The Impact of Relative Position on Attributes of an Estimated Performance Distribution

Submitted in partial fulfillment of
the requirements for the degree in
Master of Applied Science
(I/O Psychology)

Saint Mary's University
Halifax, Nova Scotia

Approved ____________
Faculty Advisor

Approved ____________
Committee Member

Approved ____________
Committee Member

April 1989

ISBN 0-315-50915-5

Acknowledgements

I would like to thank those persons who provided both assistance and support throughout this endeavour. Some deserve special recognition: Vic Catano, for providing the impetus to begin the study and the support to continue; Grace Pretty, for providing the opportunity to complete my study through the timely award of teaching assistantships; Committee members Dr. Russ Summers and Professor Valerie Bolivar; my folks for providing both financial and moral support; and, Mary, who made me believe I could finish the project.

Table of Contents

List of Tables and Figures

Abstract

Decision theoretic equations have been used to estimate the relative utility of personnel programs for years. These equations; however, have been used sparingly in I/O psychology. A number of reasons have been proposed to explain this dearth of the use. Some psychologists believe validity to be situation specific. Others believe the data required by the equations do not satisfy the assumptions of the general linear model, on which the utility equations are based. The most cited reason for the infrequent use; however, is the difficultly in obtaining one of the required parameters, namely the standard deviation of job performance in the utility metric, usually dollars (SDy). Recent studies have developed techniques for estimating this parameter. One such technique is the global estimation method proposed by Schmidt, Hunter, McKensie and Muldrow (1979). The method is based on the assumption that if job performance is normally distributed then estimates of a workers value at the 85th percentile minus that of the value of the work performed by an average employee should yield an estimate of the SDy. The technique depends on whether an individual can provide estimates that adhere to the normal distribution.

The present study examined different attributes of the performance distribution that are produced when individuals rate different positions, viz-a- viz their own. Specifically, two hypotheses were investigated. The first stated that the variance produced across percentile point estimates would increase as relative distance increased. The second predicted a significant departure from normality of the estimated performance distribution as the relative distance increased. Four hundred and forty three undergraduate students from Saint Mary's University in Halifax provided estimates for one of three positions; their own, a university professor, or an MP. The results showed a significant difference in the amount of variability as the distance between the rater and the rated position increased. As well, evidence contrary to the accepted belief that estimated performance distributions conform to the normal curve was found. In addition, two other methods of obtaining SDy values were compared to those produced by the global method. The comparison revealed little similarity in the estimates produced by the different methods These results point to some interesting aspects of utility analysis and measurement in general. First it appears that the best judge of a position may be the person functioning in that position. Second, the assumption of normality is not supported by these data.

Utility analysis is the application of the generic cost-benefits analysis to personnel problems. It has received a resurgence in interest in the past ten years (Catano, 1988). There have been a number of reasons proffered for the resurgence. Mayer (1982) suggested the rising costs associated with personnel programs has fueled the resurgence; management questions concerning the economic and productivity implications of personnel programs have been cited as another reason (Schmidt, Hunter, McKenzie, & Muldrow, 1979). Additionally, Bobko, Karren, and Parkington (1983) believe the impetus for the revival in the interest in utility analysis may be due, in part, to the decline in U.S. economic growth. The most cited reason, however, for the renewed attention to this human resource tool are the recent advances that have taken place in estimating the parameters required for these decision theoretic equations.

The generic cost-benefits analysis models are simply techniques for examining the relationship between results of personnel selection systems and their associated costs. Utility analysis, a specialized application of these analyses, allows the personnel specialist to evaluate alternative courses of action concerning personnel decisions (i.e., selection, placement, training). Alexander and Barrick (1987) state " the use of utility analysis to assess the potential dollar gains to be realized by an organization's personnel interventions is becoming an increasingly acceptable and reliable tool to the personnel psychologist" (p. 477). Blum and Naylor (1968) define the utility of a selection device as " the degree to which its use improves the quality of the individuals selected beyond what would have occurred had the device not been used." That is, the relative gain in performance of selectees chosen with a particular device can be compared to that which would have occurred if an alternative device had been used or if the selection had been carried out on a random basis. Utility models in use today can vary along two dimensions (Catano, 1988). The first concerns the manner in which the performance of candidates is quantified. Basically the performance can be viewed as a dichotomous variable or as a continuous variable. When viewed as a dichotomy, performance is considered to be either satisfactory or unsatisfactory. The alternative view, performance as a continuous variable, focuses on predicting performance over a predetermined range. The second dimension, upon which utility models can vary, concerns the presentation of the output information. In this case, the gain (or loss) in utility can be presented in dollar figures or in some other performance measure, such as percent increase in productivity. This is a particularly important aspect of utility analysis because as Cascio (1982) points out "like it or not the language of management is dollars not validity coefficients."

The present study will examine procedures used to estimate the parameters required for utility analysis. Specifically, it looks at some of the psychometric properties of one particular parameter, the standard deviation of job performance in dollars, SDy. There are two main foci of the study, one deals with the shape of the estimated performance distribution and the other deals with the variance of the SDy estimates. The study also examines how the relative distance between the rater and the rated position affects both the distribution and the variance of these estimates.

The Development of Utility Analysis

Utility analysis, "the evaluation of the benefit obtained from selection devices," has enjoyed a long history in the area of I/O psychology (Schmidt, et al., 1979). The majority of these models have focused on the use of the validity coefficient in attempts to assess utility.

The earliest of these attempts to use decision theoretic equations, as utility equations are known, was the "Index of Forecasting Efficiency" (Hull, 1928; Kelly, 1923). This index, described by the following equation:

$$E=1-\sqrt{(1-r_{xy}{}^2)}$$

allowed the comparison between two standard errors. Specifically, the standard error of the estimate or the standard error of job performance scores predicted by the mean test score with the standard deviation of job performance or the standard error that occurs when everyone's performance falls on the mean. This first attempt quickly fell into disfavor due to the pessimistic view it projected for selection tests.

This first index was followed with the "coefficient of determination" in the 1930s and 1940s (Schmidt et al., 1979). The coefficient of determination was nothing more than the square of the validity coefficient. This coefficient was considered to be an indication of the proportion of job performance variance accounted for by the selection test. Although this index certainly provided a more optimistic view of test validities, and is still referred to in some literature, it is now thought that the proportion of job performance variance accounted for by the test has no direct relationship to the utility of a personnel device (Schmidt, et al., 1979).

Both of these early attempts at utility analysis suffered from the same major flaw: neither recognized the fact that the utility of a selection device varies as a function of more than one variable in the selection procedure. Both of these indices only accounted for changes in one variable, namely the validity of the test. Other authors showed this general interpretation of the selection procedure to be inappropriate and that additional variables in the selection environment had to be considered (c.f. Brogden, 1949; Cronbach & Gleser, 1965; Taylor & Russell, 1939).

