Impact of measurement error on regression coefficients used in the State of Iowa's comparable worth system

by

Jeffrey John Greig

A Thesis Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE

Major: Economics

Iowa State University
Ames, Iowa

1987

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION AND LITERATURE REVIEW

In May of 1983, the Legislature of the State of Iowa passed House File 313, a bill requiring that state employees should be paid on the basis of comparable worth. Specifically, the bill stated:

> It is the policy of this state that a state department, board, commission or agency shall not discriminate in the employment or pay between employees on the basis of gender by paying wages to its employees at a rate less than the rate at which the employer pays wages to employees of the opposite gender for work of comparable worth. "Comparable worth" means the value of work measured by the composite of the skill, effort, responsibility and working conditions normally required in the performance of work (Arthur Young, 1984a, pg. 114).

With the passage of House File 313, the State of Iowa joined numerous other state and local governments, businesses, and individuals in the controversy of comparable worth. Exactly what is comparable worth? How can work of comparable worth be determined? What are some of the problems associated with developing a pay system based on comparable worth? These are a few of the questions that will be addressed in this thesis. To begin, it is important to have a good understanding of the term "comparable worth."

History and Definition of Comparable Worth

The concept of equal pay for work of comparable worth is actually an outgrowth of the Equal Pay Act of 1963. Originally, the bill called for equal pay for comparable work, but the legislature balked at passing a standard that was described by the National Association of Manufacturers as "...so general and so vague as to give an administrator a grant of power which would destroy the sound wage structure which many industrial companies have worked for years to perfect" (Perrin, 1985, pg. 6).

Consequently, the final passed bill stated:

> No employer having employees subject to any provisions of this section shall discriminate within any establishment in which such employees are employed, between employees on the basis of sex by paying wages to employees in such establishment for equal work [emphasis added] on jobs the performance of which requires equal skill, effort, and responsibility, and which are performed under similar working conditions...(Hartmann and Treiman quoting the Equal Pay Act of 1963, 1981, pg. 4).

The following year, the legislature again brought up discussions related to discrimination and employment. Title VII of the Civil Rights Act of 1964 prohibited discrimination based on race, color, religion, sex or national origin in all aspects of employment, including compensation. Once again, though, the act was modified before final passage, and its final version took a much weaker stance against the question of sex discrimination in employment compensation. This modification was the Bennett Amendment. Under the Bennett Amendment, pay differences were allowed in any circumstance when it could be demonstrated that the jobs were not equal. Therefore, jobs that were different in content could be paid different wages regardless of the value of the work.

Although Title VII of the Civil Rights was very successful in eliminating discrimination in employment, the Civil Rights Act and the Bennett Amendment did not apply to the issue of sex discrimination in employment compensation when the jobs were not the same. An attempt to expand the interpretation of the existing discrimination laws from jobs that were "equal" to jobs that were "comparable" occurred with the U.S. Supreme Court decision in the case of County of Washington v. Gunther. In this case, female jail patrons challenged the county on the grounds that wage discrimination existed because their job was worth as much as

the work performed by male jail patrons. The decision opened the door for further claims of sex discrimination based on jobs of similar content being compensated at different rates. The U.S. Supreme Court decision upheld:

> ...that the broader sex discrimination provision of Title VII could be used to prohibit employers from intentionally paying female employees less because of their sex than male employees who were performing different work (Kessler and Williams, 1984, pg. 32).

In the Bureau of National Affairs Special Report (1984, pg. 14), the members of the bureau state that: "Gunther marks the beginning of the modern debate on wage equality." Since the Gunther case was decided on in 1981, there have been numerous other actions concerning comparable worth. Cook (1983) identified active investigations into the issue of comparable worth in fifteen states. Indeed, Eleanor Norton's prophecy in 1979 that comparable worth would become "...the issue of the 1980s" (Norton as quoted by Hutner, 1986, pg. 1) has come of age.

Some of the debate surrounding the concept of comparable worth is due to a lack of a clear definition of comparable worth. Willborn (1986) defined comparable worth as one type of pay equity issue, consisting of disparities in compensation between two occupations that were deemed of equal value due to the occupations' sex composition. For example:

> ...comparable worth requires equal pay for work that is of comparable value to the employer. Value to the employer is defined in terms of the skill, effort, and responsibility required to do the job. Thus, if an employer employs janitors and secretaries and the two jobs, although different, require equal amounts of skill, effort, and responsibility, the employer, according to this understanding of comparable worth, should be required to pay the jobs equally. But that is a pay equity notion; it seems inequitable to differentially compensate two jobs that require equal amounts of skill, effort and responsibility. Comparable worth asks additional questions: Is the pay disparity the result of sex discrimination? Thus, theoretically, instead of comparing janitors to secretaries, comparable worth

> compares the wages of the secretary job as a sex-dominated occupation to the wages the job would be paid if it were not sex-dominated (Willborn, 1986, pg. 3).

Yet, Cook (1983, pg. 3) points out that the terms "comparable worth," "pay equity," and "comparable pay for work of comparable value" are often used interchangeably. Of more importance is the fact that while comparable worth can be defined in theory, it is difficult to set a definition for concepts such as "value" and "worth" in a practical setting. Livernash (1980, pg. 8) was the first to point out this problem when he concluded that "...the concept of comparable worth to date has not been defined in operational terms."

For the purposes of this thesis, comparable worth plans will be simply defined as plans which address differences in compensation rates across occupations that are deemed of equal value due to the sexual composition of those occupations. It is important to emphasize the fact that occupations are being evaluated and compared, not individuals. Therefore, it is not important how much a female employee is compensated as compared to a male employee (not in the comparable worth sense). It is important how much a female-dominated occupation is compensated as compared to a male-dominated occupation when those two jobs are deemed of equal value to the organization.

Now that a brief history and definition of comparable worth has been presented, an examination of the motivation for comparable worth is needed. Why is comparable worth the "issue of the 1980s?" Is there evidence of past wage discrimination by sex?

Research on Female/Male Wage Disparities

The earnings ratio of female to male wages has been used as a source of discussion and used as evidence of possible discrimination for advocates of comparable worth. The earnings ratios is the ratio of women's to men's full-time annual income. Cohen (1971), as sited by Hartmann and Treiman (1981), studied the earnings ratio for full-time salaried and self-employed workers and concluded that women only earned 55% of what men earned. Since that study, others (Blinder, 1973, Corcoran, 1979, Mellor, 1984, Sandell and Shapiro, 1978, and O'Neill, 1981), as sited by Hartmann and Treiman (1981), have estimated that the earnings ratio is between .54 and .69. Does the incidence of low earnings ratios or a "wage gap" imply that females have been discriminated against in the past? Although the data appear to suggest that this might be true, several limitations of the earnings ratio exist. Most importantly, the earnings ratio does not take into account any differences in worker characteristics. For example, if men as a group are more educated, have more experience, or work more hours then certainly it could be expected that men as a group would earn more money.

Attempts to take into consideration any differences in worker characteristics are often referred to as "human capital" approaches to explaining the wage gap. Under human capital theory, it is assumed that differences in the amounts of education, experience, etc. will translate into differences in productivity. Then, a rate of return on investments in human capital can be estimated. Finally, new comparisons between men and women can be made while holding the levels of human capital constant.

Several studies also exist that estimate an earnings ratio that has

been adjusted for differences in human capital investment. The chart below presents the findings from some of the more noted studies.

Study	Observed Earnings Ratio	Adjusted Earnings Ratio
Blinder (1973)	.54	.54
Cohen (1971)	.55	.74
Corcoran (1979)	.67	.76
Corcoran and Duncan (1979)	.74	.85
Mellor (1984)	.65	.65
Sandell and Shapiro (1978)	.66	.74

(Hartmann and Treiman, 1981, pg. 20-21)

It is obvious from the findings above that even differences in human capital investment have failed to explain the difference in earnings between females and males. The highest degree of explanation that could be accounted for by differences in human capital was found by Corcoran and Duncan (1979) to be 44%. That still leaves 56% of the wage gap unexplained.

Hartmann and Treiman (1981, pg. 41-42) conclude:

> The evidence suggests, however, that only a small part of the earnings differences between men and women can be accounted for by differences in education, labor force experience, labor force commitment, or other human capital factors believed to contribute to productivity differences among workers.

Certainly, if the differences between the earnings of males and females cannot be explained by differences in worker characteristics, then part of the residual may possibly be due to discrimination against female workers. If the differences cannot further be explained by the inability of human capital theory to control for important determinants such as the type and quality of training and experience, then comparable worth advocates can use the earnings gap as evidence of discrimination.

Aldrich and Buchele (1986) believe that this inability to explain the earnings gap is a major contributor to the motivation behind the comparable worth issue.

Because the earnings gap between males and females has caused a movement for equal pay for work of comparable value, the next question to address is how is comparable worth implemented? By what means can we determine the comparative "value" of two jobs?

Role of Job Evaluation in Comparable Worth

The discussion of the role of job evaluation in comparable worth is due to the previously discussed problems with defining comparable worth. Specifically, how does any organization measure the amount of value that each job is worth? Comparable worth is a theoretical concept that requires some form of evaluation to make it a practical policy.

An examination of the role job evaluation plays in the comparable worth issue will point out one of the fundamental differences between advocates and critics of comparable worth. While both advocates and critics agree with the numerous studies illustrating the disparity between female and male compensation, sharp differences arise when addressing the methodology of implementing comparable worth.

Job evaluation can be defined as a systematic procedure for analyzing and comparing occupations within an organization. Job evaluation in a comparable worth context refers to the evaluation of female and male-dominated occupations on the basis of the measurable amounts of skill, responsibility, effort, and working conditions involved in performing the work.

The proponents of comparable worth argue that:

> ...job evaluation determines the "worth" of a job. They believe that every job can be evaluated using a common system, that is, that dissimilar jobs can be compared, their value to the organization determined, and this value directly translated into monetary terms (Perrin, 1985, pg. 45).

Kessler and Williams (1984) also support the position that the job evaluation process can be an effective tool to measure the "value" of an occupation to its organization. Treiman (1979) notes that job evaluation systems have been in use for almost a century. He also found that employers considered job evaluation systems to be "...useful tools by which to establish a hierarchy of jobs as a basis for setting salaries" (Treiman as quoted by Remick, 1984, pg. 19). The courts have also been open to the practical use of job evaluations as a procedure to measure job content or worth. In Briggs v. City of Madison, the court held:

> ...that an inference of sex discrimination could be established by demonstrating that a female-dominated job classification was paid less than a male-dominated job classification where the two job classifications were very similar in skill, effort, responsibility, and working conditions. Thus, if male- and female-dominated jobs are determined to be comparable by a job evaluation study ... the difference in pay is the result of sex discrimination and, therefore, is illegal (Willborn, 1986, pg. 61).

For each argument in support of using job evaluation systems to measure job worth, there exists a counter-argument against the use of job evaluations by the critics of comparable worth. One of the major criticisms focuses on the inability of job evaluation to measure a concept of "value." The National Research Council of the National Academy of Sciences (1981), as sited by Livernash (1980), concluded that no standard existed by which job value or worth could be measured. Donald Schwab has also taken a definitive stance against the use of job

evaluations to measure job worth. Schwab (1981, pg. 52), as sited by Livernash (1980), states that job evaluation is:

> ...an inherently subjective set of procedures. The number of judgments made when installing and maintaining a job evaluation system is truly enormous ... as a consequence, job evaluation systems can essentially yield any result you want.

Aaron and Lougy (1986) summarize the beliefs of the critics against job evaluation by stating that overall, job evaluation is an inherently subjective process. They go on to conclude that the market wage paid to labor is the best measure of what the particular laborer is worth. The job evaluation process is too flawed to be viewed as a practical alternative of measuring worth.

Since there is such a difference between the beliefs of proponents and critics of comparable worth on the issue of job evaluations, only one conclusion is possible. Further research examining the process of job evaluation is needed to provide more insight into the possible use of job evaluation as a measurement tool. Specifically, what aspects of the job evaluation process have drawn substantial criticisms? Has there been a call for research in a particular area?

Reliability and Measurement Error

Indeed, there is one aspect of the job evaluation process that has not only drawn criticism, but is also often cited as an area for further research. Because the job evaluation process contains subjective decisions (types of job evaluation systems and the entire job evaluation process will be described in Chapter II), questions of reliability and the problem of measurement error have continually been raised; but seldom investigated.

The reliability of a job evaluation system refers to its consistency. Since subjective decisions concerning the value of the work performed in an occupation are made by an individual or a group of individuals, how likely is it that a different individual or group would arrive at the same findings? Schwab (1981, pg. 52) believes that a lack of reliability is one of the major reasons that job evaluation systems can "...yield any result that you want. Beatty and Beatty (1984), as sited by Livernash (1980), also point out that a job evaluation plan that is questioned concerning reliability is a job evaluation plan that is "fraught" with problems. Willborn (1986, pg. 64) states his position concerning reliability and job evaluation:

> If, however, it yielded different results on successive trials, [job evaluation] would be unreliable; the results on some trials would necessarily be invalid and, hence, any findings of sex discrimination based on the method would be questionable.

The problem of measurement error deals not with the consistency of the subjective decision, but with the accuracy of the decision. The accuracy of job evaluations is often referred to as the construct validity. Construct validity is the correspondence between the measure used (job evaluation) and the concept being measured (job worth). So, it is not only important that job evaluations be reliable, but it is more important that they accurately measure the concept of job worth.

Schwab (1981, pg. 62), as sited by Livernash (1980), summarizes his beliefs regarding comparable worth through the use of job evaluations and the problems of reliability and measurement error:

> ...job evaluation as theoretically prescribed is a mechanism for identifying worth based on job content. ... there currently exists no suitable basis for determining whether job evaluation measures job worth or not (i.e., whether job evaluation is construct valid).

Evidence from job evaluation investigations indicates that problems of reliability where subjective judgments are at issue are substantial, though probably not unresolvable.

Measurement error occurs in the job evaluation process because subjective decisions must be made at some time. An evaluator is faced with a complex and inherently subjective task, while at the same time is restricted by the limitations of human judgement. For example, assume that an individual must make a judgement on the amount of education needed to perform a certain job. This individual will make the decision based on the available information and his personal opinions. The resulting errors in judgement can either be an unintentional consequence of making subjective decisions on unmeasurable items or the consequence of biases influencing the decision. In either case, errors of measurement are made and the resulting findings would be subject to question. (The statistical formulation and impact of measurement error will be discussed in detail in Chapter III.) Overall, the problem of measurement error places serious limitations on the ability of job evaluation systems to accurately measure job worth.

Call for Research

Because the comparable worth advocates have found that one of the major obstacles keeping comparable worth from wide-spread implementation is the critics' continued efforts to discredit the job evaluation process, supporters of comparable worth have cited specific areas that need further research.

Hartmann (1985) points out that very little research has been conducted on the role of job evaluation systems in comparable worth.

Research on job evaluations can examine many areas, but one area in particular that Hartmann (1985, pg. 7) noted is:

> ...ways need to be devised to measure the relative worth of jobs. Since existing job evaluation procedures appear to be the principal available method, attention needs to be devoted to improving job evaluation procedures.

Schwab (1981), as sited by Livernash (1980), even points out that:

> The paucity of research on job evaluation ... may come as a surprise, given the significance of job evaluation to compensation administration.

Hartmann and Treiman (1981) have summed up the need for additional research through the following statements:

> Techniques used in job evaluation have not kept up with developments in econometrics, psychometrics, and sociological measurement. Serious attention should be given to the selection and measurement of compensable factors, the functional form of regression models, and assumptions about error structure, each of which can seriously affect the pay rates predicted by these models (pg. 80).
>
> It may be possible to improve job evaluation plans. ... we urge further research into many unresolved technical issues regarding job evaluation principles and practices (pg. 81).
>
> The development and implementation of a job evaluation plan is often a lengthy and costly process. The underdeveloped nature of the technology involved, particularly the lack of systematic testing of assumptions, does not justify the universal application of such plans. In the committee's judgement, however, the plans have a potential that deserves further experimentation and development.

Purpose and Outline of the Thesis

The purpose of this thesis is to address some of the research issues discussed in the previous section. Specifically, this thesis will examine the impact of measurement error in the job evaluation process on the comparable worth recommendations made by the Arthur Young consultants to the State of Iowa. During the job evaluation process, teams of employees rated each job classification according to the amounts of

skill, responsibility, effort and working conditions involved in performing the job. Since these ratings were the evaluation teams' judgments, the ratings could not be considered the "true" level necessary; but merely the observed level believed to be necessary. Hence, the ratings were subject to measurement error. In subsequent statistical analyses conducted by the Arthur Young firm, the ratings were treated as measured without error. Therefore, the recommendations concerning changes in appropriate pay grades made to the State of Iowa would have to be questioned.

This thesis will examine the procedures followed by the Arthur Young firm and the recommendations originally made. Instead of just considering the impact on male and female classifications, this thesis will analyze other aspects of comparable worth such as supervisory vs. nonsupervisory positions and the impact of comparable worth on different income levels. Then, the statistical analyses conducted by Arthur Young will be reworked utilizing a statistical software package developed at the Iowa State University Statistical Laboratory that will correct for the effect of measurement error.

The outline of this thesis is as follows: Chapter I is a brief background on the issues of comparable worth, job evaluation, and needed research; Chapter II is a detailed description of the procedures and results from the Arthur Young comparable worth study; Chapter III presents an explanation of the problem of measurement error; Chapter IV is statistical analysis of the Arthur Young recommendations and the results of the errors-in-variables regression; Chapter V is a summary of the findings.

CHAPTER II. ARTHUR YOUNG PROCEDURE

The Arthur Young Company of Milwaukee, Wisconsin was selected to accomplish the task of designing and analyzing a single compensation system based on the concept of the comparable worth. The major objective of the project was to "...develop a compensation system for all state of Iowa Merit Employment System employees which is internally equitable and provides comparable pay for positions of comparable value" (Arthur Young, 1984a, pg. 2).

There were several advantages to hiring an outside consulting firm to take responsibility for a project of this magnitude.

