journal of Industrial Psychology, Vol. 1 (1963) pp. 113-117.

Rawls, J. A Theory of Justice. Cambridge, Mass.: Belknap Press, 1971.

Rescher, N. Distributive Justice: A Constructive Critique of the Utilitarian Theory of Distribution. Indianapolis: Bobs-Merill, 1966.

Remick, H. Comparable Worth: Equal Pay for Equal Worth. Paper delivered at the 1977 Annual Meeting of the American Association for Affirmative Action.

Remick, H., "Strategies for Creating Sound, Bias-Free Job Evaluation Plans," in Job Evaluation and EEO: The Emerging Issues, New York: Industrial Relations Counselors, Inc., 1978.

Reynolds, L. G. and C. H. Taft. The Evolution of Wage Structure. New Haven: Yale UNiversity Press, 1956.

Richardson, R. Fair Pay and Work: An Empirical Study of Fair Pay Perception and Time Span of Discretion. Carbondale: Southern Illinois Press, 1971. 124 pp.

Robinson, D. D., O. W. Wahlstrom, and R. C. mecham, "Comparison of Job Evaluation Methods: A Policy-Capturing Approach Using the Position Analysis Questionnaire," Journal of Applied Psychology, Vol. 59 (October, 1974) pp. 633-637.

Rosen, B. and T. H. Jerdee., "Effects of Applicant's Sex and Difficulty of the Job on Evaluations of Candidates for Managerial Positions," Journal of Applied Psychology, Vol. 59 (1974) pp. 511-512(a).

Rosen, B. and T. H. Jerdee, "Influence of Sex Role Stereotypes on Personnel Decisions," Journal of Applied Psychology, Vol. 59 (February, 1974) pp. 9-14(b).

Roll, E. A History of Economic Thought. Englewood Cliffs, New Jersey: Prentice-Hall, 1956.

Runciman, W. G. Relative Deprivation and Social Justice. London: Routlege, Kegan, and Paul, 1966.

Sabine, G. H., "Justice and Equality," Ethics, Vol. 67, No. 1 (October, 1956) pp. 1-11.

Satter, G., "Method of Paired Comparison and a Specification Scoring Key in the Evaluation of Jobs," Journal of Applied Psychology, Vol. 33 (1949) pp. 212-221.

Sackett, P. R., E. T. Cornelius, III, and T. J. Carron, "A Comparison of Global Judgment vs. Task Oriented Approaches to Job Classification," Personnel Psychology, Vol. 34, No. 4 (Winter, 81) pp. 791-804.

Schwab, D. P., Construct Validity in Organizational Behavior," in B. M. Staw and L. L. Cummings, eds., Research in Organizational Behavior, Greenwich, Conn.: JAI Press (1980) pp. 3-43.

Selznick, P. Law, Society, and Industrial Justice. New York: Russell Sage Foundation, 1969. 282 p.

Shinar, E. H., "Sexual Stereotypes of Occupations," Journal of Vocational Behavior, Vol. 7 (1975) pp. 99-111.

Shrout, P. E. and J. L. Fleiss, :Intraclass Correlations: Uses in Assessing Rater REliability," Psychological Bulletin, Vol. 86 (1979) pp. 420-428.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Silvestre, J. J., "Industrial Wage Differentials: A Two-Country Comparison," International Labour Review, Vol. 110, No. 6, pp. 495-514.

Slichter, S. H., J. J. Healy and E. R. Livernash. The Impact of Collective Bargaining on Management. Washington, D.C.: Brookings, 1960.

Smith, Adam. An Inquiry into the Nature and Causes of the Wealth of Nations. New York: Random House, 1937.

Smith, Jr. and M. D. Hakel, "Convergence Among Data Sources, Response Bias, and Reliability and Validity of a Structured Job Analysis Questionnaire," Personnel Psychology, Vol. 32, No. 4 (Winter, 1979) pp. 677-692.

Stevenson, M. H., "Wage Differences Between Men and Women: Economic Theories," in A. Stromberg and S. Harkess, eds., Women Working Palo Alto, CA: Mayfield Publishing Co. (1978) pp. 89-107.

Steiber, J. The Steel Industry Wage Structure: A Study of the Joint Union Management Job Evaluation Program in the Basic Steel Industry. Cambridge: Harvard University Press, 1959. 380 pp.

Stromberg, Ann H. and Harkess, Shirley. Women Working: Theories and Facts in Perspective. Palo Alto, Calif.: Mayfield Publishing Co., 1978. 458 pp.

Sulerot, Evelyne, "Equality of Remuneration for Men and Women in Member States of the EEC," International Labour Review, Vol. 112, Nos. 2-3 (1975) pp. 87-108.

Taber, G. and J. Remick. Beyond Equal Pay for Equal Work: Comparable Worth in the State of Washington. Paper prepared for Conference on Equal Pay and Equal Opportunity Policy for Women in Europe, Canada and the United States. Wellesley, Massachusetts: Center for Research on Women, Wellesley College, 1977.

Thurow, L. C. Generating Inequality. New York: Basic Books Inc., 1975. 257 pp.