The first attempt to go beyond the use of the correlation coefficient alone in assessing the efficiency of selection procedures was introduced by Taylor and Russell (1939). This new interpretation took into account, in addition to the validity coefficient, the selection ratio and the base rate. These variables are defined as the proportion of applicants hired and the proportion of applicants who would be successful without using the test in question, respectively. Success, in these cases, was determined by using some organizationally determined job performance criterion. With the introduction of the Taylor-Russell model it became possible to demonstrate that a test of relatively low validity had high utility in certain situations. For example, consider the situation in which a test is used as an aid in selection. The test has a validity coefficient of .15. If the selection ratio is varied between .90 (high) and .10 (low) and the base rate is held constant then a discrepancy is produced in the utility of the test. This effect is outlined in Table 1. At the low selection ratio the utility of the test increases by 10%.

Table 1

Example of Taylor Russell model

Validity Coefficient	Base Rate	Selection Ratio	Utility
.15	.50	.90	.51
.15	.50	.10	.61

The Taylor-Russell model, although an improvement over earlier models, suffered from two major flaws. The first was the way in which job performance had to be viewed when using the model. Specifically, job performance was measured as a dichotomous variable. This meant incumbents could be classified as either successful or unsuccessful, with respect to job performance, without regard for the degree of success on the job. In other words, persons performing a job in a excellent manner were lumped into the same group as those barely meeting the minimum organizational criterion. This method of classifying job performance also made it very difficult, if not impossible, to compare job performance across different job situations. The second major disadvantage of the model was that it provided no basis for establishing the cutoff score which determined the dichotomy. In the model, the utility of a selection device is directly proportional to the cutoff score and the utility of the personnel selection system could thus depend on a possibly arbitrary decision.

As a next step, Brogden (1949) developed a utility model which sought to overcome some of the problems inherent in the Taylor-Russell model. Brogden used the principles of linear regression to illustrate how two variables in the selection situation, the selection ratio and the standard deviation of job performance in dollars (SDy), can affect the utility of a selection device. Brogden's utility model, thought to be a landmark in the development of utility models, is quite straight forward. It involves the derivation of the general linear model to attain the equation that yields the utility increase per selectee of a selection device. Brogden's model is represented arithmetically by the following equation:

$$M_{(y-c)} = r_{xy}\sigma_y M_{z_x} - c/p,$$

where $M_{(y-c)}$ = the gain in utility per selectee over random selection

r_{xy} = the correlation coefficient between the test and the criterion.

σ_y = the standard deviation of job performance in dollars.

M_{z_x} = the mean standardized predictor scores.

c/p = cost of testing per applicant

In addition to the simplicity of Brogden's model, it also supported his assertion that the validity coefficient of a selection device is a direct index of "selection efficiency."

Cronbach and Gleser (1965) contributed the next significant step in the development of the utility model. Their attempt was the first to branch out beyond the idea of fixed-job selection decisions to such areas as placement and classification and sequential selection decisions. Their model closely resembled that of Brogden. The Cronbach and Gleser model is represented by the following equation:

$$\Delta\bar{U} = \Delta U/N_s = r_{xy}SDy\bar{Z}_{x_s} - C_a/\emptyset,$$

where
- ΔU = the gain in utility per selectee over random selection
- r_{xy} = the correlation coefficient between the test and the criterion.
- N_s = the number of applicants selected with the test.
- SDy = the standard deviation of job performance in dollars.
- $\bar{Z}_{x_s}$ = the mean standardized predictor scores.
- C_a = the selection cost per applicant.
- $\emptyset$ = the proportion of applicants above the predictor cutting point (selection ratio)

This version of the utility equation is among the most widely used to assess the utility of personnel selection programmes (Johnson, 1989).

Given the fact that these equations have been available since 1949 one would expect their usage to be widespread in industrial psychological practice and research. This, however, is not the case. Schmidt and colleagues (1979) proffered three explanations for the dearth of research on and the application of these decision theoretic equations. First, they felt that "many psychologists believe that the utility equations are of no value unless the data exactly fit the linear homoscedastic model and all marginal distributions are normal" (Schmidt et al., 1979, p. 615). Therefore, many choose not to employ these models believing their data do not meet these assumptions. Briefly, the linear homoscedastic model contains three assumptions: linearity, equality of variances of conditional distributions, and normality of conditional distributions. Schmidt et al. (1979) address this concern for statistical assumptions and conclude there is no evidence to support the conviction that utility analysis should not be used due to failure to meet the assumptions.

Secondly, Schmidt and his colleagues argue that this lack of application resulted from the belief by some that the validity coefficients are situationally specific. That is, many psychologists hold that the validity of a predictive test is subject to so many variables indigenous to a specific situation that it is meaningless when applied to other similar situations (Schmidt, et al., 1979). Therefore, utility that is based on the validity coefficient would be situation specific. Schmidt, Hunter, Pearlman and Shane (1979) have sought to disprove this proposition through empirical study. In one such study they found that up to 87% of the variance between validity coefficients of different studies could be explained by statistical and measurement artifacts. These artifacts are given in Table 2 It must be pointed out; however, that all psychologists do not agree with the idea of validity generalization as promulgated by Schmidt and his colleagues. Many authors have taken issue with this idea (cf. Schmidt et al., 1986) and raised issues about a number of facets of validity generalization. In particular such ideas as the correction for restriction of range, the use of global criteria; and the disregard for moderator variables in the situation have been questioned. The bottom line is that validity generalization is a contentious issue that has not been resolved.

The third reason offered by Schmidt et al., (1979) and one that has been suggested by others, concerns the difficulty of obtaining the estimates needed for the model. Most often, this concern revolves around one estimate, namely, the standard deviation of job performance in dollars (SDy). The importance of this estimate is readily apparent when the equation is examined: the utility estimate is a direct multiplicative function of the SDy estimate. It follows that an inaccurate estimate of SDy will lead to an inaccurate estimate of the utility of a selection/personnel device.

Table 2. Seven sources of artifactual variance in validity studies as reported by Schmidt, Hunter, Pearlman, and Shane (1979).

1. Differences between studies in criterion reliability.
2. Differences between studies in test reliability.
3. Differences between studies in range restriction.
4. Sampling error (i.e., variance due to $N < \infty$).
5. Difference between studies in amount and kind of criterion contamination and deficiency.
6. Computational and typographical errors
7. Slight differences in the factor structure between tests of a given type.

(Schmidt, Hunter, Pearlman, & Shane, 1979).

It was once thought that SDy could only be obtained through expensive and cumbersome cost-accounting procedures (Catano, 1988). These methods required that the dollar value of an employee be costed out and the standard deviation be computed. The time and effort required for such procedures was tremendous (Roche, 1965). However, in 1979 Schmidt and his colleagues suggested a new, easier-to-use method for obtaining SDy. The method is known as the global estimation model. The authors reasoned that "if job performance in dollars is normally distributed, then the difference between the value to the organization of the products and services produced by the average employee and those produced by an employee at the 85th percentile in performance is equal to SDy."

The authors first used the model in a study of computer programmers. In this study they had supervisors estimate the worth of the employees at three levels: the 15th percentile, the 50th percentile, and the 85th percentile. They then took the mean of the 85th minus the 50th and 50th minus the 15th as the estimate of SDy. In addition, they tested the hypothesis that the distribution of estimates was normal by comparing the upper and lower estimate. They concluded that they could not reject the null hypothesis that the two estimates were equal.