1. The consultants are specialists in the job evaluation process. They have gained valuable experience from previous projects.
2. The consultants may provide a more objective point of view which is extremely important in a process that is inherently subjective.
3. If the consultants work closely with the existing employees and supervisors, any doubt concerning the accuracy and honesty of the procedures can be minimized (Lanham, 1955, pg. 21).

Although the consultants had primary responsibility for the project, it was realized that employee participation is crucial in the evaluation stages. A high degree of acceptance and support is only possible when the employees understand the procedures and results. Therefore, Arthur Young utilized state personnel throughout the project.

The technical approach itself can be broken down into several steps.

Step 1: Obtain Job Information

An objective of job evaluation has been defined as "...a method of comparing jobs by use of formal and systematic procedures in order to establish a rank order of the jobs..." (Elizur, 1980, pg. 3). It is then clear that "...the evaluation of jobs can only begin with a careful study

or analysis of their contents. Job analysis is a fundamental prerequisite of job evaluation" (Livy, 1975, pg. 41). The Arthur Young consultants needed to gather complete and unbiased information on the approximately 810 state merit job classifications. It is important to realize that the job classification is being analyzed, not the employee within the job classification. To begin, the consultants chose not to use existing information that could have possible biases against female dominated classifications. This procedure is supported by Charles Lytle (1946, pg. 133) who stated:

> Even if the jobs are well described and classified, the data may be unsatisfactory. If it is worthwhile to have a new evaluation it will usually be best to start with no limitations. The reason is that the new plan may need to bring out certain data which were wholly lacking or not distinct on the old forms. Certainly it would be foolish to economize in the matter of foundational data for anything so important as a lasting job classification.

The consultants began the information gathering process by distributing a classification analysis questionnaire to a sample of 4,500 merit system employees. The analysis questionnaire contained detailed questions concerning the duties, responsibilities, supervision, equipment, contacts and working conditions that the job classification entailed. The 4,500 questionnaires (approximately 25% of all employees) were distributed to all 810 classifications according to the following sampling scheme:

Number of employees in classification	Number sampled
5 or less	All
6 to 100	25%, but no less than 5
101 to 200	25%, up to a maximum of 35
201 to 500	20%, up to a maximum of 50
501 to 1,000	10%, up to a maximum of 75

Number of employees in classification	Number sampled
1,000+	5%, but not less than 75

(Arthur Young, 1984a, pg. 5)

It was also intended that the sample be balanced according to other criteria such as sex and salary step.

The completed questionnaires were checked by the immediate supervisor to verify accuracy and completeness. The questionnaires were further reviewed by the Arthur Young consultants. If a questionnaire was found to be incomplete or ambiguous, the consultants followed up the information with a field audit interview.

Field audit interviews were used by the consultants to obtain additional information or to clear up ambiguous information for 150 employees covering 124 job classifications. This combination method of questionnaires and interviews has several advantages which should be noted.

1. It permits good participation of employee and supervisor in determining job content.
2. The interview allows those job incumbents whose writing skills are below average to fully participate.
3. A well-written questionnaire can eliminate the inclusion of biased information (Lanham, 1955, pg. 162).

After the interviews were completed, the consultants chose five questionnaires from each job classification to forward to the evaluation teams. The criteria for selection were based on completeness of information, clarity, and representation.

Step 2: Choose Evaluation System

Once a detailed and careful description of each job is available, each job needs to be evaluated in terms of its "worth" or "value" to the organization. However, there are several different methods available to evaluate jobs, and the choice of an appropriate evaluation system can certainly affect the end results. The steering committee, set up to oversee the project and the Arthur Young consultants, wanted the evaluation system to meet the following requirements (Arthur Young, 1984a, pg. 8):

* Documented and capable of outside verification.
* Based on the principles outlined by House File 313.
* Equitable and consistent applicability for the range of job classifications evaluated.
* Easily understood by the personnel who will administer the program.
* Flexibility in responding to changes in job functions and organizational design.
* Facilitates periodic auditing.
* Minimum maintenance required on an ongoing basis.

Basically, job evaluation systems can be divided into two types: qualitative systems, and quantitative systems.

Qualitative systems

1. The Ranking Method: The jobs are compared with one another on the basis of overall worth. From the one-to-one comparisons, a overall hierarchy can be developed. The major advantage of a ranking method is its simplicity. The major disadvantage is its reliance on the subjective decisions of the evaluator.
2. The Classification or Grading Method: A series of grades or classes is devised in which the grades are based on significant differences in skill, responsibilities and requirements. Then

the evaluator "fits" the job into the most appropriate grade. Again, the major advantage of this type of system is its ease in use and understanding. The subjectivity of this type of system is its major disadvantage. Also, the classification is based on the whole job, and not on the separate job components.

Quantitative systems

1. The Factor-Comparison Method: A set of key jobs is selected to be compared. These jobs are ranked according to a set of evaluation factors. The factors are generally four or five items such as mental requirements, skills required, physical requirements, responsibility, and working conditions. Then, the money paid to each of the key jobs is allocated and assigned to the evaluation factors according to the relative importance of each item in performing the job. A table is prepared which shows the position of each key job under each factor. The other jobs are then compared to the key jobs on a factor-by-factor basis to arrive at the final hierarchy from top to bottom. The advantage of the factor-comparison system is that it eliminates much of the subjectivity associated with the qualitative systems. However, the factor-comparison system has several limitations. First, the method is difficult to use and understand. Secondly, the selection of key jobs is critical, but no clear foundation exists for selection of these jobs. Finally, if wage inequities exist in the key jobs, these inequities will be carried through the entire system.
2. The Point-Factor Method: A set of compensable factors are

selected and defined. The most common factors are skill, responsibility, effort, and working conditions. These factors are also frequently sub-divided into items such as (Elizur, 1980, pg. 25):

Skill	-	Education
		Experience
		Initiative
		Training
		Knowledge
		Manual dexterity
		Judgement
Effort	-	Mental
		Physical
Responsibility	-	Equipment
		Material or Product
		Public Relations
		Safety
		Work of others
Working Conditions	-	Risk of injury
		Hours of work

Each of these items is broken down into a level of degrees. Then, the relative weight of each factor in importance to overall job worth is determined. Based on the weights, each degree level is assigned a point value. The evaluators then determine which degree level for each factor is necessary to perform each specific job. The total points are added up and the jobs are systematically ranked. Once ranked, a group of key jobs can be priced according to the current market wages, or a pay scale can be developed to convert the total points into different pay levels. The advantages of the point-factor method are its ease in use and ease of understanding by the employees. The major disadvantages of the system are choosing the proper factors and weighing the factors correctly.

Although there are many other types of job evaluation systems such as the guide-chart profile method, the castellian method and the time span of discretion method, most job evaluation systems make use of one of the four previously described processes (Elizur, 1980, pg. 27).

The Arthur Young consultants chose to use the point-factor evaluation method to determine the job worth of the 810 job classifications. E. Lanham, Associate Professor of Management (1955, pg. 54), states a number of reasons why the point-factor method may be chosen as the appropriate system.

1. A graphic and descriptive type of scale is used which is considered by many authorities to be more reliable and valid than any other device.
2. The degree definitions are easy to use, as they are written in terms which are applicable to the types of jobs being rated.
3. The point values of jobs show the relative differences between jobs in numerical terms.
4. The plan increases in accuracy and consistency with use.
5. The plan can be understood easily by employees and supervisors.

The point-factor method has also been described as "...superior to other conventional evaluation methods in reducing the amount of subjective decisions and applying a quantitative analytical approach" (Elizur, 1980, pg. 27). Arthur Young's (1984a, pg. 8) own reason for selecting the point-factor method was:

> Determination of the relative job value was central to this project because it addressed the issue of internal equity, or fair relationship among job classifications. Our prior experience in performing similar studies for government entities suggested that the most appropriate approach was to develop and use a point-factor evaluation plan.

Once the point-factor evaluation system was chosen, emphasis shifted to developing specific steps of the point-factor plan.

Step 3: Choose Evaluation Factors

A primary step in the development of a point-factor system is selecting the appropriate factors to be rewarded or compensated. This process is not only difficult, but it is also extremely important. Livy (1975, pg. 74) points out that:

> Factor identification is most important. For emphasis, it is worth repeating the basic principle that a range of factors must be established which can be applied to the whole gamut of jobs under consideration, that too few factors will reduce the discriminatory powers of the technique, whilst too many will introduce problems of co-variance.

The Arthur Young consultants began the choice of evaluation factors by studying a job evaluation plan that was developed for Iowa merit employees but never implemented. The "Iowa Plan" was a point-factor system that had previously evaluated 113 job classifications (Arthur Young, 1984a, pg. 9). After examining the procedures and results for accuracy, the consultants conducted statistical analyses to identify the importance of the original factors used in the "Iowa Plan." From this information along with knowledge from previous experience and information gained following meetings with analysts and personnel representatives, the consultants identified nineteen potential factors (Woolsey, 1983, pg. 3). Paterson (1972, pg. 63) states: "The usual procedure is for the committee to choose a large number of appropriate factors and then to whittle these down. 'No successful plan has had more than 22 factors. From eight to 11 has proved the most successful range...'" Too many factors would lead to a complex and redundant evaluation system, while too few factors would not be capable of discerning subtle job differences. The Arthur Young consultants and the steering committee

found thirteen factors to include in the point-factor system. The consultants felt that "...the evaluation factors employed are relevant and applicable to the job classifications undergoing evaluation" (Arthur Young, 1984a, pg. 37).

The specific factors chosen and a brief definition are listed below (Arthur Young, 1984a, pg. 28-29).

Factor 1. Knowledge -- from formal training/education. This factor measures the academic preparation and/or technical training at the entry level considered to be "normal" or "typically required" to perform the work. Factor 1 represents the requirements for the job, not the particular educational background of the person holding the job.

Factor 2. Knowledge -- from experience. This factor evaluates the least amount of time normally required for a person with the "typically required" training/education to acquire the knowledge and skills to perform the job satisfactorily.

Factor 3. Job complexity, judgement, and problem-solving. This factor measures the complexity of duties and the frequency and extent of judgement used in decision-making and problem-solving.

Factor 4. Guidelines/supervision available. This factor covers the nature of guidelines and the judgement needed for application. Included are the extent of closeness of supervision required or received for methods to be followed, results to be obtained, and frequency of work progress review.

Factor 5. Personal contacts. This factor measures the responsibility for effective handling of personal contacts with persons not in the supervisory chain. Discussed is the frequency, purpose, importance, setting and person(s) contacted.

Factor 6. Physical demands. This factor measures physical effort and fatigue. Considered is the effort, strength, stamina, and endurance necessary to perform the job.

Factor 7. Mental/visual demands. This factor measures the coordination and dexterity of mind, eye and hand. Factor 7 includes duration and intensity of the coordination and not intelligence or mental development.

Factor 8. Supervision exercised. This factor measures the nature and magnitude for supervising subordinates. Indicated are the number of people supervised and the type of supervisory responsibility.

Factor 9. Scope and effect. This factor measures the relationship between the nature of the work, its purpose, breadth and depth, and the effect of work products or services within and outside the organizational unit.

Factor 10. Impact of errors. This factor measures the likely effect or probable consequences of potential errors made by an individual in the regular course of the work and the opportunity for making such errors.

Factor 11. Working environment. This factor evaluates the conditions under which the job must be performed and the extent to which conditions, i.e. heat, cold, rain, snow, dirty or bloody conditions, fumes, noises, unpleasant social encounters, etc., make the job unpleasant.

Factor 12. Unavoidable hazards/risks. This factor measures the hazards connected with the performance of the job or the extent and seriousness of potential bodily injury that normally exists in performing the job.

Factor 13. Work pace/pressures and interruptions. This factor measures the degree to which the employee is able to maintain continuity of work and to plan the scheduling and priority of job tasks in advance. Indicated are the changes in work volume and frequency of interruption.

Once the evaluation factors are chosen and clearly defined, each evaluation factor must be broken down into a series of degree levels. The job classifications are actually evaluated according to which level is most appropriate to the job for all thirteen factors. For example, all job classifications are rated according to the thirteen factors, but each job does not require the same level within each factor. Therefore, "In order to differentiate among the jobs in terms of their varying requirements, some method of gradations must be provided" (Lanham, 1955, pg. 81). The factor degree levels serve as the basis of gradation. A job that consistently requires high degree levels for all 13 factors

would be expected to rank very highly. The point factor method of job evaluation allows comparisons to be made between jobs based on the individual components of the job, not simply overall job worth. Performing an evaluation on a specific factor is much more simple than evaluating the entire job as a whole.

The Arthur Young consultants divided each evaluation factor into the different degree levels as follows:

Evaluation factor	# of degrees
1. Knowledge - from formal training/education	1 - 8
2. Knowledge - from experience	1 - 6
3. Personal contacts	1A - 5E
4. Job complexity, judgement, and problem-solving	1 - 7
5. Guidelines/supervision available	1 - 5
6. Physical demands	1 - 4
7. Mental/visual demands	1 - 4
8. Work pace/pressures and interruptions	1A - 3C
9. Supervision exercised	1A - 6F
10. Scope and effect	1 - 5
11. Impact of errors	1 - 5
12. Working environment	1 - 4
13. Unavoidable hazards/risks	1 - 5

Simply, a rating of 1 for an evaluation factor would imply that only the minimum amount of that specific factor is needed to perform the job. Conversely, the highest level for each evaluation factor represents a factor for which the maximum level is required to successfully perform the job.

Again, it is very important that these different degree levels be accurately and clearly defined. Language that is biased must be avoided so that the evaluators can make accurate judgments. Lanham (1955, pg. 82) states:

> Having been established, the degrees must be defined as clearly, concisely, and explicitly as possible if raters are to evaluate jobs

> consistently and uniformly. When degree definitions are ambiguous, interpretation will vary widely among those using the scale. As a result, job ratings will also vary widely and inequities will be perpetuated instead of reduced or eliminated.

The Arthur Young consultants recognized the possible biases and problems associated with developing vague or ambiguous degree definitions. Therefore, the consultants utilized all of their past experience to accurately define the degree levels for the thirteen evaluation factors. The consultants point out that "male bias in language defining factors and degrees has also been built into earlier systems. Our efforts were concentrated to overcome these problems and identify factors that appropriately valued all types of state jobs, irrespective of sex" (Arthur Young, 1984a, pg. 11).

Once the evaluation factors are chosen, defined, and subdivided into appropriate degrees, the actual evaluation can be performed.

Step 4: Conduct Job Evaluations

One objective of a sound job evaluation system is that the system should be accepted by both the employees and the supervisors. If the supervisors and employees do not understand the system or do not accept the system, there may be resistance to the implementation of the system. A simple method to increase the probability of acceptance is to involve the employees in the job evaluation process. Elizur (1980, pg. 40) states: "...participation would seem desirable in assuring interest and understanding of the plan and reducing resistance."

The Arthur Young consultants also believed that employee participation in the actual evaluation process was fundamental in assuring the credibility of the entire point-factor system. Thirty-six

state employees were selected to make up nine evaluation teams. Each team was composed of two males and two females in order to minimize the possibility of sex bias in the rating. These individuals were also grouped according to the type of job that they held. Each team was made up of one personnel specialist, one technical/professional employee, one support staff employee, and one supervisory/managerial employee. Furthermore, an attempt was made to balance the teams according to department, age, and geographic region. The benefit of having a committee of employees evaluate the jobs instead of a single personnel specialist can be summed up by Lanham (1955, pg. 218):

> Because participation, reduction in bias and prejudice, and provision for personal knowledge about jobs are all very important in securing accurate and acceptable ratings, the committee plan is the more popular...and is to be recommended more highly than the individual plan of rating.

Since the committees consisted of three individuals who were not familiar with job evaluation processes, it was necessary to train the evaluation teams on the objectives and procedures. The training took place over a three-day period. The major topics addressed were the concepts of comparable worth, job evaluation, group dynamics, recognizing and coping with biases, and the point-factor evaluation plan (Arthur Young, 1984a, pg. 13). A proper training program will yield many benefits. Livy (1975, pg. 122) points out:

> Training can partly overcome deficiencies. Chesler (1948) reports that raters who had been trained, assessed thirty-five jobs, and achieved high reliability co-efficients ranging from 0.93 to 0.99. Similarly, people improve with practice, and there is some evidence to suggests that experience in handling a particular technique leads to speedier, more consistent, results.

The Arthur Young consultants (1984a, pg.13) state that the purpose of

the training session was to "reduce the problem of subtle sex stereotyping and bias in job evaluation." To consistently achieve accurate and unbiased ratings, a structured procedure had to be developed and followed. Consultants supervised the evaluation process to ensure that the proper procedures were being followed.

The actual evaluation of the job classifications was relatively straight forward. Each team member rated the job classification for each of the thirteen factors independently. Then together, the evaluation team compared the individual ratings and determined the final degree level assigned for each of the evaluation factors. One consultant was assigned to three evaluation teams to provide information and supply the evaluation teams with the materials to document the ratings.

Most job classifications were analyzed and evaluated by one team. However, some classifications were evaluated by more than one team to provide information on inter-team reliability. Reliability refers to the ability to accurately rank the evaluation factors consistently. For example, if one evaluation team determines that job classification A requires level 4 of physical demands, what is the probability that other evaluation teams would determine the same ranking? To examine the inter-team reliability, several job classifications were reevaluated by another team. The sampling of job classifications was done according to the following schedule in the first few weeks as a trial run.

# of Classifications	# of Evaluations
20 classifications	2 evaluation teams
6 classifications	3 evaluation teams
2 classifications	9 evaluation teams

(Arthur Young, 1984a, pg. 14)

Following the first few weeks, at least one classification was reevaluated by a different team each day. In the end, ninety-eight job classifications were evaluated by more than one team.

While there are many different techniques to measuring reliability, one method that is frequently discussed in social science literature is the reliability coefficient that is found from the ratio of error variance to total variance for the independent variables (SUPER CARP manual, 1980, pg. 23). After discussions with Dr. Michael White, consultant for Arthur Young primarily responsible for the statistical analyses, it was discovered that this method was used by the consultants to calculate reliabilities.