Tornow, W. W. and P. R. Pinto, "The Development of a Managerial Taxonomy: A System for Describing, Classifying, and Evaluating Executive Positions," Journal of Applied Psychology, Vol. 61 (1976) pp. 410-418.

Treiman, D. J. Job Evaluation: An Analytic Review. Washington, D.C.: National Academy of Sciences, 1979. 169 pp.

Treiman, D. J. and H. I. Hartman, (eds). Women, Work and Wages: Equal Pay for Jobs of Equal Value. Washington, D.C.: National Academy Press, 1981. 145 pp.

Treiman, D. J. and K. Terrell, "Sex and the Process of Status Attainment: A Comparison of Working Women and Men," American Sociological Review, Vol. 40 (1975) pp. 174-200.

Tsanoff, R. A., "Social Moriality and the Principle of Justice," Ethics, Vol. 67, No. 1 (1956) pp. 12-16.

Turner, J. H. and A. Maryanski. Functionalism. Menlo Park, California: The Benjamin Cummings Publishing Co. (1979) 140 pp.

U. S. Department of Labor, Women's Bureau. The Earnings Gap Between Women and Men. Washington, D.C.: U.S. Government Printing Office, 1979. 22 pp.

VanOtter, Casten, "Sweden: Labor Reformism Reshapes the System," in S. Barkin, ed., Worker Militancy and Its Consequences, New York Praeger (1975) pp. 195-234.

Wiley, L. and W. S. Jenkins, "Methods for Measuring Bias in Raters who Estimate Job Qualifications," Journal of Industrial Psychology, Vol. 1 (March, 1963) pp. 16-22.

GENERAL REFERENCES

Bates, M. W. and R. G. Vail, "Job Evaluation and Equal Employment Opportunity: A Tool for Compliance - A Weapon for Defense," Employee Relations Law Journal, Vol. 1, No. 4 (1976) pp. 535-546.

Behrend, H., "The Effort Bargain," Industrial and Labor RelationsReview, Vol. 10, No. 4 (July, 1957) pp. 503-515.

Benston, Margaret Anne, "Dissonant Chords - Developments Under the Equal Pay Act of 1963," Industrial and Labor Relations Forum, Vol. 13, No. 2 (1979) pp. 67-88.

Campbell, J. P., Psychometric Theory," in M. D. Dunnette, ed., Handbook of Industrial and Organization Psychology, Chicago: Ran McNally (1976) pp. 185-222.

Cook, A. H. Collective Bargaining as a Strategy For Achieving Equal Opportunity and Equal Pay: Sweden and Germany. Paper prepared for Conference on Equal Pay and Equal Opportunity Policy for Women in Europe, Canada, and the United States, Wellesley College, May, 1978.

Cook, A. H., "Equal Pay: Where is It?" Industrial Relations Vol. 14 (May, 1975) pp. 158-177.

Dunnette, M. D., L. M. Hough and R. L. Rosse, "Task and Job Taxonomies as a Basis for Identifying Labor Supply Sources and Evaluating Employment Qualifications," in G. T. Milkovich and L. Dyer, eds., Affirmative Action Planning, New York: The Human Resource Planning Society, 1979.

Elisburg, D., "Equal Pay in the United States: The Development and Implementation of the Equal Pay Act of 1963," Labor Law Journal, Vol. 29, No. 4 (1978) pp. 195-208.

Finklestein, M. O., "The Judicia Reception of Multiple Regression Studies in Race and Sex Discrimination Cases," Columbia Law Review, LXXX (May, 1980) pp. 737-754.

Frankena, W. K. Thinking About Morality. Ann Arbor: University of Michigan Press, 1980.

Ferber, M. A. and H. M. Lowry, "The Sex Differential in Earnings: A Reappraisal," Industrial and Labor Relations Review, Vol. 29 (1976) pp. 377-387.

Freedman, S. M., "Some Determinants of Compensation Decisions," Academy of Management Journal, Vol. 21 (1978) pp. 397-409.

Gillet, Bernard and D. P. Schwab, "Convergent and Discriminant Validities of Corresponding Job Descriptive Index and Minnesota Satisfaction Questionnaire Scales," Journal of Applied Psychology, Vol. 60 (1975) pp. 313-317.

Goldfarb, R. S. and J. R. Hosek, "Explaining Male-Female Wage Differentials in the 'Same Job,'" Journal of Human Resources, Vol. 11, No. 1 (1976) pp. 98-107.

Hammond, K. P. et al., "Human Judgment and Decision Processes," in M. F. Kaplan and S. Schwartz, eds., Human Judgment and Decision Processes, New York: Academic Press, 1975.

Hay, E. N. and D. Purves, "A New Method of Job Evaluation - The Guide Chart Profile Method," Personnel, Vol. 28 (1951) pp. 162-170.

Husband, T. M. Work Analysis and Pay Structure. London, New York: McGraw-Hill, 1976. 232 pp.

Jacques, E., "Taking Time Seriously in Evaluating Jobs," Harvard Business Review (Sept-Oct, 1979) pp. 124-132.

Kerr, C. and L. Fisher, "Effect of Environment and Administration on Job Evaluation," Harvard Business Review, Vol. 28, No. 3 (1950) pp. 77-96.