The authors admitted that there were flaws in their model but discounted the effect of these flaws. They reasoned that only errors of a large enough magnitude to produce an incorrect personnel decision were of any concern and that their global estimation method was unlikely to produce such errors.

Since then the global technique has been the subject of at least 17 research studies. On utility the global technique has been used to determine SDy in a wide variety of occupations. For example Burke and Frederick, (1984, 1986); Cascio and Sibley, (1979); Weekly, Frank, O'Connor, and Peters, (1985); Reilly and Smithers, (1985); and Greer and Cascio, (1987) have used the technique in estimating utility for sales positions. Bobko, Karren, and

Parkington, (1983); Desimone, Alexander, and Cronshaw, (1986); Hunter and Schmidt, (1982); Mathieu and Tannenbaum, (1985); and Mayer, 1982 have used it to analyze utility in financial services occupations. Two papers concerned with law enforcement positions used the global model (Karren and Bobko, 1983; Schmidt, Mack, & Hunter, 1984). Tannenbaum and Dickinson, (1987) applied the technique to the nursing profession. The global estimation procedure has also been the subject of many papers dealing with military positions (Eaton, Wing, & Lau, 1985; Eaton, Wing, & Mitchell, 1985; Rossmeissel, 1984) . In many of these studies the global technique and not utility has been the focus. These studies have focussed on factors which influence the validity of the SDy estimate: factors such as accuracy, convergence, sensitivity, cognitive impact and distribution shape as well as others.

Factors Influencing SDy Estimates

Accuracy

Bobko, Karren and Kerkar (1987) state that the accuracy of the utility estimate is directly related to the accuracy of the SDy estimate. This is due to the fact that the utility estimate is a multiplicative function containing the test validity (test in this sense is considered to cover any device used in a selection procedure), the SDy and the mean standardized test scores. Therefore, it is important to examine the accuracy of the SDy estimate in order to better understand utility analysis.

The accuracy of the SDy estimate could be best studied by comparing it to some "true" dollar figure. This "true" figure, however, rarely, if ever, exists in the business world (Bobko et al., 1987). Although it may be argued that cost-accounting procedures might provide the "ultimate criterion" needed for such a study, previous researchers have found, due to the fact that these cost-accounting procedures require human judgement, that these methods do not provide better estimates of SDy (Schmidt et al., 1979). Greer and Cascio (1987) point to the fact that the human resource accounting section of the accounting field has attempted to provide such estimates without much success. They believe this lack of success to be the product of the accounting field's inability to grapple with a "soft concept", namely the value of the human employee.

A related issue to accuracy is the question of uncertainty in the estimates. The authors of the global method assert:

> It is generally not critical that estimates of utility be accurate down to the last dollar. Utility estimates are typically used to make decisions about selection procedures, and for this purpose only errors large enough to lead to incorrect decisions are of any consequence. Such errors may be very infrequent.
> (Schmidt et al., 1979, p. 619).

Bobko, et al (1987) state that more research is needed in this area to determine if this assertion is valid. Irrespective of its validity, however, the need to assess the uncertainty parameter estimates can have on utility estimates is precipitated by the fact that utility analysis attempts to bridge the gap between the Human Resource Professional's validity coefficient and the planning department's bottom line. Alexander, Cronshaw, and

Barrick (1985) sum it up well with the following statement "Financial decision makers often insist that the uncertainty associated with estimates of return to investments be accounted for in capital budgeting analyses." (Alexander, Cronshaw, & Barrick, 1985, p. 4). Given that this statement is true, Alexander et al. (1985) have performed some preliminary work in the area. They utilized the "beta risk" model in assessing the effects uncertainty in the estimates of SDy can have on the overall estimate of utility. The beta risk model, gleaned from managerial finance literature, requires that three estimates of a parameter be obtained; namely, the most likely, the pessimistic, and the optimistic estimate. The extreme values are such that no more than 1% of possible values are more extreme. The authors used this method to re-analyze the Bobko et al. (1983) data. They found the method to provide the user with an option of incorporating the uncertainty of parameter estimates into the model. This option allowed the personnel manager to work within the capital budgeting restrictions imposed on the other departments within an organization.

Convergence

In lieu of an ultimate criterion researchers have compared the global method to objective criteria. Bobko et al. (1983) compared the SDy generated by insurance sales mangers to the dollar value of insurance premiums sold by sales representatives. They found that managers could indeed provide accurate estimates of the dollar value. The problem here, however, is whether the dollar value of insurance sales is a suitable measure of the worth of an individual to the company. It is possible that there may exist some less tangible, equally important, components of worth. This is not a new idea. Katz (1964) alluded to addition components of worth when he wrote of the need for "spontaneous and innovative activity that goes beyond the role prescriptions" in order for an organization to prosper. Cronbach and Gleser (1965) also spoke of the this idea: their statement "... outcomes of hiring includes the man's hourly production, his spoilage, his length of stay, his effect on the morale and tenure of other employees, etc." (p. 22) makes reference to this fact. The most recent study on this subject was undertaken by Orr, Sackett, and Mercer, (1989). These authors found that raters employed these less tangible factors in assigning estimates. Additionally, others have found that the Bobko et al. (1983) study was subject to criterion contamination (Weekly et al., 1985). In the original study the sales managers were kept aware of insurance sales per month for the individual salesmen and therefore had prior knowledge of the criterion used in the study.

The absence of an ultimate criterion has led many researchers to adopt a psychometric method to examine the estimation of SDy. This method, convergence, is used chiefly to establish construct validity. Used in utility the general procedure is to compare multiple methods of estimating SDy to see if they provide the same outcomes; that is, to see if they provide similar estimates. In addition, a variation of this convergence method has been used that employs the same method across raters.

An application of this convergence method was undertaken by Weekly, et al., (1985) who compared three methods of estimating SDy: the Schmidt et al., (1979) global procedure; the CREPID method developed by Cascio (1982); and the Schmidt, Hunter, and Pearlman (1982) 40% method. The Schmidt et al. (1979) method, as previously mentioned, requires an estimate of the value of the performance at the 15th, 50th and 85th percentile and the subsequent computation of SDy. Cascio's (1982) method is based on a job analysis that requires the job be broken down into its constituent factors, assignment of weights to the identified job factors, rating the incumbent population on these factors, and the subsequent computation of SDy. The 40% method, (Schmidt et al., 1982), requires only that 40% of the annual salary be employed as the estimate of SDy. The results of the Weekly et al. study showed that convergence existed between the 40% method and the CREPID method. Estimates produced by these two procedures, however, did not converge with the global method. Bobko and his colleagues offered the following explanation for these results:

> ... the convergence of two of the methods may be due solely to fundamental assumptions about the role of salary in overall worth. That is, the two methods which did converge both use salary as their basis for computation: the Schmidt et al., (1982) procedure by a direct proportion of average salary; the CREPID method by a direct apportionment of mean salary across critical job dimensions. Thus, the convergence of these methods is not surprising and may simply reflect an assumption that overall worth is directly related to salary. In contrast, the Schmidt et al. (1979) allows judges to estimate percentile values, using whatever cues the judges wish ...
>
> (Bobko et al., 1987, p. 8).