The Arthur Young consultants report the following reliability coefficients:

Evaluation factor	Reliability coefficient
Knowledge - from education	.92
Knowledge - from experience	.75
Complexity, judgement, and problem solving	.85
Guidelines/supervision available	.73
Personal contacts - purpose	.77
Personal contacts - type	.78
Physical demands	.84
Mental/visual demands	.55
Supervision exercised - nature	.91
Supervision exercised - number	.94
Scope and effect	.73
Impact of errors	.74
Working environment	.71
Unavoidable hazards/risks	.86
Work pace/pressures	.61
Interruptions	.48
Total evaluation	.89

(Arthur Young, 1984a, pg. 15)

The purpose of checking the evaluators for inter-team reliability was

summed up by the Arthur Young consultants (1984a, pg. 14):

> First, it provided an early indicator if there was a systematic difference in evaluation by any team. Second, it served to provide confidence in the system and documentation in establishing that job classifications were being evaluated in accordance with the concept of comparable worth.

From their past experience, the consultants had determined that a reliability coefficient of greater than .70 was acceptable. Only three of the sixteen evaluation factors fell below the .70 reliability coefficient criteria. The reliability coefficient for the total rating was .89. Overall, after some adjustments were made to those factors whose reliability was below .70, the Arthur Young consultants were satisfied with the evaluation process. All of the evaluations were reviewed by the consultants to "...ensure a uniform application system..." (Arthur Young, 1984a, pg. 15). Further reviews were conducted by the steering committee, the Iowa Merit Employment department, and selected department heads. Any inconsistencies were referred back to the evaluation teams for reconsideration.

Step 5: Determine Factor Weights

Once each job classification has been rated in each of the thirteen compensable factors, the relative importance of each factor needs to be determined. Simply adding up the ratings for each job classification to achieve the total rank would be assuming that each of the thirteen factors was equally as important in determining overall job worth. Since this assumption of equality would be incorrect and naive, some method of determining the relative weight of each factor must be developed.

The importance of the relative factor weights cannot be

overemphasized. Treiman (1979) demonstrated that even subtle differences in the weights of the factors can result in major differences in the final hierarchy of job classifications. On the subject of factor weights, Treiman (1979, pg. 6) states:

> The choice of factors, and the choice of how heavily to weigh each factor's contribution to the total score (in point systems), are at the heart of the design of job evaluation systems. For it is the choice of factors and factor weights that determines the relative ordering of jobs on the job worth scale. One set of factors and factor weights may produce a particular ordering of jobs while a different set of factors or a different weighing of factors may produce quite a different ordering.

Patton, Littlefield, and Self's (1964, pg. 148) view on the process of determining factor weights also emphasizes the importance of appropriately deriving the weights.

> Assigning value weights to the factors chosen for the point plan is of key importance. This step, more than any succeeding decision, produces a fundamental effect on the final evaluations of the jobs. Improper weighting decision made at this step can cause completely unforeseen aberrations in relative job values when the point totals for each job are determined.

Given the importance of deriving the appropriate factor weights, what methodology was used by the Arthur Young consultants? Treiman (1979, pg. 7) points out that there are two basic methods:

> They may be assigned directly by the designers of the system to reflect a priori judgments about how much each factor should contribute to the total worth of a job. Alternately, they may be derived empirically....

Actually, the Arthur Young consultants used both methods to determine the factor weights. In the first approach the steering committee, with the help of the consultants, selected a set of factor weights that they believed were most appropriate. Their decision was based upon:

> ...the different impacts on male and female jobs, the reliability in the use of the factors, intercorrelation among factors or factor

redundancy, the statistically derived weights for predicting current grade levels, and the ways the factors actually acted in determining the final point totals for all jobs (Arthur Young, 1984a, pg. 30).

This set of factor weights is listed below:

Evaluation factor	Factor weight
1. Knowledge - from education	15%
2. Knowledge - from experience	10%
3. Complexity, judgement, and problem solving	12%
4. Guidelines/supervision available	5%
5. Personal contacts	10%
6. Physical demands	5%
7. Mental/visual demands	5%
8. Supervision exercised	8%
9. Scope and effect	10%
10. Impact of errors	5%
11. Working environment	5%
12. Unavoidable hazards/risks	5%
13. Work pace/pressures and interruptions	5%
	100%

(Arthur Young, 1984a, pg. 31)

The Arthur Young consultants also statistically derived a set of factor weights. One of the major criticisms of job evaluation systems in general and the point-factor system in particular is the system's reliance on subjective decision-making. In order to avoid these possible criticisms, the consultants wanted an alternative set of weights to be derived objectively through the use of a statistical analysis.

In this statistical analysis, the consultants regressed the thirteen current pay of each job classification on the thirteen evaluation factors. This regression model can be illustrated below:

$$Y_i = X_{jB_j} + e_i$$

where, Y_i = current job classification pay grade
$i = 1, 2, \ldots\ldots 758$ job classifications

X_j = evaluation factor rating
$j = 1, 2 \ldots\ldots 13$ evaluation factors

B_j = evaluation factor weight
j = 1, 2.....13 evaluation factors

e_i = random error term
i = 1, 2.....758 job classifications

The results of this regression would be the relative weight of each factor in determining current pay. Only one obstacle had to be cleared before the analysis could be conducted. The objective of the project was to evaluate all of the State of Iowa job classifications according to the principles of comparable worth. However, a regression analysis that uses current pay grade as the dependent variable will not predict pay on a comparable worth basis, but pay according to the existing system. Therefore, any biases that currently exist will be perpetuated instead of eliminated. The Arthur Young consultants realized this potential problem and took steps to avoid biasing the estimators.

The consultants referred to research conducted by Hartmann and Treiman (1981):

> Treiman and Hartmann suggest two possible statistical methods for determining factor weights which are unbiased on terms of sex. The authors suggest these procedures may be used to create bias-free job evaluation plans. The procedures may also prove helpful in identifying specific instances of pay discrimination. The first method suggested by the authors is to use a multiple regression approach in which current pay or pay grade is predicted from job evaluation factors and an additional variable defined as the percent of female incumbency in each job classification under study. The second approach uses the pay or pay grades of jobs held mainly by men as the standard of "fair" wages (Arthur Young, 1984a, pg. 36).

The Arthur Young consultants divided the job classifications into three groups: male-dominated, female-dominated, and mixed. A male-dominated or female-dominated job classification is a classification containing more than 70% of one sex as incumbents.

The principles underlying the two statistical approaches are clear.

In order to avoid simply perpetuating the existing system, some adjustment must be made to include the possibility that sex bias exists. In the first method, it is believed that the percentage of females in the job classification can be used as a proxy for the degree of discrimination. Once the regression is completed with the additional variable included, the significance of the percent of female incumbency can be tested. If the variable was not statistically significant, then the hypothesis that the sexual makeup of the job classification influenced the pay could not be supported. If the percent female variable was statistically significant, then the hypothesis that sex makeup did not influence job pay would have to be rejected. Also, the percent of female incumbents variable may be controlling for other variables not included in the analysis. If the variable is significant, it could be an indication that a significant determinant ofpay grade has been omitted. As for the second method, if it is believed that males are not discriminated against, then the results obtained from regressing the evaluation factors on the pay grade of male classifications should result in bias-free estimates of the relative weights.

The regression analyses were completed utilizing both methods outlined by Hartmann and Treiman. The relative weights derived from the model based on all jobs including the percent of female incumbents and the model based on male job classifications only are presented below. The corresponding t-statistics are presented in parentheses.

Evaluation Factor	Relative weights	
	All jobs	Male dominated
Knowledge - from education	24.5 (15.67)	27.0 (12.31)

Evaluation Factor	Relative weights	
	All jobs	Male dominated
Knowledge - from experience	12.9 (9.59)	15.9 (8.51)
Complexity, judgement, and problem solving	18.3 (7.76)	18.7 (5.64)
Guidelines/supervision available	6.6 (3.35)	1.6 (.58)
Personal contacts - purpose	9.5 (6.28)	7.8 (3.49)
Personal contacts - type	-.1 (-.36)	1.6 (1.95)
Physical demands	.2 (-.20)	-.2 (.11)
Mental/visual demands	-.6 (3.09)	0.1 (3.66)
Supervision exercised - nature	1.2 (.80)	4.6 (2.06)
Supervision exercised - number	.9 (.73)	-1.1 (-.61)
Scope and effect	9.9 (5.13)	13.6 (4.85)
Impact of errors	10.0 (5.93)	9.7 (4.26)
Working environment	0.04 (-.21)	1.1 (-.76)
Unavoidable hazards/risks	-.1 (.31)	-1.0 (1.28)
Working pace/pressures	.2 (.31)	-.6 (-.70)
Interruptions	0.01 (-.01)	1.3 (1.63)
Percent female incumbents	-6.3 (-12.25)	---

Evaluation Factor	Relative weights All jobs	Male dominated
R^2	93%	92%

(Arthur Young, 1984a, pg. 39)

An interesting result of the statistical analysis was that the percent of female incumbents variable was statistically significant and negative. This would suggest that a hypothesis stating that the male/female composition of a job would have no impact on pay would have to be rejected. Other evaluation factors that were significant under both models were knowledge-from education, knowledge-from experience, job complexity, personal contacts-purpose, mental/visual demands, and the scope and effect factor. Several of the other factors had low relative weights and were not statisitcally significant. So, factors such as personal contacts-type, physical demands, supervision exercised-number, working environment, unavoidable hazards, work pace, and work interruptions did very little to explain differences in pay grades. It is apparent that the evaluation factors cited as statistically significant combine to actually determine pay grade.

To further examine the effects of different sets of factor weights, the Arthur Young consultants did a statistical analysis of the factor weights that were determined subjectively by the steering committee. Although the steering committee set the weights according to the relative importance of each factor as viewed by the committee, variations in degree utilization and the effect of intercorrelations among the factors caused the actual relative weight of each factor to differ from its assigned weight. For example, the committee assigned a weight of 15% to

knowledge-from education and 5% to physical demands. If most job classifications receive the same rating, then the relative impact of the evaluation factor may be smaller than the assigned weight. In essence, the committee assigned a set of weights subjectively, but the relative impact of each of the evaluation factors cannot be determined by examining the assigned weights. To examine the relative impact of the evaluation factors, the consultants did a statistical analysis on the committee assigned weights. This was accomplished by regressing the total points calculated from the assigned weights on the evaluation ratings. The relative weights for the model based on all jobs and the model based on male jobs are presented below with the original assigned weight.

	Relative weights		
Evaluation factor	Committee assigned	Impact on all jobs	Impact on male-dominated
Knowledge - from education	15%	23.7	22.7
knowledge - from experience	10%	15.1	15.7
Complexity, judgement, problem solving	12%	14.6	15.9
Guidelines/supervision available	5%	7.4	7.5
Personal contacts	10%	11.9	12.5
Physical demands	5%	-3.5	-4.8
Mental/visual demands	5%	.8	-.2
Supervision exercised	8%	.5	8.0
Scope and effect	10%	14.5	15.5
Impact of errors	5%	7.4	7.5
Working environment	5%	1.4	-3.0
Unavoidable hazards/risks	5%	-.3	-1.2
Work pace/pressures and interruptions	5%	4.2	3.9

(Arthur Young, 1984a, pg.40)

It is obvious that the committee had one set of relative weights in theory, but in reality the impact of those assigned weights was entirely

different. This result could be a further criticism of determining factor weights by means of a subjective decision.

Other evaluation factors that were significant under both models were knowledge-from education, knowledge-from experience, job complexity, personal contacts-purpose, mental/visual demands, and the scope and effect factor. Several of the other factors had low relative weights and were not statistically significant. So, factors such as personal contacts-type, physical demands, supervision exercised-number, working environment, unavoidable hazards, work pace, and work interruptions did very little to explain differences in pay grades. It is apparent that the evaluation factors cited as statistically significant combine to actually determine pay grade.

Step 6: Determine Total Points

Now that the evaluation of each of the factors is completed and the weighting of the evaluation factors has been determined (or alternatives developed), the total points of each job classification can be derived and hierarchally ranked. The procedure is relatively simple. The weights for each factor determine the number of points to be assigned to the highest degree level. For example, if knowledge - from education has a weight of 15%, then 150 points are assigned to the highest degree level of knowledge - from education (level 8). This system assumes that a job requiring the highest degree level of each factor would rate a perfect score of 1,000. To assign point values to the lower degree levels, two procedures can be used. Lanham (1955, pg. 89) states:

> Two basic approaches to the assignment of point value to degrees of the factors exist; namely, arithmetic and geometric. Under

arithmetic progression, the points between the degrees of a factor are constant. When the geometric approach is followed, the points between degrees increase progressively.

The Arthur Young consultants chose to use a system of geometric progression. To determine the degree of progression, the consultants first determined the ratio of pay grades. Dividing the maximum salary available ($2,612 biweekly) by the lowest salary available ($403.20 biweekly) resulted in a multiple of 6.478. However, since it was unlikely that the highest paying jobs received the highest degree levels for all factors or the lowest paying jobs received the lowest degree levels, a multiple from eight to ten was deemed more appropriate. The final value of progression was determined to be 1.66. Therefore, in most instances, the lower degree levels were assigned a point value that was 1.66 times lower than the point value of the next highest level. In cases where the evaluation factor was a matrix of two sub-factors, a multiple of 1.66 was used along the diagonal and a multiple of 1.288 (square root of 1.66) was used for the off-diagonal elements. The final point structure is given in the following exhibit.

STATE OF IOWA

JOB EVALUATION SYSTEM

FINAL POINT STRUCTURE

Factor	Degree 1	2	3	4	5	6	7	8	Matrix maximum
1. Knowledge-Education	6	10	17	29	46	77	129	150	150
2. Knowledge-Experience	8	13	22	36	60	100			100

State of Iowa job evaluation system final point structure (continued)

Factor	Degree 1	2	3	4	5	6	7	8	Matrix maximum
3. Complexity Judgement-Problem Solving	6	10	16	26	43	72	120		120
4. Guideline/ Supervisor	6	11	18	20	50				50
5. Personal Contacts									
			A	B	C	D			
1.			17	22	28	36			
2.			22	28	36	47			
3.			28	36	47	60			
4.			36	47	60	78			
5.			47	60	78	100			100
6. Physical Demands		11	18	20	50				50
7. Mental Visual		11	18	30	50				50
8. Supervision Exercised									
	A	B	C	D	E	F			
1.	0	0	0	0	0	0			
2.	0	11	14	17	23	29			
3.	0	14	17	23	29	37			
4.	0	17	23	29	37	48			
5.	0	23	29	37	48	62			
6.	0	29	37	48	62	80			80
9. Scope and Effect	13	22	36	60	100				100
10. Impact Errors	6	11	18	30	50				50
11. Work Environment	11	18	30	50					50
12. Hazards-Risks	6	11	18	30	50				50
13. Pace/Interruptions									
			A	B	C				
1.			18	23	30				
2.			23	30	39				
3.			30	39	50				50

(Arthur Young, 1984a, pg. 34)

Once the final point structure was determined, the total points of each job classification could be easily derived. For each evaluation factor the appropriate number of points in accordance with the degree level chosen by the evaluation teams is selected and totaled. Based on the total points, the job classifications can be ranked from top to bottom. The resulting ranking along with the total points can be found in Appendix A.

Step 7: Determine Impact of the Proposed System

The final step was to convert the total point scores into pay grades, and then examine the results for the extent of changes recommended. To begin the process of converting total points to pay grades, the consultants first tried to group similar job classifications. The reasons stated for this procedure consisted of:

1. Relatively modest differences in evaluation results may not indicate true differences in value of positions.
2. Salary administration is simplified when there is a limited number of job classification grades and associated rates.
3. There must be a sufficient number of salary grades or groups of job classifications to reflect differences in pay levels that would normally be expected based on differences in overall job worth (Arthur Young, 1984a, pg. 16).

The pay grade structure recommended by the consultants was based on a 5.2% increase in evaluation points for each increase in pay grade. Therefore, the lowest evaluation total point score was assigned to the lowest available pay grade. Each 5.2% increase in evaluation points was assigned to the next highest pay grade. The evaluation point ranges and corresponding pay grades are presented below:

Grade	Point Ranges
10	142 - 149

Grade	Point Ranges
11	150 - 157
12	158 - 165
13	166 - 174
14	175 - 183
15	184 - 193
16	194 - 203
17	204 - 213
18	214 - 244
19	225 - 236
20	237 - 248
21	249 - 261
22	262 - 275
23	276 - 289
24	290 - 307
25	305 - 320
26	321 - 336
27	337 - 354
28	355 - 372
29	373 - 392
30	393 - 412
31	413 - 433
32	434 - 456
33	457 - 480
34	481 - 504
35	505 - 531
36	532 - 558
37	559 - 587
38	588 - 618
39	588 - 618
40	651 - 684
41	685 - 719
42	720 - 757
43	758 - 796
44	797 - 837

(Arthur Young, 1984a, pg. 19)

In addition to the primary pay recommendations, the consultants also made another set of recommendations based on the statistical analysis conducted previously. The advantage of these recommendations lies in the method of determining the factor weights. Of six models in total, three models predicted pay grade according to the committee assigned weights and three models predicted pay grade using the statistically derived

factor weights. Explanations of the six models are given below.

1. A regression equation developed on all jobs in the study and using total point scores derived from committee assigned factor weights [COMMITTEE WEIGHTS/ALL JOBS].

2. A regression equation based on male-dominated jobs only and using the total point scores derived from committee assigned factor weights [COMMITTEE WEIGHTS/MALE JOBS].

3. A regression equation based on all jobs in the study using total point scores derived from committee assigned factor as a predicator; in determining predicted grade, the regression weight for "percent female" was multiplied times the average percent of female incumbents in all jobs (i.e., 33.5%) [COMMITTEE WEIGHTS/PERCENT FEMALE].