Krzystofiak, F., J. M. Newman and G. Anderson, "A Quantified approach to Measurement of Job Content: Procedures and Payoffs," Personnel Psychology, Vol. 32, No. 2 (1979) pp. 341-357.

Lawler, E. E. II. Pay and Organizational Effectiveness: A Psychological View. New York: McGraw-Hill, 1971.

Leach, D. E., "An Emerging Concept: Equal Pay for Work of Equal Value," EEOC Vice Chairman Remarks to American Compensation Association Conference, Dallas, 1978.

Leventhal, G. S., "The Distribution of Rewards and Resources in Groups and Organizations," in L. Berkowitz, ed., Advances in Experimental Social Psychological, Vol. 9 (1976) New York: Academic Press.

Levine, E. L., R. A. Ash, and N. Bennett, "Exploratory Comparative Study of Four Job Analysis Methods," Journal Of Applied Psychology, Vol. 65 (1980) pp. 524-535.

Lloyd, C. B. and B. T. Niemi. The Economics of Sex Differentials. New York: Columbia University Press, 1979. 355 pp.

McDermott, A. A., "Lemons versus the City and County of Denver," Personnel Admininstrator, Vol. 25, No. 10 (1980) pp. 95-106.

McLean, R. A. et al., "Compensating Wage Differentials for Hazardous Work: An Empirical Analysis," Quarterly Review of Economics and Business, Vol. 18 (1978) pp. 97-107.

Milkovich, George T., "Pay Inequalities and Comparable Worth," in Industrial Relations Research Association, Proceedings of the Thirty-Third Annual Meeting (September, 1980) pp. 147-154.

Nelson, B., "Wage Discrimination and Comparable Worth Theory in Perspective," University of Michigan Journal of Law Reform, Vol. 13, No.2 (1980) pp. 231-301.

Nieva, V. F. and B. A. Gutek, "Sex Effects on Evaluation," Academy of Management Review, Vol. 5, No. 2 (1980) pp. 267-276.

O'Kelly, C. G., "The Impact of Equal Employment Legislation on Women's Earnings," American Journal of Economics and Sociology, Vol. 38 (1979) pp. 419-430.

Ontario Ministry of Labor. Equal Pay for Work of Equal Value: A Discussion Paper. Toronto, Ontario Ministry of Labor, 1976. 106 pp.

Phelps Brown, E. H. UThe Inequality of Pay. Oxford, 1977.

Pien, E. P. and W. W. Ronan, :Job Analysis: A Review of Research Findings," Personnel Psychology, Vol. 24 (1971) pp. 371-396.

Robertson, P. C., "Compensation EEOC Style," in 1979 National Conference Proceedings, Scottsdale, AZ: American Compensation Association, pp. 1015.

Rosenbaum, J. E., "Hierarchial and Individual Effects on Earnings," Industrial Relations, Vol. 19, No. 1 (1980) pp. 1-14.

Schaefer, D. L., (ed.). The New Egalitarianism. Port Wasington, New York: Kennikat Press, 1979. 264 pp.

Smith, L., "The EEOC's Bold Foray into Job Evaluation," Fortune, (September 11, 1978) pp. 58-60, 64.

Steinberg-Ratner, R. Labor Market Inequality and Equal Employment Policy for Women. Paper prepared for the Organization for Economic Co-operation and Development, September, 1979. 35 pp.

Suzuki, H., "Age, Seniority and Wages," International Labour Review, Vol. 113, No. 1, pp. 67-84.

Taviernier, Gerard, "German Company Sets Equal Pay Standards for All Employees," Compensation Review, Vol. 8, No. 3 (1976) pp. 51-54.

Terborg, J. R. and D. R. Ilgen, "A Theoretical Approach to Sex Discrimination in Traditionally Masculine Occupations," Organizational Behavior and Human Performance, Vol. 13 (1975) pp. 353-376.

Thomsen, D. J. "Discrimination and Job Evaluation," in B. R. Ellig and D. J. Thomsen, Compensation and Benefits: Analytical Strategies, Scottsdale, Arizona: The American Compensation Association (1977) pp. 137-147.

Thomsen, D. J. "Evaluating Job Evaluation Plans," in Non Discriminatory Pay, Los Angeles: Compensation Institute, 1979.

Wallach, M. A. and N. Kogan, "Sex Differences and Judgment Process," Journal of Personality, Vol. 29 (1959) pp. 555-564.

"West German Government Report on Job Evaluation Systems,"

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

Willis, N. D. and A. O. Worcester. State of Washington Comparable Worth Study. Norman D. Willis & Associates, September, 1974.

Wolf, W. C. and N. D. Fligstein, "Sexual Stratification: Differences in Power in the Work Setting, "Social Forces, Vol. 58, No. 1 (September, 1979) pp. 94-107.

Wood, A. A Theory of Pay. New York: Cambridge University Press, 1978. 251 pp.

Zalusky, J., "Job Evaluation: An Uneven World," AFL-CIO American Federationist, Vol. 88, No. 4 (April, 1981) pp. 11-20.