Bobko and colleagues conclude their analysis of the Weekly et al. (1985) study by stating the results point to two areas of needed research. Future convergence research must examine an increasingly larger number of different methods of estimating SDy and research must be undertaken which illuminates the cues used by raters in their estimates of SDy (Bobko et al., 1987).

Rater's Position

A factor that can influence the the SDy estimation is the relative distance of the rater from the position to be rated. In utility analysis, as with other personnel tools, the supervisor serves as the source for parameter estimation. The question is, what level of supervisor generates the best estimates of SDy – immediate supervisor, middle management, or top management? Although there is a scarcity of research in this area two studies exist that attempt to answer the above question. Mayer (1982) had employees at different organizational levels in a bank (branch/district manager, head teller, teller) estimate the overall worth of the position of bank teller. These individuals used the global method for generating their estimates. As well, Mayer obtained "accounting estimates" of the SDy. He found significant differences in the SDy estimates across organizational levels. Specifically, he found only the branch/district managers' estimates approximated the accounting estimates. The accuracy of the managers' estimates must be tempered, however, by the fact that the accounting estimates were, in themselves, only imperfect estimates. This study did, however, provide an interesting contrast to that which would be expected from an intuitive standpoint. That is, it might be thought that the closer the rater to the incumbent population the better the estimate. If that were the case then it would be expected that the head teller's estimates would be the closest to the accounting estimates.

Johnson (1989) examined the role of rank and experience in the estimation of SDy in a military situation. He found significant differences in percentile estimate variance across ranks. Specifically, the author found that variance in judgements increased rather than decreased as rank increased, that is, as the relative distance between rater and rated position increased so did variance. Johnson postulated three reasons for this "unexpected finding." First, the chance that judges at different levels may use different processes or populations in making estimates. Second, the impact of the judges' experience in the rated position and the recency of this experience. Third, the degree of interaction of the judges with the rated position.

Inter-rater Reliability

Another area of research needed in the estimation of SDy concerns the judgements made by individual raters. Specifically, inter-rater reliability must be examined further. Previous research studies have found, in some cases, that the interrater variability is greater than or equal to the estimates of SDy (Bobko, et al., 1983; Reilly & Smither, 1985; Schmidt et al., 1979). In the Bobko study the authors found that the raters appeared to be employing different rating scales in making their judgements. In this instance the authors found that raters were using widely discrepant distributions as evidenced by the fact that one rater placed a value of $16,000 on the worth of the 50th percentile performer and another placed the value of $300,000 on this estimate (Bobko et al., 1983). Furthermore, these authors found a significant correlation between the rater's estimate of the 50th percentile and the SDy estimate (r ≈ .70). This correlation leads to a confounding effect on the utility estimate since the distribution of SDy is assumed to be normal and a defining property of a normal distribution is the independence of the mean and standard deviation.

Recent studies have attempted to combat this estimate variability by employing a feedback procedure. These procedures involve providing the mean of the 50th percentile estimates to the raters before they make the additional estimates (Bobko et al., 1987). Two such studies have produced conflicting results, Burke and Frederick (1984) found that the provision of feedback substantially reduced inter-rater variability whereas Wroten (1984) found no such decrease Both studies used the global method for estimating SDy. The most recent attempt to assess the efficacy of feedback in this area was carried out by Tannenbaum and Dickinson (1987) who used a group consensus technique, the Delphi method, along with the global method in an attempt to reduce inter-rater variability. Their results showed an increase in inter-rater reliability and, thus, a decrease in the between rater variability. It must be noted, however, that the study also used a critical incidents methodology to facilitate the judges' percentile estimates. It is, therefore, difficult to pinpoint the exact source of improvement.

Directly linked to the research on inter rater variability is research on the incidence of inconsistent judgements in the estimation of SDy. Two studies, Bobko et al. (1983) and Karren and Bobko (1983), found that 20 % of the research participants, managers, provided inconsistent judgements. They defined inconsistent judgements as those where estimates for the 85th percentile were lower than the accompanying 50th percentile estimate. The authors suggest that the cure may be as simple as informing the raters that their estimate for the 15th percentile should be lower than their estimate for the 50th percentile, and so on. They also believe that raters may be making the individual percentile judgements independently and should be warned to check for inconsistent estimates.

The inconsistency of percentile judgements can be viewed as the manifest portion of a larger problem; that is, do raters understand the meaning of percentiles and do they assume an underlying normal distribution in performance? Two studies have attempted to investigate the rater's understanding of the normal distribution by having the raters provide, in addition to the usual three percentile estimates, an estimate for the 97th percentile. The rationale being that judges who understand the normal distribution will produce estimates that result in equal differences between the 97th and 85th, the 85th and the 50th, and the 50th and the 15th percentile estimates. Both studies found estimates of the 97th percentile to be less than what would be expected (Bobko et al., 1983; Burke & Frederick, 1984). These results were not consistent with those that should have occurred if subjects were operating from a normal distribution. To explain this finding, Bobko and his colleagues suggested that judges may have rated performance using a rectangular distribution rather than a normal distribution since the former is less cognitively taxing. The importance of this phenomenon can not be emphasized enough since one of the rudimentary assumptions of utility analysis, when employing the Cronbach-Gleser-Brogden model, is that all underlying distributions are normal. Some authors (c.f. Burke & Frederick, 1984) have suggested this non-normal estimate of percentile points may be due simply to the temporal order of the requested estimates. That is, they suggest that the occurrence may be due to the fact that the judges were asked to estimate the 85th then the 97th given the 50th. This suggestion is, however, repudiated by Bobko et al. (1983); they replicated the study varying the order and obtained similar results.

One final issue, concerning the estimate of percentiles, involves the verbal cues provided to the judges. The request for an estimate of the worth of the performer at the 85th percentile is anchored to the adjective 'superior'. The request for an estimate of the worth of the performer at the 15th percentile is anchored to the adjective 'inferior'. The question concerns the use of the words "superior" and "inferior". Specifically, what is the effect of using language as a metric? Bobko et al (1987) state that there is a complete lack of research in this area.

Estimates Limits

One characteristic associated with the present methods of estimating the percentile point values is the fact that negative estimates are not sought and, for all intents and purposes, are not possible. Although these estimates are not sought they are possible. For example, it is possible that a supervisor will associate the employee working at the 15th percentile with negative worth. That is, this employee costs the organization more than he/she brings in. It would be highly implausible to expect a judge to come up with a negative estimate using the Schmidt et al. (1979) procedure or the CREPID method. The Schmidt et al. (1979) procedure contains, as an aid, the statement "... consider what the cost would be of having an outside firm ... provide the products and services" (Schmidt et al., 1979, p. 621). Therefore a judge would have to consider a situation in which an outside firm was paid to provide products or services in order to reach a negative estimate. The CREPID method, which involves the fractionation of salary, will always provide a non-negative result.