4. A regression equation based on all jobs in the study using statistically derived weights for the job evaluation factors [STATISTICAL WEIGHTS/ALL JOBS].

5. A regression equation based on male-dominated jobs only using statistically derived weights for the job evaluation factors [STATISTICAL WEIGHTS/MALE JOBS].

6. A regression equation based on all jobs in the study using statistically derived weights for the job evaluation factors and including "percent female" as a predictor; in determining predicted pay grade, the regression weight for "percent female" was multiplied times the average percent of female incumbents in all jobs (i.e., 33.5%) [STATISTICAL WEIGHTS/PERCENT FEMALE] (Arthur Young, 1984a, pg. 42).

The model for these six prediction equations is presented below.

PREDICTION EQUATIONS

$$Y_i = X_j B_j$$

where, Y_i = job classification predicted pay grade

$i = 1, 2.....758$ job classifications

X_j = evaluation factor rating

$j = 1, 2.....13$ evaluation factors

B_j = evaluation factor weights under the six models

1. [COMMITTEE WEIGHTS/ALL JOBS]:
   relative impact of committee assigned weights on all jobs.

2. [COMMITTEE WEIGHTS/MALE JOBS]:
   relative impact of committee assigned weights on male jobs.

3. [COMMITTEE WEIGHTS/PERCENT FEMALE]:
   relative impact of comittee assigned weights on all jobs plus the impact of percent of female incumbents.

4. [STATISTICAL WEIGHTS/ALL JOBS]:
   relative impact of statistically derived weights on all jobs.

5. [STATISTICAL WEIGHTS/MALE JOBS]:
   relative impact of statistically derived weights on male jobs.

6. [STATISTICAL WEIGHTS/PERCENT FEMALE]
   relative impact of statistically derived weights on all jobs plus the impact of percent of female incumbents.

j = 1, 2.....13 evaluation factors

The results of these predicted pay grades are listed in Appendix B.

Step 8: Determine Impact of Recommendations

Because a number of different methods were used to derive the factor weights, an entire spectrum of recommended pay grades is available. Since the implications of these various methods is the main issue of this thesis, a more detailed investigation of the different plans can be discussed now.

For the primary pay structure recommended by the Arthur Young consultants, the following chart summarizes pay grade changes according to sex. A classification is considered female-dominated if 70% of the incumbents are female. The same criterion applies to male-dominated

classifications. If a classification is neither 70% male or female, it is considered a mixed classification.

	Female %	Male %
Increase	78.9%	53.9%
Stay the same	10.0%	18.0%
Decrease	11.1%	28.1%

(Arthur Young, 1984a, pg. 42)

Although both male and female dominated classifications would benefit overall, 78.9% of female-dominated classifications would receive improved pay while only 53.8% of male-dominated classifications would move up.

The impact of the alternative recommendations based on the six regression models is summarized below:

		Female %	Male %
COMMITTEE WEIGHTS/ ALL JOBS	Increase	66.10%	31.73%
	Stay the same	8.89%	17.35%
	Decrease	24.98%	50.91%
COMMITTEE WEIGHTS/ MALE JOBS	Increase	72.79%	39.96%
	Stay the same	10.56%	16.67%
	Decrease	16.68%	43.38%
COMMITTEE WEIGHTS/ PERCENT FEMALE	Increase	68.90%	34.69%
	Stay the same	7.78%	15.98%
	Decrease	23.34%	49.32%
STATISTICAL WEIGHTS/ ALL JOBS	Increase	63.89%	28.31%
	Stay the same	15.00%	22.83%
	Decrease	21.11%	48.87%
STATISTICAL WEIGHTS/ MALE JOBS	Increase	81.11%	37.87%
	Stay the same	10.00%	23.97%
	Decrease	8.90%	38.14%
STATISTICAL WEIGHTS/ PERCENT FEMALE	Increase	71.66%	23.07%
	Stay the same	12.22%	21.69%
	Decrease	16.12%	55.24%

(Arthur Young, 1984a, pg. 43)

In all of the alternative pay structures, female-dominated classifications benefit to a greater extent than male-dominated classifications. The consultants point out that the regression analyses for both committee assigned weights and statistically derived weights based on male-dominated classifications would be the most costly structures to implement.

The reason behind this finding is probably due to the method of determining a fair wage. In the first case, the current pay grade was adjusted by the percentage of female incumbents. An alternative method was based on the belief that male job classifications represent the fair wage. The subsequent recommendations led to significant downgrading of male classifications. Therefore, male job classifications may not be a good measure of the fair wage because it appears that male job classifications are overvalued. Hence, any plan based on male job classifications only would tend to overvalue all jobs and prove quite costly to implement. Since the two analyses that are based on all jobs would tend to perpetuate any existing biases instead of eliminating biases, the consultants believe that these plans are not appropriate either. Overall, the consultants would recommend that either plan based on the percent of female incumbents regressions are the best alternatives to the initial pay grade changes that were proposed.

Because of the previously noted problems with implementing the models based on male classifications only and the belief that a model based on all job classifications would perpetuate existing discrimination, the models based on all classifications plus an additional variable representing the percent of female incumbents would be the most

appropriate pay plans to implement.

Although the consultants conducted a thorough study into the problems and procedures associated with developing and implementing a comparable worth system, several questions remain to be answered. First, the possibility of a measurement error problem was not addressed. How would the Arthur Young recommendations change if the information known on the reliability coefficients is incorporated into the analysis? Also, comparable worth is not simply a question of male versus female classifications. What would be the impact of comparable worth on other aspects of the State of Iowa employees? These questions will be examined in the subsequent chapters. To begin, the problem of measurement error needs to be more clearly defined.

CHAPTER III. MEASUREMENT ERROR

The general linear model assumes that the independent variables are measured without error, an assuption that is often not met. Measurement in the social sciences will generally result in measurement error. In the case of the Arthur Young job evaluation process, one source of the measurement error is the evaluation of the job classifications.

During the evaluation process, each evaluation team was directed to rate the job classification in thirteen different categories. Each team came to a consensus concerning the amounts of skill, responsibility, effort, and working conditions involved to perform the job. But the evaluation teams could not have known the "true" amounts required, only their opinions on the amounts of skill, responsibility, effort, and working conditions involved. Therefore, the ratings could not and should not have been treated as variables measured without error. To examine the potential problem of measurement error, a comparison of the results of utilizing ordinary least squares when measurement error does not exist and when measurement error does exist is in order.

Ordinary Least Squares Without Measurement Error

To understand the effect of measurement error in regression analysis, it will be helpful to review the results of using ordinary least squares as a regression tool when no measurement error exists. To simplify the example, assume that only one independent variable is utilized in the regression.

The classical linear regression model would then be illustrated as:

$$Y_t = B_0 + B_1 X_t + e_t \qquad t = 1, 2.....n \qquad (3.1)$$

where, Y_t = dependent variable
(current pay grade in the Arthur Young approach)

B_0 = intercept

B_1 = regression coefficient
(evaluation factor weight in the Arthur Young approach)

X_t = independent variable
(evaluation factor rating in the Arthur Young approach)

e_t = random error

The resulting estimate of the regression coefficient (B_1) can be illustrated as:

$$B_1 = \frac{\sum (X_t - \bar{X})(Y_t - \bar{X})}{\sum (X_t - \bar{X})^2} = \frac{\sigma XY}{\sigma X^2} \qquad (3.2)$$

This estimated coefficient will have the properties of unbiasedness and the smallest variance of any linear unbiased estimator. These properties will only hold true if the following assumption are met:

1. Y is a linear function of X. The linear model must be the best model to represent the relationship between X and Y. If a nonlinear model is more appropriate, then the estimated regression coefficient will no longer be the best linear unbiased estimator of B_1.
2. The error term must be identically and independently distributed with a mean of zero and a constant variance.
3. X must be non-stochastic. X must be fixed and measured without error.

It is obvious from earlier discussions that the Arthur Young consultants used an improper methodology to derive the weights of the

evaluation factors. Subsequently, their findings and pay grade recommendations have to be questioned.

Ordinary Least Squares with Measurement Error

Given the fact that the independent variables (evaluation ratings) were not measured without error, what is the impact of proceeding with ordinary least squares?

Measurement error can be illustrated by the following equation:

$$X_o = X_u + e_m \qquad (3.1)$$

where, X_o = observed value of X
X_u = true, unknown value of X
e_m = measurement error

The classical linear regression model would now be illustrated as:

$$Y_t = B_o + B_1 X_t + e_t \quad t = 1, 2.....n \qquad (3.4)$$

with, $$X_t = X_u + e_m \qquad (3.1)$$

where, Y_t = dependent variable (current pay grade)

B_o = intercept

B_1 = regression coefficient (evaluation factor weight)

X_t = observed independent variable (evaluation factor rating)

X_u = true independent variable (true evaluation rating)

e_t = random error

e_m = measurement error

The resulting estimate of the regression coefficient (B_1) will no longer be equal to formula (3.3), but it will now be:

$$\hat{B}_m = \hat{B}_1 \left(\sigma^2_X + \sigma^2_{e_m} \right)^{-1} \sigma^2_X \qquad (3.5)$$

where, $\hat{B}_m$ = estimated regression coefficient when measurement error exists

$\hat{B}_1$ = unbiased regression coefficient from model with no measurement error (3.3)

σ^2_X = variance of X

$\sigma^2_{e_m}$ = variance of measurement error

Therefore, the estimated coefficient when measurement error exists is biased by the ratio of the true variance of X to total variance of X. The total variance of X is now equal to the true variance of X plus the variance of the measurement error. Since the ratio will be less than one, the direction of bias is towards zero. This only holds true in the single variable case or in the multivariable case when the covariances are equal to zero and the measurement errors are independent. For multivariate analysis, the direction of bias cannot be determined prior to the analysis.

Overall, the effect of measurement error in the independent variable will result in biased coefficients when the model is estimated using ordinary least squares. Any analysis or inferences made using the biased coefficients will necessarily be made in error. In the Arthur Young comparable worth project, the predicted pay grades under six different statistical models are all flawed by the existence of measurement error. How serious a problem actually existed in the Arthur Young procedures? That question will be answered in the next chapter, but it is known that:

> ...measurement error will result in regression estimates different from those that would be obtained had the independent variables been measured without error. This in turn means that the presence of measurement error makes it more difficult to draw sound, substantive conclusions on the basis of regression estimates (Asher, 1983, pg. 251).

Measurement Error Without Statistical Analysis

Since the statistical analyses conducted by the Arthur Young consultants appear to be flawed by their failure to recognize the problem of measurement error, it may seem as though this author is supporting the use of committee assigned factor weights instead of statistically derived factor weights. However, this is not true. Although it is apparent that the statistically derived weights were biased, using subjectively assigned committee weights does not eliminate the problem of measurement error.

It must be remembered that the measurement errors occur when the evaluation team makes subjective decisions concerning the amounts of each compensable factor required to perform a job. The choice of how to weight these factors will not eliminate the errors that have already been made. Under the objective approach of determining the weights statistically, the effect of the measurement error is to bias the resulting coefficients. Under the subjective approach of determining the weights by committee assignment, the weights will not be directly influenced by the measurement error, but the errors are not eliminated.

When the regression coefficients are biased due to measurement error following a statistical approach, procedures now exist to derive unbiased coefficients. Such a procedure will be utilized in the next chapter to determine unbiased factor weights. When the factor weights are assigned by a committee, the only possible correction would be in the committee's wisdom. If the committee recognized that the evaluation ratings were subject to measurement error, it may be possible that they utilized this knowledge appropriately and assigned factor weights to correct the

problem. The chances of assigning appropriate factor weights in the context of a measurement error problem seem highly unlikely. In fact, the standard procedure is to assign weights independent of the measurement process which ignores the possibility of measurement error.

Therefore, I will continue to support the use of statistical weights over committee assigned weights for two basic reasons:

1. A major obstacle in the accepted use of job evaluation as a procedure to determine job worth is the perceived subjectivity involved. Assigning statistically developed factor weights will reduce the amount of subjective decisions made. Using committee assigned weights will only add to the problem.
2. Statistical procedures exist to correct the problem of measurement error. Unbiased statistical weights can be derived. For the committee assigned weights, the only chance for correction lies with the committee's wisdom in the area of measurement error.

Errors-in-Variables Regression Program (EVCARP)

A statistical program designed in the Iowa State University Statistical Laboratory called EVCARP has been developed to solve the problem of measurement error. From the previous discussion of measurement error, the regression coefficient in the single variable case would be biased toward zero. Depending on the information known concerning the measurement error, EVCARP can derive new regression coefficients that correct the measurement error problem and produce unbiased estimates of the coefficients.

There are three different regression programs that are dependent upon what is known or can be assumed about the measurement error. These three types are:

Type 1: Error Variance Known or Estimated

Type 2: Ratio of Error Variance to Total Variance Known

Type 3: Ratio of Error Variances Known

To use EVCARP Type 1, the error variance must be known. It is assumed that the measuring technique can be repeated on the same object in independent trials. If enough independent trials are conducted, the variance of the measurement error ($\sigma^2_{e_m}$) can be estimated. For a sufficient number of independent trials, it can further be assumed that $\sigma^2_{e_m}$ is known. Because the evaluation ratings were conducted by more than one evaluation team, repeated independent measurements were not taken. Therefore, EVCARP Type 1 does not apply to the Arthur Young procedure.

To use EVCARP Type 3, the ratio of error variances must be known for both the independent variables and the dependent variables. The ratio of the error variances is the ratio of the measurement error to the total error. If independent trials had been conducted, information on the measurement error could have been obtained. Information on the error variances of the independent variables was not obtained through a series of repeated trials. Therefore, EVCARP Type 3 does not apply to the Arthur Young procedure.

For the EVCARP Type 2 regression, it is not necessary to assume that $\sigma^2_{e_m}$ is known or estimated. The information needed to use Type 2 concerns the ratio of error variance to total variance for the

independent variables. This ratio is illustrated below:

$$r_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_{e_m}^2} \qquad (3.6)$$

where, r_X^2 = ratio of error variance to total variance

σ_X^2 = total variance of x

$\sigma_{e_m}^2$ = variance of measurement error

How does this pertain to the Arthur Young procedure? In social science literature, the ratio of error variance to total variance is more commonly known as a reliability coefficient. The Arthur Young consultants actually reported the reliability coefficients for the evaluation factors, but they chose to ignore the impending problem of measurement error. However, with the reported reliabilities and EVCARP Type 2, new unbiased coefficients can be determined and used to revise the pay grade recommendations.

To calculate the unbiased coefficients, EVCARP2 transforms the input data before regressing the dependent variable on the independent variables. The first transformation occurs because of the presence of an intercept term. The transformation that is performed corresponds to computing the sums of squares corrected for the mean. Then, a second transformation is performed that adjusts the input data for the presence of measurement error. The adjustment in EVCARP2 is $(1 - r_X^2)$, which is one minus the reliability coefficient. If the reliability coefficient is an indication of how well the variable is measured, then one minus the reliability coefficient is an indication of the amount of measurement

error. After these transformations have been performed, the new coefficients are derived utilizing the transformed input data. As noted before, these new coefficients will be unbiased estimators.

CHAPTER IV. STATISTICAL ANALYSIS

The major emphasis of the statistical analysis is to derive unbiased coefficient estimates for the factor evaluations. These coefficients will be free from the problem of measurement error. A new set of recommendations can then be developed and compared to Arthur Young's previous pay grade recommendations. However, it will be worthwhile to first analyze the six Arthur Young recommendations according to several different criteria. Then, comparison of the new recommendations derived with EVCARP with Arthur Young's recommendations will be more insightful.

Most Beneficial to Female Classifications

Under the premise of comparable worth, the State of Iowa job classifications were evaluated and ranked by the Arthur Young consultants. It should be interesting to see which of Arthur Young's six pay plans would have been most beneficial to the female classifications. Table 1 presents the ranking from most beneficial to least beneficial to female job classifications in terms of net percentage increases most beneficial to least beneficial to female job classifications in terms of net percentage increases for the six pay recommendations.

Table 1. Female job classification changes

Pay recommendation		Percent of female classifications
[STATISTICAL WEIGHTS/ MALE JOBS]	Increase	81.1%
	Stay the same	10.0%
	Decrease	8.7%
[COMMITTEE WEIGHTS/ MALE JOBS]	Increase	72.8%
	Stay the same	10.6%
	Decrease	16.7%

Table 1. continued

Pay recommendation		Percent of female classifications
[STATISTICAL WEIGHTS/ PERCENT FEMALE]	Increase	71.7%
	Stay the same	12.2%
	Decrease	16.1%
[COMMITTEE WEIGHTS/ PERCENT FEMALE]	Increase	68.9%
	Stay the same	7.8%
	Decrease	23.3%
[STATISTICAL WEIGHTS/ ALL JOBS]	Increase	63.9%
	Stay the same	15.0%
	Decrease	21.1%
[COMMITTEE WEIGHTS/ ALL JOBS]	Increase	66.1%
	Stay the same	8.9%
	Decrease	25.0%

The results show a definite consistency between the pay plans. The pay recommendation plans developed with weights derived from male jobs only led to the greatest benefit for female classifications. If the [STATISTICAL WEIGHTS/MALE JOBS] plan would have been adopted, a stunning 81% of all female job classifications would have been increased at least one pay grade.

In every case at least 60% of the female job classifications would have been improved by at least one pay grade. For an advocate of comparable worth, it would definitely appear that work done by mainly female classifications was being undervalued. It is worth pointing out that job evaluations based on comparable worth will not necessarily raise the pay for all females. In five of the six pay plans, at least 15% of the female classifications would have been downgraded.

It is not surprising to find the pay plans based on all job classifications at the bottom of the list. A regression analysis based

on all jobs would yield coefficients that would tend to perpetuate the existing system, not cause major changes. On the other hand, the regressions based on the jobs of male classifications only does not include all of the job classifications, only a selected subset. The pay recommendations based on the pay of male job classifications led to the highest percentage of increases. This is probably due to the possibility that male job classifications are currently overvalued. This question will be examined more closely in the next section.