With respect to this idea of negative worth some unusual findings have been reported by researchers. Specifically, some researchers have reported negative estimates at the 15th percentile (Burke & Frederick, 1984; Karren & Bobko, 1983; Mathieu & Tannenbaum, 1985; Weekly et al. 1985). In the Mathieu and Tannenbaum (1985) study of bank branch managers, the authors found a negative estimate for the 15th percentile, probably due to the fact that this manager was associated with a branch that was losing money. Karren and Bobko (1983) looked at the performance of U.S. Marshalls assigned to the Witness Security Division. Again the 15th percentile performer was associated with a negative estimate. Burke and Frederick (1984) found negative estimates for the 15th percentile performing district sales manager and Weekly and colleagues (1985) also reported negative estimates at the 15th percentile. In this case the performers were convenience store managers.

The important point here is these negative estimates were obtained through methods which have a "demand characteristic" (Bobko et al., 1987) for positive values. Future research is needed to examine what happens to SDy estimates when these bounds are removed. Do they increase, does the distribution become positively skewed, do managers accept the idea of negative worth, etc?

Performance Distributions

One of the assumptions underlying the utility equation is the normality of the underlying distributions. Some researchers (c.f. Schmidt et al., 1979) have used the following method for testing the normality of the SDy distribution. The differences between the 50th and the 15th and the 50th and the 85th have been compared. A non significant difference has been taken as evidence of the normality of the distribution. This, however, is not a rigorous enough test. Any symmetrical distribution would satisfy this criterion. In other words this test is necessary but not sufficient.

Another concern with the underlying distribution is what population are the judges using in the estimation. It has been suggested that raters use the incumbent job population rather than the candidate population as a frame of reference (Bobko et al., 1987). The effect of such a reference frame is a negatively skewed distribution due to the organizational constraints imposed on the lower limit of production. In other words, employees must meet some minimum standard in order to retain their jobs.
Ancillary knowledge held by the raters may also act to skew the

distribution. For example, Karren and Bobko (1983) had supervisors rate the performance of U.S. Marshalls. The supervisors were aware of the extremely low selection ratio (5%) and the intensive post selection training program. The net effect was that supervisors equated average and superior performance and gave extremely low estimates for the worth of the inferior employee's performance; thus, producing a negatively skewed distribution.

In a related study, Schmidt, Mack, and Hunter (1984) obtained a skewed distribution of the performance for Park rangers. Again, the authors proposed an explanation focusing on some aspect of the rating process to explain the skewed results rather than the idea that the distribution was, in reality, skewed. These authors thought the managers could not differentiate along the entire performance continuum equally, thus, producing negatively skewed results.

The results of studies of the estimated performance distribution are inconclusive. In some cases the authors state the distribution can be considered normal (Schmidt et al., 1979) in others they conclude the distribution is non-normal. The question is who is right? The present study undertook to answer this question by investigating the distribution shape by using a large sample size and by varying distances between the rater and the rated position.

Summary

The preceding discussion is meant to be a summary of the research concerning psychometric issues as they apply to the estimation of SDy.

There are a number of other concerns with the estimation of SDy, as pointed out earlier, such as contextual factors and cognitive components of SDy estimation. The present research, however, is primarily concerned with the psychometric aspect.

Alternative Methods of Estimating SDy

The preceding sections dealt with problems inherent in the estimate of SDy. Alternative methods of assessing utility exist that do not require such an estimate. For the purpose of the present study the introduction of two of these alternative methods is required. As stated earlier some authors have borrowed the technique of psychometric convergence. Convergence requires that a number of methods be used to measure a phenomenon and the results be compared to see if the different methods provide the same conclusions. In the case of utility research a number of methods are used to compute the SDy value and the results are compared. The alternative methods used here are the Superior Equivalents Technique and the 40% method.

The Superior Equivalents Technique was developed by Eaton and his colleagues to deal with instances where applying dollar metrics to point estimates would be difficult or impossible (Eaton, Wing, & Mitchell, 1985). The authors were primarily concerned with the estimation of Tank Commander worth in the U.S. Army. To combat the intransigence on the part of some of their past subjects, these subjects refused to provide dollar estimates, they developed an alternative that required only that raters make relative judgements in terms of numbers. Specifically, the Superior Equivalents Technique requires raters to estimate the number of superior performers (85th percentile) it would take to do the work of 10 average performers (50th percentile). The SDy of job performance is then computed using the average salary paid. It is important to note that the entire technique is built on the belief that the average performer is paid what he/she is worth. The authors underscore this assumption in the statement " one might assume organizations pay average employees about what they are worth" (Eaton, et al., 1985, p. 31).

The 40% method was developed by Schmidt, Hunter and Pearlman (1982). These authors found in the their previous research that estimates of SDy fell between 40% and 70% of the salary of the position rated. Previous research had revealed SDy to be 42% of salary for entry-level Park rangers (Schmidt et al., 1984); 55% of salary for computer programmers (Schmidt et al., 1979); and 62% for budget analysts (Hunter et al., 1982). They argued that 40% of the salary could be employed as a conservative estimate of SDy when the situation was such that alternative methods could not be employed.

Recent Studies

A recent study was conducted by Johnson under the auspices of the Canadian military and Saint Mary's University. Briefly, Johnson examined the impact of a number of factors on SDy estimation: the judges' position relative to the judged position, contextual factors, the judge's understanding of percentile points and distributions, order of estimates, and cognitive dimensions (Johnson, 1989). In his study Johnson had upper level officers in the Canadian military, Lieutenant, Lieutenant Commander, and Commander make estimates about the position of Sub Lieutenant. A number of interesting results occurred. Johnson found that between judge variance of estimates was related to both position and experience, with position impacting more than experience. He found the variance to be greatest among the most experienced group, the group who were also the farthest distance from the rated position. The implications being that the closer the rater is to the position to be rated the less the variability across the estimates. Johnson also found that the assumption of normality could not be supported from his data. He found the overall distribution to be negatively skewed. Specifically, he asked respondents for three estimates, the 85th, the 50th, and the 15th percentiles. From these he estimated two values of SDy. He found the estimates of the lower SDy (50th - 15th) to be larger than the upper SDy estimate. This result again indicated the mercurial nature of the estimated performance distribution.

The Present Study

The present study examined a number of aspects of SDy estimation. It paralleled Johnson's (1989) study in a number of ways ; but it departed from that study in two important respects, namely the position of the raters and the position rated. Whereas Johnson's study concerned itself with one position to be rated and with the raters from varying positions, the present study was concerned with multiple rated positions and one level of rater.

In this way an attempt was made to replicate the findings of Johnson that relative position affects the variance of the estimates of the individual percentile points. The study also sought to investigate the shape of the estimated performance distribution. It departed from some previous studies in that an additional estimate was required. That is, whereas past research has asked respondents for estimates at the 15th, 50th, and 85th percentiles this study required an additional estimate be made at the 97th percentile. In this way the study followed the course set by Burke and Frederick (1984) who added the additional requirement.

In order to test the idea that relative position affected SDy values raters were asked to provide estimates for one of three different positions. The first was one that they should have had great experience with, the second, one they should have had moderate experience with, and the third, one in which they were likely to have little experience with. For the first, the position in which the raters were to have had a great deal of experience, the respondents were asked to provide estimates on a position they had held in the past, either as a full time, part time, or summer job. The moderate experience condition required that a position be chosen that respondents were readily familiar with but had not actually functioned in. In this case the position of university professor was chosen. Since the respondents were undergraduate students it was felt that this position would satisfy the requirements. For the final position it was necessary to choose a job in which respondents would recognize, but not be totally familiar with (i.e, not totally familiar was operationally defined as not having regular contact). The position of Member of Parliament was selected. It was hypothesized that by varying the position of the rater viz a viz the estimated position the variance of the percentile point estimates would increase as relative distance increased.

The second purpose of the present study was to answer the questions 'What is the shape of the SDy distribution?' and 'How does it vary with varying relative distances?' Does the distribution approximate the normal distribution or is it of a non-normal form? As stated earlier there is some argument over the shape of the SDy distribution generated by raters. One such argument states that raters will produce a rectangular or flat distribution since this shape is easier to process than a normal distribution. The present study is a fundamental examination of the shape of the distribution in that it is not so much concerned with the factors leading to the distribution shape but addresses itself only to the shape of the distribution.

The method of determining the shape will be a variant of the global method. The procedure, in its pristine form, requires that raters make three estimates of the worth of a position to an organization. Specifically, judges are asked to give estimates of the worth of individuals performing at the 15th, 50th and 85th percentiles. The present study will ask for, in addition to the regular three estimates, an estimate of the worth of an individual who is performing at the 97th percentile. The rationale for requesting this additional estimate is as follows. Previous attempts to examine the shape of the estimated performance distribution have taken, as evidence of normality, a non significant difference between the upper estimate minus the average and the average minus the lower. The problem is a non significant difference does not support the hypothesis that the distribution is normal. It only supports the hypothesis that the distribution is symmetrical. It is believed that the additional required estimate will overcome this criticism. Also, for exploratory purposes, the present study will require that participants choose, from a number of options, a curve that best fits their interpretation of how performance is distributed.

The hypotheses of the present study are as follows:

> Hypothesis 1 — The between judge variance in point estimates of the dollar value of performance will significantly increase as judge's experience with the performance being estimated decreases.
>
> Hypothesis 2 — The shape of the distribution will depart from normal as the relative distance of the rater from the rated position increases.

In addition to the data collected for testing of the above hypotheses, data will be collected which will allow a convergence study of three alternative methods of SDy, a variant of the Schmidt et al (1979) procedure, the Superior Equivalents Technique, and the 40% method.

Method

Participants

The participants for the study were students of Saint Mary's University in Halifax, Nova Scotia. Their participation was on a strictly voluntary basis. Those participants who were enrolled in Introductory Psychology received some course credit as an incentive. The sample contained 443 students from varying years, faculties, and genders.

Description of the Sample

Figure 1 shows the composition of the sample with respect to gender. Approximately fifty eight percent of the sample were female while forty-two percent of the sample were male.

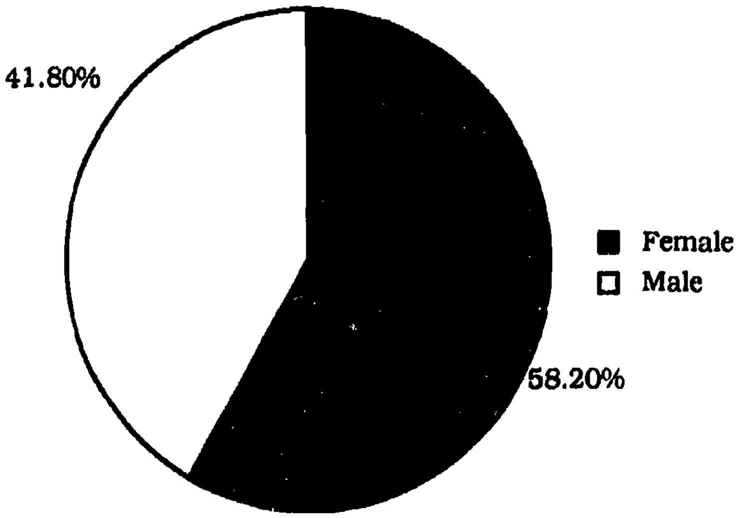


Figure 1. Breakdown of sample with respect to gender.

Figure 2 shows the distribution of the different faculties in the sample. The breakdown of the sample with respect to faculty was as follows: the majority of the respondents were enrolled in the Arts faculty (59.4%), the faculty of Commerce provided 22% of the sample, Science made up 16.94%, Engineering contributed a little over 1% (1.39%), and Education students made up .23% of the sample.

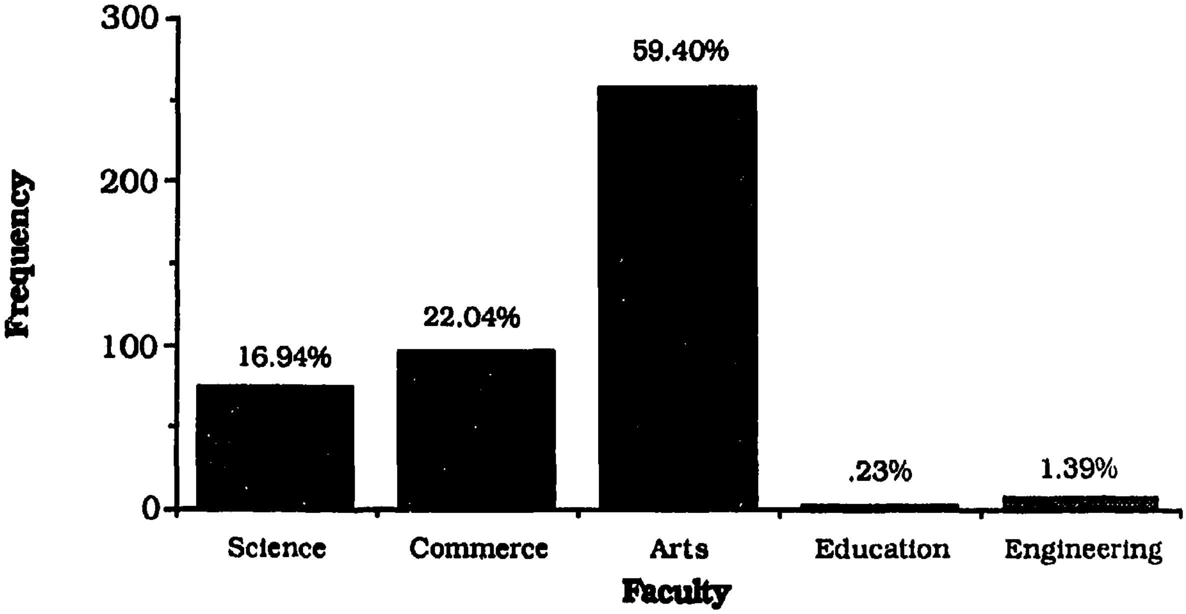


Figure 2. Distribution of faculties in sample.

Figure 3 shows that the sample consisted primarily of first year students. There was a steady dropoff of students from the first (63.21%) to second (20.52%) year of studies, second to third (11.79%), and higher (4.48%). The breakdown of the sample with respect to year is shown in Figure 3.