Most Beneficial to Male Classifications

Although comparable worth is a cause primarily supported by those who would like to see female wages rise, it may be interesting to study the effect of comparable worth on the male classifications. Actually, the effect of the Arthur Young study on male classifications may be of more importance because two-thirds of the state employees are male and 57% of the job classifications are male-dominated. Table 2 presents the ranking from most beneficial to least beneficial to male job classifications for the six pay recommendations.

Table 2. Male job classification changes

Pay recommendation		Percent of Male Classifications
[STATISTICAL WEIGHTS/ MALE JOBS]	Increase	37.9%
	Stay the same	24.0%
	Decrease	38.1%
[COMMITTEE WEIGHTS/ MALE JOBS]	Increase	40.0%
	Stay the same	16.7%
	Decrease	43.4%
[COMMITTEE WEIGHTS/ PERCENT FEMALE]	Increase	34.7%
	Stay the same	16.0%
	Decrease	49.3%

Table 2. continued

Pay recommendation		Percent of Male Classifications
[COMMITTEE WEIGHTS/ ALL JOBS]	Increase	31.7%
	Stay the same	17.4%
	Decrease	50.9%
[STATISTICAL WEIGHTS/ ALL JOBS]	Increase	28.3%
	Stay the same	22.8%
	Decrease	48.9%
[STATISTICAL WEIGHTS/ PERCENT FEMALE]	Increase	23.1%
	Stay the same	21.7%
	Decrease	55.3%

From the data presented in Table 2, it appears inappropriate to discuss changes in male classifications in terms of which recommendation was most beneficial. In all six pay recommendation plans, the overall effect on male classifications was pay grade decreases. As with the female classifications, the pay plans that afforded the most benefit were those based on male jobs solely.

For an advocate of comparable worth, the possibility of discrimination against female-dominated classifications appears to be supported by the Arthur Young recommendations. When rated by an independent evaluation team, a substantial portion of the female classifications are recommended for pay increases while the male classifications are recommended for net decreases. The existing pay plan would seem to highly overvalue work done by male classifications at the expense of female classifications. In four of the six recommended pay plans, approximately 50% of male classifications are targeted for downgrading.

Variation in Male/Female Recommendations

Besides the overall impact on male and female job classifications, a study of the variation in pay grade recommendations by sex may provide additional insights. Did the evaluation teams have a more difficult time in rating female or male jobs? Table 3 presents the average range of the pay grade recommendations for females and males.

Table 3. Variation of pay recommendations by sex

Sex	Average range
Male	2.42 pay grades
Female	2.84 pay grades

The data on pay grade variability by sex would indicate that female job classifications were rated with a little more variation than the male job classifications. However, I would not consider the difference in the average range to be substantial enough to infer any major difference in the variation of pay grade recommendations.

Most Costly Program to the State

Job evaluation on the basis of comparable worth in theory can be totally different than comparable worth in reality. The reality of comparable worth means that a change in the pay structure is going to translate into a change in the wage bill for the organization. Arguments concerning the economic impact of implementing comparable worth on a widespread level are still being debated. Therefore, it may be interesting to see which pay plans would require the largest increases in the State of Iowa payroll. Table 4 presents a ranking from

highest cost to lowest cost for the six pay grade recommendations.

Table 4. Cost of implementation

Pay recommendation	Total pay grade changes
[COMMITTEE WEIGHTS/MALE JOBS]	43,236 pay grades
[COMMITTEE WEIGHTS/PERCENT FEMALE]	31,059 pay grades
[COMMITTEE WEIGHTS/ALL JOBS]	25,150 pay grades
[STATISTICAL WEIGHTS/MALE JOBS]	24,302 pay grades
[STATISTICAL WEIGHTS/PERCENT FEMALE]	10,109 pay grades
[STATISTICAL WEIGHTS/ALL JOBS]	5,854 pay grades

The results presented in Table 4 are consistent with the beliefs of many critics of comparable worth. Although arguments on the exact economic impact of comparable worth implementation are still being debated, critics point out that one sure impact will be an increase in the payroll costs of the organizations that implement comparable worth. Four of the six Arthur Young pay recommendations called for payroll increases of over 20,000 pay grades. The most expensive plan, [COMMITTEE WEIGHTS/MALE JOBS], would require a total of 43,236 pay grade increases to implement. With a state workforce of approximately 24,000, the average increase per worker is close to two pay grades. Certainly, the state could not sustain such a large increase in the wage bill without making adjustments in other areas. A potential reaction by the state may be a reduction in the number of workers employed.

It is interesting to note that all three of the pay recommendations based on committee assigned weights were the most expensive plans.

Also, the statistically derived weights based on all jobs would have led to the smallest number of changes. This could be due to Treiman and Hartmann's belief that this type of model perpetuates the existing system instead of changing it.

Overall, it is obvious that comparable worth cannot be achieved without cost. In times of tightening fiscal budgets and concern over the public sector debts, the cost of implementing comparable worth may be a greater obstacle than the issues concerning the validity of comparable worth. The predicted effect of the Arthur Young recommendations seems to support a statement that Willborn (1986, pg. 29) made concerning comparable worth implementation:

> Employer organizations have estimated that the cost of implementation would be $320 billion and there would be dire economic consequences. Opponents of comparable worth contend that the direct costs are only one of the economic consequences of implementation. Inflation and unemployment are indirect costs of implementation that must also be considered.

It should be pointed out that the final compromise worked out between the state and the employees did not involve the implementation of any of the Arthur Young recommendations. First, it was agreed that none of the job classifications would be downgraded. Obviously, this would in itself cause the cost of implementing one of the recommendations to rise substantially higher. However, the final agreement reduced the number of pay grade increases. This specific case serves as a prime example of how the cost of comparable worth can be a major deterrent to comparable worth implementation.

Effect of Different Levels of Information

In this section it is hypothesized that there may be significant differences in pay grade recommendations between job classifications that have relatively few incumbents and job classifications that are well represented. When the Arthur Young consultants gathered information on job content, questionnaires were used to obtain information on the duties and responsibilities of each job classification. These questionnaires were reviewed by the incumbents' immediate supervisors to ensure accuracy. The evaluation ratings were made based on the pool of information gathered from these questionnaires. Typically, the evaluation team had the five most representative questionnaires available. However, for those job classifications with less than five incumbents, all questionnaires were passed on to the evaluation teams. So, for some classifications, the evaluation teams had less information to base their important decisions on. This is important because the availability of information may have an impact on the degree of measurement error. For classifications with five incumbents or more, the possible measurement errors may be reduced because the errors may average out over the five questionnaires. The "unique" classifications do not have this opportunity. Also, for some high-level positions with only one incumbent, what supervisor examined the responses? For example, who reviews the Director of Human Services' questionnaire for accuracy? The existence of an unintentional "halo bias," perceiving yourself in the best light, has been demonstrated in social science literature (Farr and Landy, 1983, pg. 147).

It is obvious that the evaluation teams made ratings based on two

distinct pools of information. The first was the complete pool of information gathered from a random sample of job incumbents in largely populated claassifications whose questionnaires were reviewed. The second pool of information is made up of a select few questionnaires whose review process is not as clear. Could it be possible that the resulting recommendations reflect differences in the amounts of information available? Table 5 presents the percentage of pay grade increases and decreases of the six recommended plans for the two groups: under five incumbents and over five incumbents.

The results presented in Table 5 do not support the earlier hypothesis that job classifications would benefit under comparable worth because they are evaluated on a smaller base of information. In all six pay recommendations, the job classifications with five or more

Table 5. Pay grade changes by number of incumbents

Pay recommendation		Under five incumbents	Five or more incumbents
[COMMITTEE WEIGHTS/ ALL JOBS]	Increases	33.0%	43.4%
	Decreases	51.0%	42.5%
[COMMITTEE WEIGHTS/ MALE JOBS]	Increases	40.2%	51.6%
	Decreases	42.8%	33.2%
[COMMITTEE WEIGHTS/ PERCENT FEMALE]	Increases	34.6%	47.4%
	Decreases	49.7%	40.5%
[STATISTICAL WEIGHTS/ ALL JOBS]	Increases	33.3%	42.0%
	Decreases	47.7%	34.0%
[STATISTICAL WEIGHTS/ MALE JOBS]	Increases	43.5%	59.7%
	Decreases	37.6%	19.0%
[STATISTICAL WEIGHTS/ PERCENT FEMALE]	Increases	29.1%	43.8%
	Decreases	51.0%	33.9%

incumbents benefited to a greater extent. This result is probably due to the fact that most "unique" jobs are in the high-level positions. Since males dominate the composition of high-level positions, it is not surprising that "unique" jobs did not fare as well as classifications with five or more incumbents. Therefore it appears that the existence of a "halo" bias that would benefit "unique" jobs cannot be supported. However, the question of differences in information will not be dropped yet.

Variation in Pay Recommendations and Differences in Information

Although the results presented in the previous section did not support the hypothesis that "unique" jobs had an advantage over well represented jobs, the variation in pay grade recommendations may provide additional information. Table 6 presents the average range of pay grade recommendations for job classifications with under five incumbents and classifications with five or more incumbents.

Table 6. Variation by number of incumbents

Job classification	Average range
Under five incumbents	2.65 pay grades
Five or more incumbents	2.40 pay grades

Based on the variation in pay grade recommendations, there is some support for the earlier hypothesis. This support comes from the finding that job classifications with under five incumbents appear to be harder to measure than their counterparts with more than five incumbents. With more than five incumbents, there is more information available to make

proper evaluation ratings. Hence, the variation in pay grade recommendations is lower.

To be sure that this difference in the average range is due to the difference in the number of incumbents and not due to the sex of the classification, another analysis was conducted to hold sex of the classification constant. Table 7 presents these results.

Table 7. Variation by number of incumbents and sex

Job Classification	Sex	Average Range
Under five incumbents	F	3.04 pay grades
Under five incumbents	M	2.47 pay grades
Five or more incumbents	F	2.50 pay grades
Five or more incumbents	M	2.36 pay grades

Even when holding the sex of the classification constant, the job classifications with under five incumbents had an average range that was higher than the classifications with more than four incumbents.

Overall, the percentage of changes in pay grades did not support the hypothesis, but the average range of pay grade recommendations did lend some support to the hypothesis that a difference in the level of information exists. When studies are conducted in the future, it may be worthwhile to examine this question more closely.

Pay Grade Recommendations and Income Distribution

Although it is obvious that the comparable worth issue is supported mainly by those who would like to see the pay of women rise, there is some dissention among advocates and critics concerning which income level would benefit the most. Aldrich and Buchele (1986, pg. 133) quote

Clarence Pendleton as stating that comparable worth is "...reparations for middle-class white women." This opinion that the middle-class will reap most of the benefits of comparable worth is also shared by Michael Horowitz. Horowitz argues that comparable worth would "...help middle-class white women at the expense of blacks" (Aldrich and Buchele, 1986, pg. 133). These criticisms of comparable worth are quite important to the chances of the comparable worth issue gaining further acceptance. If it can be shown that only a subset of females (middle-class) will benefit, then the comparable worth movement would lose a large base of support.

What light can the Arthur Young study shed on this issue? In Table 8 the pay grade recommendations are broken down by income distribution. To begin, the 758 job classifications ranged from a low of pay grade 8 to a high of pay grade 46. Breaking the job classifications down into three nearly equal groups was relatively easy. The low income group was pay grade 8 to pay grade 20 and covered 240 classifications. The middle income group covered pay grades 21 through 26 and 250 classifications. The upper income group was classified as classifications with a pay grade at least 27 or higher. This group contained 268 classifications.

Table 8. Pay recommendations by income distribution

Pay recommendation		Upper level	Middle level	Lower level
[COMMITTEE WEIGHTS/ ALL JOBS]	Increase	23.5%	24.0%	72.5%
	Decrease	63.1%	59.6%	12.5%
[COMMITTEE WEIGHTS/ MALE JOBS]	Increase	19.8%	34.8%	90.0%
	Decrease	63.1%	43.6%	1.3%
[COMMITTEE WEIGHTS/ PERCENT FEMALE]	Increase	22.4%	27.6%	79.6%
	Decrease	63.1%	57.2%	9.6%

Table 8. continued

Pay recommendation		Upper level	Middle level	Lower level
[STATISTICAL WEIGHTS/ ALL JOBS]	Increase	23.5%	38.4%	55.4%
	Decrease	50.4%	37.2%	30.0%
[STATISTICAL WEIGHTS/ MALE JOBS]	Increase	31.7%	58.0%	72.1%
	Decrease	42.9%	29.3%	15.8%
[STATISTICAL WEIGHTS/ PERCENT FEMALE]	Increase	17.9%	47.6%	60.4%
	Decrease	57.8%	36.8%	25.8%

From the data presented in Table 8, it is fairly obvious that Pendleton and Horowitz's comments cannot be substantiated. In all six pay grade recommendations, it is not the case that the middle-class benefits at the expense of the lower-class. The lower-class consistently has a higher percentage of classifications increasing and a lower percentage of classifications decreasing.

To be sure that the findings in Table 8 are not being influenced by differences in the sex of the job classifications, another analysis was conducted to hold differences in sex constant. Table 9 presents the average pay grade changes of the six Arthur Young recommendations by income distribution.

Table 9. Pay recommendation by income distribution and sex

Income distribution	Sex	Pay recommendation	
Upper level	F	Increase	32.1%
		Decrease	60.7%
Middle level	F	Increase	24.4%
		Decrease	63.4%
Lower level	F	Increase	67.6%
		Decrease	18.0%

Table 9. continued

Income distribution	Sex	Pay recommendation	
Upper level	M	Increase	5.1%
		Decrease	75.5%
Middle level	M	Increase	14.8%
		Decrease	67.1%
Lower level	M	Increase	24.7%
		Decrease	53.8%

The data in Table 9 reinforce the earlier finding that the lower-class classifications would benefit to the greatest extent if a comparable worth plan were implemented in the State of Iowa. Contrary to previous statements made by Pendleton and Horowitz, comparable worth is not merely a vehicle for the middle-class female. Actually, upper and middle level females do not benefit at all. Table 9 illustrates the fact that upper level and mmiddle level females are targeted for net decreases, much the same as the male classifications. Aldrich and Buchele's own findings on this issue compare with the results above. In a study that utilized U.S. Census data from 1980, Aldrich and Buchele estimated the impact of comparable worth on the distribution of income in the United States. Aldrich and Buchele (1986, pg. 146) found:

> Comparable worth is, of course, necessarily a women's issue. Is comparable worth a middle-class, white women's issue? Our answer is a qualified no.

Variation in Pay Recommendations by Income Distribution

Since the issue of the variation in pay recommendations has come up in previous sections of this chapter, it may be interesting to see if the distribution of income has any impact on variation. Table 10

presents the average range of the six pay recommendations classified by income distribution.

Table 10. Variation in pay recommendations by income level

Income distribution	Average Range
Upper level	2.32 pay grades
Middle level	2.28 pay grades
Lower level	3.10 pay grades

From the data in Table 10, there is a significant difference in the average range of the low end as compared to the middle and high groups. Some of the difference can be attributed to differences in the average range of females and males. It was discovered earlier that the variation in female pay grade recommendations was higher than the variation for males. Since the lower end of the income distribution is dominated by the female classifications, some of that variation is carrying through. Also, a previous analysis discovered that the low-class occupations would benefit the most from the implementation of comparable worth. This may also add to the variation in pay grade recommendations. This may be due to the possibility that it is easier for classifications at the lower end of the scale to change pay grades. Because pay grade levels are set at a 5.2% increase, lower pay grades have a smaller range of points than the higher pay grades. However, such a significant difference cannot be easily explained away. This may be one aspect of comparable worth that requires further investigation.

Pay Grade Recommendations by Management Level

In the process of examining for differences in pay grade recommendations attributed to each income distribution, the method of separating the job classifications was somewhat arbitrary. Although dividing the range of job classification pay grades into three groups led to interesting findings, there are alternative methods of grouping the job classifications. One alternative is to examine the differences in pay grade recommendations for supervisory positions as compared to non-supervisory positions. This comparison can be evaluated to determine if the findings that low-class jobs benefited the most under comparable worth can be supported. Table 11 presents the pay grade recommendations for supervisors versus nonsupervisors.

Table 11. Pay grade recommendations by management level

Pay recommendation		Supervisors	Nonsupervisors
[COMMITTEE WEIGHTS/ ALL JOBS]	Increase	36.4%	40.9%
	Decrease	49.1%	43.9%
[COMMITTEE WEIGHTS/ MALE JOBS]	Increase	35.4%	54.2%
	Decrease	45.4%	31.9%
[COMMITTEE WEIGHTS/ PERCENT FEMALE]	Increase	35.4%	46.5%
	Decrease	48.5%	41.5%
[STATISTICAL WEIGHTS/ ALL JOBS]	Increase	36.8%	39.6%
	Decrease	39.2%	39.8%
[STATISTICAL WEIGHTS/ MALE JOBS	Increase	57.4%	46.4%
	Decrease	24.6%	29.6%
[STATISTICAL WEIGHTS/ PERCENT FEMALE]	Increase	32.0%	41.5%
	Decrease	45.0%	38.1%

The data in Table 11 lend support to the previous findings. Again, the lower income classifications would benefit to a greater extent than the upper income classifications. Under five of these six pay recommendations presented in Table 11, the nonsupervisory positions increased at a higher percentage than the supervisory positions. The method of break-down between upper and lower income classifications for the two approaches was different, but the results were the same. Again, based on supervisory versus nonsupervisory positions, the impacts predicted by Pendleton and Horowitz on income distribution would not necessarily occur.

Variation in Pay Recommendations by Management Level

Since the use of management level as a factor separating income levels led to the reinforcement of the earlier findings, it may be interesting to see if division by management level can help explain some of the variation in pay recommendations for different income levels. From an earlier section, it was discovered that a substantial variation in pay grade recommendations existed for lower income classifications. At the time the only possible explanations that could be forwarded were the effect of a large concentration of female classifications in the lower level or the overall effect of a high percentage of increases recommended. Now that the income distribution is broken down by another criteria, does the supervisory versus nonsupervisory data exhibit the same variation? Table 12 presents the average range of pay grade recommendations for supervisory and nonsupervisory classifications.