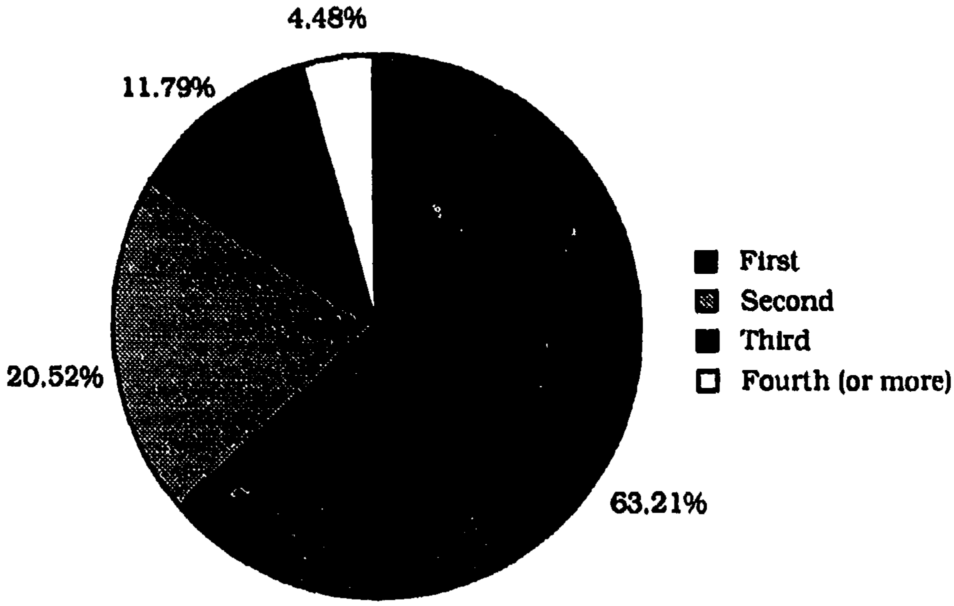


Figure 3. Breakdown of the sample with respect to year of study.

Since one of the hypothesis was concerned with the subjects' understanding of the normal distribution, subjects' knowledge of statistics could have an effect on their estimates. Therefore, the sample was examined to determine how many participants had taken or were taking a statistics course. Nearly three quarters of the respondents reported that they had not taken or were not taking a statistics course (Figure 4).

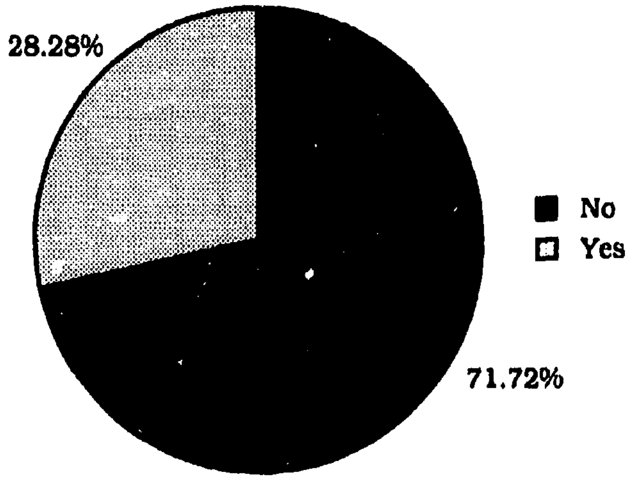


Figure 4. Participants' response to the statistics question.

Design

A between group design was used in the study. Participants were randomly assigned to one of three conditions. These conditions differed in the position the respondents were asked to rate. A respondent could be asked to make estimates about a position that he/she was readily familiar with (self estimates), the position of university professor, or the position of member of parliament. This was the distinguishing characteristic across the groups. All other factors were held constant (i.e., the questions, the data collection, instrument, design, etc.). The total number in each group varied slightly due to the random assignment of respondents to groups. Group 1 which asked for the self estimates had 152 respondents, Group 2 concerned with university professors had 148, and Group 3 requiring estimates about MPs, had 143.

Instruments

A four part questionnaire was used to collect data. The first part of the questionnaire was a variant of the one used by Schmidt et al., (1979) with some notable departures. The second part consisted of a page of graphically depicted distributions from which the participants were to choose one to indicate their idea of the performance distribution for the particular job. The third part was a variation of the Eaton et al. (1985) questionnaire. It was used to collect the data for the Superior Equivalents Technique. The fourth part was a short demographic questionnaire. Three variations of this questionnaire were used to collect the data for the three different positions. Samples of the questionnaires are shown in the appendices. The questionnaires used to collect data for the self estimating group, for the professor group, and for the MP position are shown in Appendix A, B, and C, respectively.

Part one of the questionnaires was used to collect the estimates of performance at the different percentiles. As mentioned earlier there were some significant departures from the original questionnaire used by Schmidt et al. (1979). The first concerned the number of estimates requested. Specifically, whereas the original procedure asked for three estimates, the 15th, the 50th, and the 85th percentiles, the present asked for an estimate at the 97th percentile as well. The second departure concerned the wording of the original instrument. As an aid to making the estimates, it contained the following sentence: "In placing an overall dollar value on this output, it may help to consider what the cost would be of having an outside firm provide these products and services." It was thought that this statement would not help the participants generate an estimate but would only serve to confuse them given the nature of the jobs being used here. Also, previous studies used managers as raters whereas the present study utilized students who were less likely to be familiar with the costs associated with contracting for services.

The second part consisted of a page of different distributions, skewed positive and negative, rectangular, bi-modal, and normal. Participants were asked to check the one that best represented their idea of the distribution of performance for the particular position they estimated.

Part three required the participants to make an estimate of the number of superior performers it would take to do the work of a set number of average performers.

Part four collected demographic information from the subjects.

Procedure

The questionnaire was pilot tested with two different groups of subjects in order to ensure clarity and understanding in its wording.

Data were collected in classrooms of professors who agreed to cooperate in the study. Questionnaires were distributed at the beginning of the class period and were immediately filled out and returned by those students who agreed to participate.

Participants in the study received a questionnaire for only one of the three aforementioned positions.

Analysis

Before any formal analyses were undertaken the data were examined for outliers. The removal of outliers involved the computation of means and standard deviations for each of the conditions and the subsequent removal of any estimates which fell outside of three standard deviations. The means and standard deviations were recomputed and the data again checked for outliers. This process was iterated until there were no outliers in any of the estimate distributions. In total the data from 15 respondents had to be cut, eight respondent's data in the self condition, three respondent's data from the professor condition, and four respondent's data from the MP condition. The removal of these data did not change the results of the analyses. That is, the data analyzed with and without the outliers provided the same results.