Table 12. Variation in pay grade recommendation by management level

Management Level	Average Range
Supervisory	2.28 pay grades
Nonsupervisory	2.72 pay grades

The data on the variation of pay grade recommendations for supervisory and nonsupervisory classifications also support the previous findings that lower income classifications have a higher variation. The fact that female classifications are concentrated in nonsupervisory positions is the best explanation for the differences in variation by management level.

Pay Grade Recommendations by Pay Plan

Currently, a single compensation plan does not exist in the State of Iowa. One of the goals of the Arthur Young project was to develop one plan that could encompass all of the different job classifications currently covered by a variety of pay plans. Certainly then, any plan that was recommended for implementation would have to be evaluated for its effect on the existing pay plans. The state and Arthur Young may find that several current pay units would be extremely opposed to the implementation of a comparable worth plan because of the plan's projected impact on their salaries. This possibility is important because some of these different pay plans are determined through collective bargaining.

What role have unions played in the issue of comparable worth? Several major unions have made comparable worth a bargaining issue. The Bureau of National Affairs stated in their Special Report (1984, pg. 73-

74) concerning unions that:

> The American Federation of State, County and Municipal Employees (AFSCME), one of the largest public sector unions in the country, has taken such a leading role in the pay equity area that some observers refer to it as the "shadow EEOC." According to reliable estimates, most of the more than 260 charges alleging sex-based wage discrimination that are currently backlogged at EEOC have been filed by the union. Other unions' strategies include: negotiating collective bargaining agreements that include equitable wage increases, supporting pay equity legislation, and filing litigation against employers.

What effect would the six pay recommendations have on some of the current pay plans that are strongly unionized? Table 13 presents the pay recommendations effects on the following current bargaining units:

1 - Clerical
2 - Technical
3 - Blue collar
4 - Fiscal and staff
5 - Human services
6 - Security
7 - Public Safety
8 - Professionals
0 - Not designated

I believe that the data in Table 13 indicate the reason why unions have taken a pro-comparable worth stance. The existing bargaining unit that would benefit the most from the implementation of a comparable plan is the clerical unit. In all six of the pay recommendations, at least 80% of the job classifications would be upgraded. Another bargaining unit that could benefit a great deal is the blue collar unit. If any of the pay recommendations based on committee weights are chosen, at least 60% of the blue collar classifications would be increased at least one pay grade.

What do the clerical and blue collar bargaining units have in common? They are probably the two units where union power would be at

Table 13. Pay grade recommendations by bargaining unit

Pay recommendation		Bargaining unit 1	2	3	4	5	6	7	8	0
[COMMITTEE WEIGHTS/ ALL JOBS]	Increase	91.5%	39.1%	64.0%	19.4%	36.0%	38.9%	25.0%	27.0%	36.4%
	Decrease	6.4%	47.9%	15.3%	67.5%	48.0%	38.9%	50.0%	58.2%	45.5%
[COMMITTEE WEIGHTS/ MALE JOBS]	Increase	95.7%	52.1%	83.8%	17.5%	38.0%	53.7%	25.0%	27.7%	63.6%
	Decrease	2.1%	33.1%	6.3%	66.3%	30.0%	27.8%	50.0%	52.5%	22.7%
[COMMITTEE WEIGHTS/ PERCENT FEMALE]	Increase	93.6%	44.4%	73.0%	18.8%	38.0%	40.7%	25.0%	27.7%	40.9%
	Decrease	4.3%	45.6%	14.4%	67.5%	46.0%	33.3%	50.0%	57.5%	36.4%
[STATISTICAL WEIGHTS/ ALL JOBS]	Increase	80.9%	44.4%	43.2%	23.8%	50.0%	27.8%	0.0%	34.0%	22.7%
	Decrease	6.4%	38.5%	42.3%	45.6%	28.0%	40.7%	50.0%	43.3%	59.1%
[STATISTICAL WEIGHTS/ MALE JOBS]	Increase	95.7%	61.5%	52.3%	39.4%	68.0%	48.2%	25.0%	44.0%	45.5%
	Decrease	4.3%	20.7%	28.8%	33.8%	16.0%	31.5%	50.0%	30.5%	36.4%
[STATISTICAL WEIGHTS/ PERCENT FEMALE]	Increase	91.5%	46.2%	37.8%	24.4%	44.0%	27.8%	0.0%	28.4%	36.4%
	Decrease	4.3%	34.9%	43.2%	48.8%	28.0%	46.3%	50.0%	47.5%	63.6%

1 - Clerical.
2 - Technical.
3 - Blue collar.
4 - Fiscal and staff.
5 - Human services.
6 - Security.
7 - Public safety.
8 - Professionals.
0 - Not designated.

its greatest. It is not surprising to see unions take a stance in favor of comparable worth when the possible benefits of comparable worth implementation are so great.

The two bargaining units that would do the worst under a new professionals unit. For fiscal and staff, any implementation would comparable worth compensation plan are the fiscal and staff unit and the professionals unit. For fiscal and staff, any implementation would translate into a downgrading of at least 33% of the job classifications. The professionals unit would have at least 30% of their classifications downgraded if any of the pay recommendations were implemented. Not surprisingly, these two units probably have the lowest union power of all the bargaining units.

Overall, it is clear that the different pay recommendations will have different impacts on the current bargaining units. Any plan for implementation of one of these recommendations must take into account the impact and possible response of a bargaining that is well represented by a union or collective bargaining. As comparable worth becomes a larger issue in the future, unions will probably take a more active stance in shaping the future of comparable worth. Hutner (1986, pg. 209) summarizes the current position of unions as:

> Faced with severe declines in union membership in the traditional union strongholds of heavy industry, mining, and transportation, some unions have mounted vigorous campaigns to organize workers in the service sector. In the process, they have discovered that eliminating sex-based wage discrimination is an issue of primary concern to women workers. Consequently, they have made it a key organizing issue.

Variation in Pay Recommendations by Bargaining Unit

It is not only useful to examine the impact of the six recommended pay plans on the different bargaining units, but examining the variation of these recommendations can provide additional information. Remember, one of the main criticisms against using job evaluations as an indication of job worth is the question of the job evaluation's ability to yield consistent results. The job evaluation process is based on the information collected concerning the job classification's duties and responsibilities. Any technique or method that can improve upon this information base has the possibility of improving the entire process. Examining the variation of the pay recommendations may give indications of where the information pool is at its lowest.

Table 14 presents the average range of pay recommendations according to the bargaining unit involved.

Table 14. Variation in pay recommendations by bargaining unit

Bargaining unit	Average Range
Clerical	2.87 pay grades
Technical	2.57 pay grades
Blue collar	3.30 pay grades
Fiscal and staff	2.75 pay grades
Human services	2.06 pay grades
Security	2.19 pay grades
Public safety	1.75 pay grades
Professionals	2.01 pay grades
Not designated	2.13 pay grades

Table 14 provides some interesting data. You would have to identify blue collar and public safety bargaining units as outliers from the other bargaining unit variations.

The public safety variation is due to small sample size. Only four classifications are covered by the public safety bargaining unit. Therefore, the sample size is too small to make any inferences into the reasons why the variation is so low.

However, the variation associated with the blue collar bargaining unit cannot be as easily explained. This variation is nearly one-half a pay grade higher than the next closest unit; clerical. A priori, I would have thought that the clerical unit would have the highest variation, although it does have the second highest variation. Based on the other findings, groups that were dominated by female classifications tended to have the highest average range. One possible explanation is that the variation is due primarily to the extreme female and male percent incumbents in the model based on percent female. The clerical unit may have a high variation because the incumbents are almost totally women. Along the same line, the blue collar unit may have a high variation because the incumbents include almost no women. These two extremes of female/male composition may be the cause of high variation. Again, this is an area to which more resources and attention could be devoted.

Pay Recommendations and the 70 Percent Rule

Throughout the Arthur Young analysis, job classifications were divided into three groups: male-dominated, female-dominated, and mixed. The criterion for this break-down were the 70 percent rule. If more than 70 percent of the incumbents of a classification are of one sex, then that classification is "dominated" by that sex.

I do not know how this seemingly arbitrary number was derived, but it has become a standard measurement in comparable worth literature. While the advantages and disadvantages of setting such an arbitrary criterion will not be discussed here, it would be interesting to examine the impact if this rule is not used.

When the 70 percent rule was utilized for the first time, it may have been the ideal criterion to use. For job classifications in the State of Iowa though, the 70 percent rule does not seem to be a good measure of "dominance." The reason is simple; a majority of the job classifications for the state are dominated by males. In fact, male job classifications make up approximately 67 percent of the state's work force. If 70 percent is used as the criterion for dominance, it is very easy to obtain male-dominated classifications due to randomness.

For example, assume that job classification A has ten incumbents. What are the probabilities that this classification will be male- or female-dominated? Using the formula for the probability of an event occurring from a binominal distribution:

$$[\ P(X=x) = f(x) = \binom{n}{x} p^x q^{n-x}\ ],$$

where, $P(X = x)$ probability of exactly x successes

P = probability of success: males 67%, females 33%

q = probability of failure: males 33%, females 67%

x = number of successes

n = number of trials: $n = 10$

it can be calculated that the probability that the classification will be male-dominated is 57 percent and the probability that the

classification will be dominated by females is 2 percent. Certainly, if a simple random distribution of males and females could lead to such a large number of classifications being dominated by males, then the 70 percent rule may not be the appropriate criterion for the State of Iowa.

To examine the impact of another method, all of the job classifications with under five incumbents were eliminated from the study. This would keep the "unique" jobs from overly influencing the results. By eliminating these jobs, we can actually examine two questions simultaneously. The effect of changing the job dominance criteria can be studied and the effect of an equal information base for the evaluations can also be examined. Eliminating the classifications with under five incumbents reduces the number of remaining job classifications to 306. Of these 306, the number of male-dominated job classifications using the 70 percent rule is 207, the number of female-dominated is 66, and the number of mixed is 33.

Eliminating the "unique" classifications may correct for any problem due to a difference in information between classification types. To correct for the problem of the arbitrary 70 percent rule, the following analysis was conducted. Nothing was done to change the 66 female-dominated classifications. These classifications represent "true" female dominance. For females to compromise 70 percent of a classification given the fact that only one-third of the state employees are female constitutes dominance. For the males the 66 classifications with the highest percentage of males were chosen. The classifications ranged from 100 percent male to approximately 85 percent male. It may be more accurate to consider 85 percent male incumbents as dominance

when 67 percent of the state employees are male. Since it is relatively easy for male dominance to occur, only the highest degree of male dominance was used.

Table 15 presents the pay recommendations for male and female classifications that fit the above-mentioned requirements and the original recommendations.

Table 15. Pay grade recommendations by Sex[a]

Pay recommendation		Original 70% rule Male	Original 70% rule Female	New Dominance Criterion Male	New Dominance Criterion Female
[COMMITTEE WEIGHTS/ ALL JOBS]	Increase	31.7%	66.1%	30.3%	76.2%
	Decrease	50.0%	25.0%	52.8%	15.4%
[COMMITTEE WEIGHTS/ MALE JOBS]	Increase	40.0%	72.8%	41.4%	84.0%
	Decrease	43.3%	16.7%	43.3%	11.4%
[COMMITTEE WEIGHTS/ PERCENT FEMALE]	Increase	34.7%	68.9%	36.5%	80.7%
	Decrease	49.3%	23.3%	48.8%	14.7%
[STATISTICAL WEIGHTS/ ALL JOBS]	Increase	28.3%	63.9%	24.8%	71.5%
	Decrease	48.9%	21.1%	46.0%	13.9%
[STATISTICAL WEIGHTS/ MALE JOBS]	Increase	37.9%	81.1%	34.6%	91.4%
	Decrease	38.1%	8.9%	35.2%	2.6%
[STATISTICAL WEIGHTS/ PERCENT FEMALE]	Increase	23.1%	71.7%	22.2%	83.8%
	Decrease	55.2%	16.1%	51.7%	9.6%

[a]The male and female classifications in this analysis have over five incumbents. The male classifications are the 66 classifications with the highest percent male.

From the information in Table 15, it does not appear that changing the criteria for determining male and female classifications has affected the outcome of comparable worth. Despite the change of the dominance criteria, women still benefited to a greater extent than the

male classifications. The impact of changing the dominance criteria was for women to benefit to even a greater extent. The male classifications did not exhibit very much change from the original recommendations. Overall, changing the 70% did not dramatically alter the results. Perhaps other analyses had been conducted in previous projects, and it was found that altering the 70 percent rule did not change the results substantially. Actually, when the weights are determined from a regression based on all classifications, whether or not the classifications are dominated does not matter. The sex of the composition was only important when weights were determined on the male classifications and for use in descriptive statistics. Any criticism of using the arbitrary 70 percent rule should be directed at its use in the regression procedures that set factor weights based on the pay of male classifications. It was demonstrated earlier that achieving a classification of 70 percent males when two-thirds of the employees are already male is not an accurate determinant of job classification dominance.

Weighted Least Squares

Before moving on to the analyses that examine the impact of measurement error on the evaluation factor coefficients, one more study of the differences in information level is appropriate. From the earlier results, it was discovered that job classifications with five or more incumbents had a lower average range of pay recommendations than the job classifications with under five incumbents. Various other examinations pointed out differences in the average range, but they were

unable to provide any clear explanations.

The difference in information levels could be a major source of the existing measurement error. It would be very interesting to compare reliability coefficients for job classifications with five or more incumbents to the other. Based on the previous findings in this thesis, I would hypothesis that job classifications with five or more incumbents would consistently be evaluated with more accuracy.

Unfortunately, such information does not exist. One analysis that can be conducted is a weighted least squares regression. Under weighted least squares, job classifications with five or more incumbents would be weighted five times as much as a classification with only one incumbent. Classifications with four incumbents would be weighted four times and so on for classifications with three and two incumbents. By weighting the job classifications according to the number of incumbents, more weight is given to the evaluations that were made with a full level of information available.

Although this method does not explicitly correct for measurement error, if the hypothesis that more information will lead to more reliable ratings is true, then weighted least squares will indirectly adjust for measurement error. Even if the hypothesis is not true, the results will be interesting from the standpoint of reducing the impact of the "unique" jobs.

Table 16 presents the results of the weighted least squares regression. For comparison purposes, the relative weights of the statistical model based on all jobs is also presented. The corresponding t-statistics are in parentheses.

Table 16. Weighted Least Squares

Factor	Relative weights Statistical weights/ all jobs	Weighted Least Squares
1. Knowledge-education	25.2% (14.8)	23.1% (13.2)
2. Knowledge-experience	15.1% (10.4)	14.7% (9.6)
3. Job Complexity	20.0% (7.9)	25.1% (9.0)
4. Guidelines-supervision	9.2% (4.2)	9.3% (4.1)
5. Personal contacts	7.6% (2.8)	5.9% (2.6)
6. Physical demands	-1.3% (1.2)	-1.19% (1.0)
7. Mental demands	-.3% (1.6)	-.29% (.9)
8. Supervision exercised	1.9% (1.0)	2.8% (1.3)
9. Scope and effect	10.1% (4.7)	8.8% (3.8)
10. Impact of errors	12.2% (6.5)	10.9% (5.6)
11. Working environment	-.7% (1.5)	-1.3% (2.4)
12. Unavoidable hazards	-.2% (0.8)	-.4% (1.3)
13. Work pace/interruptions	-.3% (-.4)	-.1% (.3)
R2	92%	90%

The results in Table 16 are very consistent with the earlier weights except for a couple of evaluation factors. The weight given to the evaluation factor job complexity rose from 20% to 25%. Job complexity is now the largest single determinant of job pay. On the other hand, the weights for personal contacts, impact of errors, and scope and effect decreased slightly under the weighted least squares regression. Examining the t-statistics, not much change is in evidence. It is still true that only a handful of evaluation factors significantly determine current pay.

What does this information tell us? When job classifications are weighted according to the amount of information the evaluators will have the impact of job complexity rises to the forefront as the most important determinant. By itself, it is difficult to determine whether the weighted least squares regression is able to compensate for measurement error by giving more weight to those classifications that are well represented. Comparison of the results obtained here with the results of the following sections may provide more insight.

Coefficients Adjusted for Measurement Error

Now that a thorough examination of Arthur Young's methods and results has been completed, the analysis can proceed to the problem of measurement error. In the third chapter a brief introduction and discussion of measurement error was provided. Basically, the use of ordinary least squares as a regression technique has to be questioned when the independent variables are not measured precisely. Classical linear regression requires the assumption that the independent variables

are measured without error. Since this assumption is clearly not met when evaluators rate job classifications, the resulting evaluation factor coefficients are biased. The use of biased coefficients to predict job classification pay under comparable worth is also questionable. Comparable worth advocates argue that a job evaluation system is an appropriate method of determining the value or worth of the job. Given the fact that the evaluation factor coefficients are biased due to measurement error, the ability of the job evaluation system to measure job worth is quite suspect.

Yet, statistical programs have been developed to adjust the evaluation factor coefficients for the amount of measurement error that exists in the model. Based on the amount of information that can be obtained or assumed concerning the measurement error, the appropriate model (EVCARP1, EVCARP2, EVCARP3) can be specified and analyzed.

The appropriate model for the Arthur Young job evaluation system is EVCARP2. For this model it is assumed that the reliability coefficients of the independent variable are known or estimated. The reliability of measure is an indication of the amount of measurement error that can exist. Since the Arthur Young consultants provided the reliability coefficients of the evaluation factors, the proper procedure is to use this additional information to estimate coefficients that are no longer biased. It should be noted that the reliabilities reported by the Arthur Young consultants may overstate the actual reliability. This is because the consultants let the evaluation teams examine previously conducted evaluations as the process was occurring. For example, during the second week of the evaluation process, the evaluation teams got to

examine all of the evaluations conducted in the first week. The effect of this procedure should be to raise the reliability of subsequent evaluations. Since the final reported reliabilities are the only data available, they will be utilized despite the possibility that they may overstate the true reliabilities.

In the first regression, the current pay was regressed on the evaluation factor ratings for the classifications that were dominated by females. The results obtained from this model would provide an indication of which evaluation factors were most important in determining the pay of female-dominated classifications. Table 17 presents the original evaluation factor weights and the coefficients that were estimated with EVCARP2. The corresponding t-statistics are presented in parentheses.

Table 17. EVCARP coefficients for female classifications

Evaluation factor	Statistical weights	Committee weights	EVCARP weights
1. Knowledge-education	28.8% (6.0)	27.2%	29.4 (2.2)
2 Knowledge-experience	8.2% (3.0)	11.9%	11.1% (1.3)
3. Job complexity	15.8% (2.8)	12.2%	44.5% (2.4)
4. Guidelines-supervision	11.6% (2.4)	8.0%	9.8% (0.9)
5. Personal contacts	12.3% (2.4)	12.9%	3.1% (0.3)
6. Physical demands	-.4% (0.2)	-2.0%	-.1% (0.01)

Table 17. continued

Evaluation factor	Statistical weights	Committee weights	EVCARP weights
7. Mental demands	.7% (-0.8)	-1.8%	-4.3% (0.07)
8. Supervision exercised	2.3% (1.0)	6.3%	2.3% (0.1)
9. Scope and effect	15.1% (3.1)	13.3%	2.5% (1.0)
10. Impact of errors	8.9% (2.8)	6.3%	-15.6% (-1.6)
11. Working environment	-.1% (0.9)	.7%	.4% (0.2)
12. Unavoidable hazards	0.0% (-1.2)	.6%	0.0% (0.1)
13. Work pace/interruptions	-3.0% (-1.0)	4.6%	-1.0% (-0.08)
R^2	92%		90%

The results in Table 17 have two interesting findings. Overall, the weights derived using EVCARP are similar to the original weights with a few major exceptions. The evaluation factor weight for job complexity rose from an average of 14 percent to 44 percent. Remember, the weighted least squares regression produced a similar finding. The weighted least squares was run under the belief that job classifications with five or more incumbents might be measured more accurately because there was more information to make the ratings. The weighted least squares regression suggested that job complexity is more important than knowledge from education in determining job pay. Although no

information was available at the time to further investigate the finding, the EVCARP regression results support the weighted least squares results.

The other major difference is with the weighting of the impact of errors factor. Under EVCARP, the weighting has changed from significant and positive to a highly significant negative weight. It is difficult to find an explanation for this result. The reliability coefficient of the factor was .74 which is about average for all the reliabilities taken together. So it does not appear that the negative weight could be caused primarily by a low reliability measure. An examination of the correlation between the impact of errors and job complexity did not reveal any additional information. The correlation coefficient between the two factors was .77. Overall, no obvious explanation can be found. However, the t-statistic for the impact of errors evaluation factor was insignificant. It will be interesting to discover if other models conclude similar results.

A third interesting finding of the EVCARP regression was that the statistical significance of the evaluation factors fell. This result is due to the fact that the intercept variable, which is measured without error, becomes a highly significant determinant of current pay. The independent variables, which are measured with error, fall in statistical significance. This finding has important implications when compared with the original results. When the factor ratings were treated as measured without error, only a handful of factors were statistically significant. Now that the additional information concerning the measurement error is included, even fewer of the factors

significantly determine current pay.

The coefficients for female classifications are presented for informational purposes only. None of the Arthur Young recommended pay plans were based on the weights of female job classifications. If it is believed that females are discriminated against, why base a comparable worth system on the pay of females?

The next model examined was the pay of male job classifications. The original weighting of the job evaluation factors of male classifications was used to predict the effects of an implemented comparable worth plan. The results of this model are presented in Table 18. The derived EVCARP weights and corresponding t-statistics can be compared to the original findings.

Table 18. EVCARP coefficients for male classification

Evaluation factor	Statistical weights	Committee weights	EVCARP weights
1. Knowledge-education	27.0% (12.3)	22.7%	32.1% (4.1)
2. Knowledge-experience	15.9% (8.5)	15.7%	16.3% (2.6)
3. Job complexity	18.7% (5.6)	15.9%	21.4% (2.2)
4. Guidelines-supervision	1.6% (.6)	7.5%	3.1% (.8)
5. Personal contacts	9.4% (2.4)	12.5%	7.4% (1.4)
6. Physical demands	-.2% (.1)	-4.8%	-5.3% (-1.0)
7. Mental demands	0.0% (3.7)	-.2%	0.0% (.6)

Table 18. continued

Evaluation factor	Statistical weights	Committee weights	EVCARP weights
8. Supervision exercised	3.5% (1.4)	8.0%	2.4% (.08)
9. Scope and effect	13.6% (4.9)	15.5%	6.6% (1.9)
10. Impact of errors	9.7% (4.3)	7.5%	3.6% (.2)
11. Working environment	1.1% (.8)	-3.0%	-2.8% (-.1)
12. Unavoidable hazards	-1.0% (1.2)	-1.2%	-.9% (-.09)
13. Work pace/interruptions	.7%	3.9%	-.1%
13. Work pace/interruptions	.7% (1.0)	3.9%	-.1% (-.2)
R^2	92%		90%

The results in Table 18 lend more support to the earlier weighted least squares regression and the analysis on female classifications. Although the weight on job complexity is not the highest, it has shown an increase. As more and more evidence supports this finding, it appears that the original weights may have underestimated the contribution of job complexity. When estimated with a program that adjusts for measurement error, job complexity is emerging as one of the more important factors. Because the pay grades are determined over a range of points, the exact weights are not as important as determining which evaluation factors contribute the most.

More evidence against the importance of the impact of errors and

scope and effect is also found. Although the estimated coefficient on impact of errors is not negative, it has again been given less weight than the original estimated coefficients. Overall, the first two EVCARP models have provided some consistent and interesting results.

The previous findings concerning the statistical significance of the evaluation factors also receives more support. Under the model based on male classifications when the ratings were taken as measured without error, the following factors were statistically significant: knowledge-education, knowledge-experience, job complexity, personal contacts, mental demands, scope and effect, and imipact of errors. Under the EVCARP regression, the following factors were significant: knowledge-education, knowledge-experience, and job complexity. Only these three factors significantly determine current pay for males. As before, the intercept term becomes highly significant.

The third model to be examined was the relative weight of the evaluation factors in determining the pay of all job classifications when an extra variable for the percentage of female incumbents is included. Table 19 presents the relative evaluation factor weights from the original Arthur Young system and the EVCARP regression with the corresponding t-statistics.

Table 19. EVCARP coefficients for all jobs with percent female

Evaluation factor	Statistical weights[a]	EVCARP weights
1. Knowledge-education	24.5% (15.7)	26.7% (6.2)

[a]The Arthur Young consultants did not present the factor weights of this model with committee weights.

Table 19. continued

Evaluation factor	Statistical weights[a]	EVCARP weights
2. Knowledge-experience	12.9% (9.6)	22.7% (5.5)
3. Job complexity	18.3% (7.8)	22.9% (5.0)
4. Guidelines-supervision	6.6% (3.4)	7.9% (2.1)
5. Personal contacts	9.4% (5.8)	14.0% (2.6)
6. Physical demands	.2% (.2)	3.2% (.9)
7. Mental demands	-.6% (-3.1)	-.6% (-1.3)
8. Supervision exercised	2.1% (.8)	-2.5% (.7)
9. Scope and effect	9.9% 95.1)	10.0% (1.6)
10. Impact of errors	10.0% (5.9)	12.1% (2.3)
11. Working environment	0.0% (-.2)	-4.1% (-.40)
12. Unavoidable hazards	-.1% (.3)	1.6% (.1)
13. Work pace/interruptions	.2% (.2)	-3.5% (-.1)
14. Percent female	-6.3% (-12.3)	-3.0% (-8.9)
R_2	93%	90%

The results in Table 19 are another confirmation of earlier findings. Once again, the weight on job complexity has risen. It is

not the most important factor in determining current pay, but this is the fourth instance where job complexity has increased. The other results are similar to what had been derived originally by the Arthur Young consultants.

The biggest change occurred with the knowledge from experience variable. Originally, the relative weight was 12.9 percent. From the EVCARP regression, the relative weight is estimated at 22.7 percent. The factor is still statistically significant.

In terms of statistical significance, the earlier findings are somewhat supported. Originally, nine evaluation factors were statistically significant. The most important determinants were knowledge-education, knowledge-experience, job complexity, impact of errors, personal contacts, and the percent of female incumbents. When the regression was run under EVCARP, the intercept term and the percent of female incumbents remained highly significant. Other factors such as knowledge-education, knowledge-experience, and job complexity are still significant determinants of current pay.

Prior to this analysis, the results had shown a tendency for the original regressions to undervalue the contribution of job complexity and overvalue the contribution of scope and effect and impact of errors. This model can only slightly reinforce those findings. Although job complexity increased again, the weights on scope and effect and impact of errors were relatively stable.

Despite being the only variable that was measured without error, the weight on percent of female incumbents did not change dramatically. It is still negative and significant which implies that the sexual

composition of a classification is still important in determining how much the classification is paid. The fact that the weight is negative implies that the more a classification is made up of females, the less it is paid. However, the value of the coefficient did fall by one-half. This is a significant finding that would suggest that the earlier findings may have over-emphasized the impact of the male/female composition. These two results together would be used by comparable worth advocates as evidence of discrimination.

To summarize, what can be said about the impact of measurement error? First, the measurement error problem did not significantly change the evaluation factor weights. There was evidence that the job complexity factor had been previously undervalued. Evidence also existed that the impact of errors and scope and effect may have originally been overvalued. The results in this area are not nearly as strong. Although the relative weights are not significantly changed, the statistical significance of several factors was changed. Under the original regressions, about one-half of the evaluation factors significantly determined pay grade. These evaluation factors included knowledge-education, knowledge-experience, job complexity, guidelines-supervision, personal contacts, scope and effect, impact of errors, and the percent of female incumbents. The remaining evaluation factors do not significantly explain differences in pay. However, the number of significant evaluation factors under the EVCARP regressions is definitely less. Only factors such as knowledge-education, knowledge-experience, job complexity, and percent of female incumbents remained significant when the measurement error problem was corrected.

Therefore, it is obvious that the problem of measurement error cannot be ignored. Any opportunity to derive factor evaluation weights that utilize any available information should be used. EVCARP and other statistical packages that take into account information known or estimated about the quality of the input data are tools that the comparable worth advocates should use in the future. If they choose not to, then criticism concerning reliability, consistency, validity, and measurement error will impede their progress.

On the other hand, if the statistical regressions are conducted to not only determine which factors contribute the most, but exactly how much each factor contributes, then measurement error is a much more serious problem. The analyses conducted in this section have shown that even with measurement error, a handful of evaluation factors contribute the most in explaining current pay. However, the specific weights on these factors were not consistent. Therefore, a regression designed to determine the exact "optimal" weight of each factor could be seriously flawed if the problem of measurement error is ignored.

The next section demonstrates how serious of a problem measurement error can be.

Average Evaluation Rating by Sex

In the next section a set of derived EVCARP weights will be used to examine the impact of implementing the EVCARP model. To clearly understand how changes in the evaluation factor weights affect the pay grade changes for male and female classifications, it is important to know how male and female classifications were rated. Table 20 presents

the average evaluation rating for the male and female classifications.

From Table 20, several observations can be made. First, male jobs classifications received higher average ratings for every job evaluation

Table 20. Average evaluation rating

Evaluation Factor	Average rating Male	Female
1. Knowledge-education	4.74	3.95
2. Knowledge-experience	4.02	3.12
3. Job complexity	3.65	2.75
4. Guidelines-supervision	2.95	2.28
5. Personal contacts-purpose	2.61	2.19
6. Personal contacts-type	3.21	3.30
7. Physical demands	1.83	1.73
8. Mental demands	2.13	2.29
9. Supervision exercised-nature	2.19	1.63
10. Sueprvision exercised-number	2.29	1.78
11. Scope and effect	2.72	2.11
12. Impact of errors	3.04	2.27
13. Working environment	1.88	1.61
14. Hazards and Risks	1.74	1.45
15. Working conditions-Pace	1.98	1.78
16. Working conditions-interruptions	2.10	1.94

factor except mental/visual demands and personal contacts-type. The evaluation factors that male classifications held the greatestadvantages were knowledge-education, knowledge-experience, job complexity, guidelines-supervision, scope and effect, and impact of errors. Therefore, any changes in relative weights that give more emphasis to those factors would be expected to benefit male classifications.

In the next section the unbiased coefficients derived with EVCARP will be used to recommend a set of pay grade changes.

Unbiased Coefficients and Pay Grade Recommendations

By using statistical packages such as EVCARP, unbiased coefficients can be derived that are adjusted for the additional information on the quality of measure for the independent variables. It would then be interesting to examine what pay grade recommendations would be made utilizing the unbiased coefficients. The choice of the model to demonstrate is between the model based on the pay of males or the model based on the pay of all classifications plus an additional variable for percent of female incumbents. Both models have been forwarded by comparable worth advocates as appropriate techniques to obtain "unbiased" estimators. In this setting "unbiased" refers to estimators that are free of sexual bias.

Because the percent of female incumbents model is frequently used in other job evaluation studies, it will be the model of choice to examine the effect of using the unbiased evaluation factor weights determined by EVCARP.

EVCARP Coefficients and the Pay Recommendations

How will the use of the unbiased coefficients affect the pay recommendations for male and female classifications? The original six recommendations all called for substantial increases in the pay of female classifications while also calling for net decreases in the pay of male classifications. Table 21 presents the pay recommendations under the EVCARP weights for female and male classifications and the original recommendations of the statistical percent female model.

Table 21. EVCARP pay recommendations

Pay recommendations		Pay grade changes Male	Female
Original weights from statistical percent female model	Increase	23.1%	71.7%
	Decrease	55.2%	16.1%
EVCARP weights based on percent female model	Increase	34.2%	65.3%
	Decrease	43.3%	21.9%

It certainly appears that the unbiased EVCARP weights tempered the increases for females slightly and also eased the reductions for males. It is not surprising that the recommendations did not change significantly since the weights derived under the EVCARP model were very similar with a couple of exceptions. The fact that males improved their position under the EVCARP weights is probably due to the increases in the weights on the evaluation factors knowledge from experience and physical demands. Male classifications held an advantage in both of those categories.

EVCARP Coefficients and Variation

In the previous section, a lot of time was spent investigating differences in variation. The average range of the recommendations was deemed important because of the issues of reliability, consistency, and validity. Although different regressions will certainly yield different coefficients, it would be ideal if the recommendations based on the derived coefficients did not suffer from a high average range. When the goal of job evaluation is to provide an indication of the worth of a job, recommended job values that vary significantly will be viewed as unreliable or invalid. Therefore, any method or technique that can

produce unbiased results and lower the range of the recommendations would be of great advantage to comparable worth advocates.

From the earlier examinations, it was found that female classifications had a slightly higher range of recommendations than male classifications. Table 22 presents the average range of the recommendations with EVCARP results utilizing the percent female model included with the original recommendations.

Table 22. EVCARP recommendations included in the average range

Pay recommendations	Male	Female
Original recommendations	2.42 pay grades	2.84 pay grades
Original recommendations plus EVCARP recommendations	2.46 pay grades	2.89 pay grades

The results of Table 22 show that when the EVCARP recommendations were also included with the original six recommendations, the average range of the recommendations rose slightly. This implies that most of the EVCARP recommendations are within the range of recommendations already determined by the six original recommendations. Therefore, inclusion of the EVCARP model with the previous models has not damaged the reliability or validity of the earlier findings.

EVCARP and Cost to the State

One of the more interesting findings earlier in this section was the discovery that comparable worth implementation will cause the wage bill of the organization to rise. Four of the six Arthur Young pay recommendations called for increases of over 20,000 pay grades. How

does the EVCARP recommendations compare to the earlier findings? The cost of the system is an important facet that must be taken into account. If the best comparable worth system ever devised is not politically or economically acceptable on a cost basis, the system doesn't have much of a chance of being implemented. Table 23 presents the original cost data and the cost of the EVCARP recommendations.

Table 23. Cost of Implementation

Pay recommendations	Total pay grade changes
[COMMITTEE WEIGHTS/MALE JOBS]	43,236 pay grades
[COMMITTEE WEIGHTS/PERCENT FEMALE]	31,059 pay grades
[COMMITTEE WEIGHTS/ALL JOBS]	25,150 pay grades
[STATISTICAL WEIGHTS/MALE JOBS]	24,302 pay grades
[EVCARP WEIGHTS/PERCENT FEMALE]	13,287 pay grades
[STATISTICAL WEIGHTS/PERCENT FEMALE]	10,109 pay grades
[STATISTICAL WEIGHTS/ALL JOBS]	5,854 pay grades

The results in Table 23 demonstrate that the two statistical models based on all jobs with percent female (EVCARP and original Arthur Young) would not require the significant increases in pay grades that the models based on committee weights or the pay of male classifications required. The EVCARP model is slightly more expensive than the original Arthur Young recommendations based on the same model. This is due to the fact that the EVCARP recommendations were slightly more beneficial to male classifications, and male classifications outnumber female classifications. Overall, the EVCARP recommendations appear to be as economically and politically acceptable as the original recommendations.

EVCARP and the Level of Information

The issue of different information levels was another topic that has been investigated a great deal in this thesis. Are there two distinct levels of information that can significantly affect the outcome of job evaluations? I still believe that this is one question that requires further research. As for the EVCARP recommendations, Table 24 presents the pay grade recommendations by number of incumbents.

The results in Table 24 show that the EVCARP recommendations also were more beneficial to the classifications with five or more incumbents. Again, this finding is contrary to prior expectations. Before the analysis, the impact of the "unique" jobs would have been

Table 24. EVCARP recommendations by number of incumbents

Pay recommendation		Under five incumbents	Five or more incumbents
Original recommendations based on percent female	Increase	29.1%	43.8%
	Decrease	51.0%	33.9%
EVCARP recommendations based on percent female	Increase	34.8%	48.2%
	Decrease	44.7%	30.3%

hypothesized to lead to larger increases for classifications with under five incumbents. In comparing the original recommendations with the EVCARP recommendations, the changes under EVCARP were slightly more beneficial than the original recommendations.

EVCARP Recommendations and Income Distribution

Early indications seem to imply that the unbiased coefficients generated by EVCARP have not led to major differences in the earlier

findings. As for the effect of the pay recommendations on the income distribution, the earlier results contradicted statements made by Pendleton and Horowitz who believed that middle-class white women would benefit the most. The Arthur Young pay recommendations would have benefited the lower-class jobs the most. Table 25 presents the EVCARP recommendations and the effect on the income distribution.

Table 25. EVCARP recommendations by income level

Pay recommendation		Upper level	Middle level	Lower level
Original recommendations based on percent female	Increase	17.9%	47.6%	60.4%
	Decrease	57.8%	36.8%	25.8%
EVCARP recommendations based on percent female	Increase	26.4%	39.1%	67.4%
	Decrease	48.0%	40.1%	17.2%

The EVCARP recommendations are consistent with the original findings that low-income classifications would benefit the most from the implementation of a comparable worth plan. However, EVCARP shifts the degree of benefit to the group a little more than the original recommendations. This is probably due to the fact that the EVCARP recommendations are slightly more beneficial to male classifications than some of the other pay recommendations.

Another method of determining the break down of income distribution was to examine the pay recommendations according to the management level. The early indications supported the finding that lower-income classifications would benefit the most from the implementation of a comparable worth plan. Table 26 presents the EVCARP pay recommendations and the original results for the two management levels: supervisory and

nonsupervisory.

Table 26. EVCARP recommendations by management level

Pay recommendation		Supervisors	Nonsupervisors
Original recommendations based on percent female	Increase	32.0%	41.5%
	Decrease	45.0%	38.1%
EVCARP recommendations based on percent female	Increase	39.4%	42.7%
	Decrease	41.1%	40. %

From the results presented in Table 26, it is obvious that definite support for the earlier findings exists. First, the EVCARP recommendations that were based on the unbiased evaluation factor coefficients have supported the earlier findings in every case. The inclusion of the additional information on measurement error changed some of the evaluation factor weights, but the overall impact on the subsequent recommendations is slight. Second, the recommendations based on the EVCARP model have shown a tendency to lessen the impact of some of the original results. In most cases presented, the benefits of females and low-income classifications have been reduced slightly in favor of male and upper-income classifications.

EVCARP Recommendations and Pay Plan

In the earlier analysis of the six recommended plans and their effect on different bargaining units, it was found that the clerical and blue collar bargaining units would benefit the most if a comparable worth plan was implemented. Fiscal and staff and professionals were targeted for overall pay grade decreases. Based on the results to this point, it would be expected that the EVCARP results would support the

earlier findings. Table 27 presents the results of analyzing the EVCARP recommendations by bargaining unit.

Table 27. EVCARP recommendations and bargaining unit

Bargaining unit	Original recommendation based on percent female		EVCARP recommendation based on percent female
Clerical	Increase	91.5%	87.8%
	Decrease	4.3%	6.7%
Technical	Increase	46.2%	47.4%
	Decrease	34.9%	38.0%
Blue collar	Increase	37.8%	50.4%
	Decrease	43.2%	28.3%
Fiscal and staff	Increase	24.4%	26.2%
	Decrease	48.8%	49.6%
Human services	Increase	44.0%	46.0%
	Decrease	32.7%	30.0%
Security	Increase	27.8%	36.3%
	Decrease	46.3%	39.2%
Public safety	Increase	0.0%	25.0%
	Decrease	50.0%	50.0%
Professionals	Increase	28.4%	32.5%
	Decrease	47.5%	44.8%
Not designated	Increase	36.4%	38.3%
	Decrease	63.6%	50.9%

The information presented in Table 27 supports the earlier results as expected. These results again emphasize the increasing role that unions may play in the comparable worth movement. The bargaining unit that will benefit the most, clerical, is the bargaining unit that is characterized by strong union influence. The fact that the professionals and the fiscal and staff units would receive the most

downgrading combined with the fact that those two bargaining units are not highly unionized is also worth repeating.

Also, the shift towards male classifications is again evident. The clerical bargaining unit, dominated by female classifications, does not benefit as much under EVCARP as it did under the original recommmendations. On the other hand, male-dominated classifications such as the professionals and blue-collar will benefit more under the EVCARP recommendations.

CHAPTER V. SUMMARY AND CONCLUSIONS

Throughout this thesis, a great deal of information has been presented concerning the impacts of the original Arthur Young recommendations on various segments of the State of Iowa's work force and the effect that measurement error could have had in biasing the original estimated evaluation factor coefficients. In the previous statistical section various analyses were conducted to examine the predicted impact of a comparable worth system beyond the simple question of male versus female. Then, the original results were modified with a statistical package designed to determine unbiased estimators when a problem of measurement error exists. What conclusions can be drawn from the results that were presented?

From the introduction and literature review, the following statements summarize the background of comparable worth and the need for this thesis.

1. Historically, the average earnings of women have consistently been only a fraction of what men earn. The earnings ratio has been estimated at approximately 50 to 70 percent.
2. The low earnings ratio combined with the inability of human capital differences to explain the wage gap have been two primary motivators of the modern comparable worth movement.
3. Comparable worth advocates argue that job evaluation systems can accurately and consistently measure the value of a job to its organization.
4. Critics point out that the job evaluation process is highly subjective. They argue that problems with reliability,

consistency, and validity discount the job evaluation's ability to measure job worth.

5. Research into the aspects of measurement error and its impact on the job evaluation process have been called for, but not done at the time of this thesis.

Chapter II of the thesis explored the methods and results that the Arthur Young consultants originally reported to the State of Iowa following their study concerning the implementation of a comparable worth plan. Overall, the consultants seemed to have developed and conducted a thorough and objective examination into the procedures needed to implement a large comparable worth compensation system.

The problem of measurement error was discussed in Chapter III. The major points brought out are noted in the following statements.

1. Measurement error occurs when the independent variables are not known with certainty.
2. The presence of measurement error will lead to biased coefficients in a statistical analysis.
3. Even when a statistical analysis is not conducted, the problem of measurement error cannot be ignored.
4. Measurement error can be a significant problem in social science studies. When the independent variable is determined through the subjective judgement of an individual(s), the observed value of the variable may significantly differ from the true value.
5. Although it is impossible to know the exact amount of measurement error made, the reliability of the measurement for

the independent variable gives an indication of the quality of the data.

6. Statistical packages have been developed that utilize the information on the quality of the input data and determine unbiased coefficient estimates. One such program is EVCARP, an errors-in-variables regression program developed in the statistical laboratory at Iowa State University.

From the statistical analyses that were conducted on the original Arthur Young recommendations in Chapter IV, the following findings can be concluded.

1. It is not surprising that the comparable worth recommendations are most beneficial to female classifications. Overall, an average of 71 percent of female classifications were to be upgraded. The impact of comparable worth on male classifications is quite the opposite. On the average, 48 percent of the male classifications are overvalued according to the comparable worth results.
2. On the average, the female classification recommendations were more variable then the male classifications.
3. More evidence was found supporting the contention that comparable worth plans would prove very costly to implement. Four of the six Arthur Young recommendations would require an increase of over 20,000 pay grades to implement. This finding coupled with the often used practice of implementing comparable worth without any downgrading translates into significantly high payroll costs. The economic implications of the increase

in labor costs is still being debated.

4. The question of the existence of different levels of information was never fully settled. The results based on the variation of pay recommendations supported the hypothesis that jobs with under five incumbents are not measured as accurately. However, based on pay recommendations, the "unique" classifications did not benefit to a greater extent.
5. Another criticism of comparable worth was dispelled with an investigation into the impact of the pay recommendations on the distribution of income for state employees. Critics of comparable worth have stated that the issue was primarily a vehicle of middle-class white women. An examination of the original Arthur Young recommendations did not support the statements. It was discovered that the lower-income classifications would reap the most benefit of comparable worth implementation. This result was independent of the methods used to break-down the classifications into income levels.
6. The analysis of the pay recommendations and their impact on different bargaining units also revealed some interesting findings. It was discovered that the bargaining units that stood to gain the most from comparable worth were the most organized. The bargaining units that would bear the most decreases in pay were generally characterized as not highly unionized. It is apparent that a level of causality between union strength and comparable worth benefits may exist. However, the direction of causality is not known. Do these

bargaining units benefit because of union strength or are unions realizing the advantages of organizing those sectors that will benefit from comparable worth? I believe that it is obvious that along with governments and the courts, unions are going to be highly involved in the future of comparable worth.

7. The use of an arbitrary 70 percent rule as the criterion of a dominated classification does not appear to be an accurate measure of "dominance." In a state that has roughly two-thirds of its work force male, it was shown that 57 percent of the classifications would be male-dominated if males and females were randomly assigned to classifications. Along the same lines, only two percent of the classifications would be female-dominated. With such a large disparity of odds, using the 70 percent rule does not seem accurate. However, an alternative method did not significantly change the original findings. It was also pointed out that the sex of the classification may not have an impact on the results at all.

8. To further examine the impact of different levels of information, a weighted least squares regression was conducted that gave more weight to those classifications with more than five incumbents. The results gave the first indication that job complexity may have been undervalued by the original models. Of all the issues examined, I believe that the question of different levels of information is an issue that needs to be further researched.

After the original recommendations were fully investigated, EVCARP

was utilized to determine an alternative set of estimated coefficients. With these unbiased coefficients and the percent female model, a new set of pay grade recommendations were developed. The major findings of the impact of measurement error are summarized below.

1. The new estimated coefficients did not significantly alter the original results. There was some further evidence that suggested that job complexity may have been a more important evaluation factor than earlier studies suggested. Although the evidence was not as strong, some results showed the possibility that scope and effect and impact of errors were overvalued. Overall, the possibility that the job complexity factor was underrated by the Arthur Young approach appeared quite likely.
2. However, even though the original results were not significantly affected, ignoring measurement error is not a desirable alternative. Even if the goal of the statistical analyses is simply to determine which factors are most important, the critics of comparable worth will continue to discount the role of job evaluation as long as questions of reliability and validity are not fully addressed. As long as it is cost efficient to perform the more detailed EVCARP statistical analyses, measurement error should not be ignored.
3. It was found that the statistical significance of the evaluation factors changed under the EVCARP regression. Previously, only one-half of the factors significantly determined current pay. When the fact that the evaluation ratings were measured with error was included, the significance

of the ratings in explaining differences in pay fell. Only knowledge-education, knowledge-experience, and job complexity remained statistically significant. When the independent variables are measured with error, the intercept term and the percent of female incumbents become highly significant because they are not subject to measurement error.

4. Of course, if the goal of the statistical analyses is to determine the exact contribution of each evaluation factor, the effect of measurement error can certainly not be ignored. The EVCARP program will yield different results that are based on the quality of the input data.
5. The use of the new coefficients to make subsequent pay grade recommendations did not alter any of the previous findings. There was a slight shift towards increased benefits for male classifications. This was probably due to the increased weight on the job complexity factor. Male classifications held an advantage in the average rating for the job complexity factor.

Overall, this thesis addressed the impact of some serious criticisms of job evaluation systems and comparable worth. It was interesting to use a relatively new statistical approach to examine an issue that will certainly continue to grow in importance in the future. Comparable worth is a political, economic, and moral issue that is being debated more and more each day. Once statistical packages such as EVCARP are fully developed and available, their impact on current statistical methods should also increase in the future.

BIBLIOGRAPHY

Aaron, H. J., and Lougy, C. M. The Comparable Worth Controversy. Washington, D.C.: The Brookings Institution, 1986.

Aldrich, M., and Buchele, R. The Economics of Comparable Worth. Cambridge: Ballinger Publishing Company, 1986.

Asher, Herbert B. Casual Modeling. 2nd edition. Beverly Hills: Sage Publications, 1983.

Arthur Young. Study to Establish an Evaluation System for State of Iowa Merit Employment System Classifications on the Basis of Comparable Worth: Final Report. Milwaukee, Wisconsin: Arthur Young Company, April 1984a.

Arthur Young. Study to Establish an Evaluation System for State of Iowa Merit Employment System Classifications on the Basis of Comparable Worth: Statistical Supplement. Milwaukee, Wisconsin: Arthur Young Company, April, 1984b.

Bailey, C. T. The Measurement of Job Performance. Aldershot, England: Gower Publishing Company Limited, 1983.

Barrett, R. S. Performace Rating. Chicago: Science Research Associates, Inc., 1966.

Bartley, D. L. Job Evaluation: Wage and Salary Administration. Reading, Massachusetts: Addison-Wesley Publishing Co., 1981.

Berk, R. A.(ed.). Performance Assessment. Baltimore: John Hopkins University Press, 1986.

Blalock, H. M. Basic Dilemmas in the Social Sciences. Beverly Hills: Sage Publications, 1984.

Blau, F. D. Equal Pay in the Office. Toronto: Lexington Books, 1977.

Bureau of National Affairs. VDTS in the workplace: A study of the effects on employment. A Special Report. Washington D. C., 1984.

Cook, A. H. Comparable Worth: The Problem and States' Approaches to Wage Equity. Manoa: Industrial Relations Center, University of Hawaii, 1983.

Dunnette, M. D., and Fleishman, E. A. (eds.). Human Performance and Productivity: Human Capital Assessment. Volume I. Hillsdale, New Jersey: World Publishing Company, 1982.

Elizur, Dov. Job Evaluation: A Systematic Approach. Westmead, England: Gower Publishing Company Limited, 1980.

Farr, J. L., and Landy, F. J. The Measurement of Work Performance. New York: Academic Press, 1983.

Ghiselli, E. E., Campbell, J. P., and Zedeck, S. Measurement Theory for the Behavioral Sciences. San Francisco: W. H. Freeman and Company, 1981.

Hartmann, H. I. (ed.). Comparable Worth: New Directions for Research. Washington, D.C.: National Academy Press, 1985.

Hartmann, H. I., and Treiman, D. J. Women, Work, and Wages: Equal Pay for Jobs of Equal Value. Washington, D.C.: National Academy Press, 1981.

Henderson, R. I., and Clarke, K. L. Job Pay for Job Worth. Atlanta: Business Publishing Division, College of Business Administration, Georgia State University, 1981.

Hidiroglou, M. A., Fuller, W. A., and Hickman, R. D. SUPER CARP. Ames, Iowa: Statistical Department, Iowa State University, 1980.

Hutner, F. C. Equal Pay for Comparable Worth. New York: Praeger Publishers, 1986.

Jaech, J. L. Statistical Analysis of Measurement Errors. New York: John Wiley and Sons, 1985.

Johansen, Elaine. Comparable Worth: The Myth and the Movement. Boulder: Westview Press, 1984.

Kessler, L. L., and Williams, R. E. A Closer Look at Comparable Worth. Washington, D.C.: National Foundation for the Study of Equal Employment Policy. 1984.

Lanham, E. Job Evaluation. New York: McGraw-Hill Book Company, Inc., 1955.

Latham, G. P., and Wexley, K. N. Increasing Productivity Through Performance Appraisal. Reading, Massachusetts: Addison-Wesley Publishing Company, 1982.

Leukart, R. H., and Otis, J. L. Job Evaluation: A Basis for Sound Wage Administration. New York: Prentice-Hall, Inc., 1948.

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

Livy, Bryan. Job Evaluation: A Critical Review. London: George Allen and Unwin Limited, 1975.

Lopez, F. M. Evaluating Employee Performance. Chicago: Public Personnel Association, 1968.

Lytle, C. W. Job Evaluation Methods. New York: The Ronald Press Company, 1946.

Miller, R. L., and North, D. C. The Economics of Public Issues. New York: Harper and Row, Publishers, 1987.

Paterson, T. T. Job Evaluations. London: Business Books Limited, 1972.

Patton, J. A., Littlefield, C. L., and Self, S. A. Job Evaluation: Text and Cases. Homewood, Illinois: Richard D. Irwin, Inc., 1964.

Perrin, S. M. Comparable Worth and Public Policy: The Case of Pennsylvania. Philadelphia: Industrial Research Unit, University of Pennsylvania, 1985.

Remick, Helen (ed.). Comparable Worth and Wage Discrimination: Technical Possibilities and Political Realities. Philadelphia: Temple University Press, 1984.

Schlafly, Phyllis (ed.). Equal Pay for Unequal Work. Washington, D.C.: Eagle Forum Education and Legal Defense Fund, 1984.

Schwab, D. P., and Cummings, L. L. Performance in Organizations: Determinants and Appraisal. Glenview, Illinois: Scott, Foresman and Company, 1973.

Topping, J. Errors in Observation and Their Treatment. London: Chapman and Hall, 1972.

Treiman, D. J. "Job Evaluation: An Analytic Review." Paper prepared for the Committee on Occupational Classification and Analysis Assembly of Behavioral and Social Sciences National Research Council. Washington, D.C.: National Academy of Sciences, 1979.

U.S. Commission on Civil Rights. Comparable Worth: Issue for the '80's. A Consultation of the U.S. Commission on Civil Rights, June 6-7, 1984. Washington, D.C.: U.S. Commission on Civil Rights, 1984.

Velikanov, M. A. Measurement Errors and Empirical Relations. Jerusalem: Israel Program for Scientific Translations, 1965.

Willborn, S. L. A Comparable Worth Primer. Toronto: Lexington Books, 1986.

Woolsey, F. E. "An Analysis of the Arthur Young Job Evaluation Study and Its Effect on Comparable Worth Implementation in the State of Iowa." Unpublished Creative Component, Department of Industrial Relations, Iowa State University, 1983.