The next step consisted of the computation of the SDy values for the three conditions. The values were calculated in the following manner. The difference between the 50th percentile estimate and the 15th percentile estimate became SDy1, the difference between the 85th percentile estimate and the 50th percentile estimate became SDy2, and the difference between the 97th percentile estimate and the 85th percentile estimate became SDy3. These values were then checked for inconsistency. Inconsistency was operationally defined as negative values of SDy. These negative values occur when a respondent gives a higher dollar value for a lower estimate than he/she give for a higher estimate. For example, a respondent might estimate the value of a 15th percentile performer at $20,000 per year and the value of a 50th percentile performer at $15,000 per year. The resulting SDy1 value would be -$5,000 per year. The idea of inconsistency, then, comes from the inconsistent assignment of values to the original estimates. The gleaning of inconsistent responses coupled with the removal of outliers resulted in the data from 25 respondents being dropped. Fourteen respondents from the self condition, seven from the MP condition, and four respondents from the professor condition were removed under this criterion.

In addition to treating the responses of the participants as a group the data were broken down with respect to a number of demographic variables and analyzed for support of the hypotheses. The respondents were placed in groups based on their gender, statistical background, year of study, and faculty. These groups were then compared on the basis of support for the hypotheses. It was found that the support for the hypotheses did not change, therefore the results for the entire group are reported here.

Hypothesis 1 was tested by comparing the variances at the four percentile points for two of the estimated positions, professor and MP. The data from condition 1, self estimates, were not used here. The rationale for omitting the data from condition 1 was as follows. The hypothesis stated that variance in the point estimates would increase as the relative distance between the rater from the rated position increased. At first glance it might seem appropriate to include the data from the first condition. It is important to note, however, that the hypothesis dealt with variance of the point estimates and therefore required that the judges be judging the same position. That is, it was necessary for each judge to be using the same anchor. The hypothesis was tested using two statistical tests, Cochranes C and Bartlett-Box F.

Analysis of the data for Hypothesis 2 required the computed values for SDy1, SDy2, and SDy3. The Wilcoxon Matched Pairs Signed Rank test was used to evaluate the normality of the distribution in the three conditions.

Chi square tests were used to analyze the relationship of the respondents' statistical knowledge and the perceived performance distribution for the position they estimated. As well, a chi square test was used to determine the relationship between the position rated the performance distribution selected as representative of that position..

The convergence of the different methods of estimating SDy was measured by comparing the results of the global method, the Superior Equivalents technique, and the 40% method. The 40% method and the superior equivalents technique called for an estimate of average salary. In the case of the participants in the self estimate condition the salaries were obtained from the questionnaire and averaged across all respondents. The mean salary was used. For the professor's position the average salary was obtained from the administration of Saint Mary's University. The average salary of Members of Parliament was obtained from Public Information Office of the Canadian House of Commons.

Results

Overview

The first hypothesis predicted that the amount of variance within each of the point estimates would significantly increase as the relative distance between the rater and the rated position increased. This was supported by the data.

The second hypothesis dealt with the shape of the estimated performance distribution. This hypothesis predicted an increasingly non-normal distribution as the relative distance between rater and the rated position increased. Respondents in Group 1, who made estimates for their own position, produced SDy1 and SDy2 values that did not differ significantly, supporting the normality assumption. However, their estimates produced SDy2 and SDy3 values that did differ significantly from one another. Respondents in Group 2 and Group 3 made estimates that led to significant differences among all SDy levels.

Participants chose the normal distribution as indicative of performance in approximately 40% of the cases. There was a relationship between the condition and the distribution chosen but none between the respondents' statistical background and the distribution of choice.

The study showed little convergence among the three methods of obtaining SDy estimates. This lack of convergence was accompanied by a lack of consistency across the estimates. The superior equivalents technique provided the highest SDy estimates in Groups 2 and 3 and the moderate estimate for Group 1.

Hypothesis 1

The first hypothesis stated that the within point estimate variance of the dollar value of performance would significantly increase as judge's experience decreased. In order to evaluate the validity of this hypothesis two tests for the homogeneity of variance were performed for each of the percentile estimates between the positions of professor and MP. The means and standard deviations for these estimates are given in Table 3. The results of the two tests are given in Table 4.

A visual inspection of Table 3 reveals a discrepancy in the variance at the four point estimates between the two positions. The standard deviation for the point estimates of the position of MP are approximately 3 to 6 times larger than the estimates for the position of professor. The results of the statistical tests confirm this visual inspection. In all cases there are significant differences between the different position point estimates (See Table 4).

Table 3. The means and standard deviations of the subjects estimates for the four percentile points broken down by position rated.

		Estimates			
	n	15th	50th	85th	97th
Own Job					
Mean	138	11480	19986	26961	33219
Standard Deviation		12264	20774	25686	31668
Professor					
Mean	144	28302	39479	52736	62897
Standard Deviation		9609	10723	15399	19920
MP					
Mean	136	33563	46147	72985	92988
Standard Deviation		35457	34436	82773	119146

Note: All figures are in dollars

Table 4. Tests of the homogeneity of variance between professor point estimates and MP point estimates.

Estimate	Cochrane's C	Bartlett-Box F
15th percentile	.9290*	192.025*
50th percentile	.8993*	145.864*
85th percentile	.9639*	285.422*
97th percentile	.9710*	316.244*

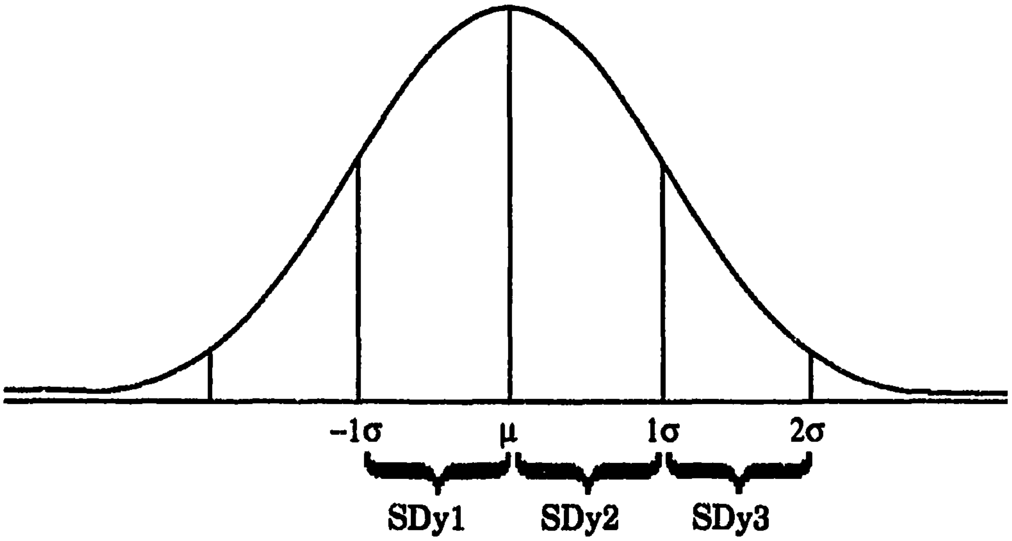
*$p < .001$

The results of these two tests of variance support the first hypothesis. That is, the data show significantly more variable estimates of worth, at all percentile points, between the two positions. An inspection of Table 3 confirms the direction of the hypothesis.

Hypothesis 2

The second hypothesis predicted that the shape of the estimated performance distribution would depart from normal as the relative distance between the rater and the rated position increased. Specifically, the distribution will be more aberrant for the SDy values of professor and MP than for the participants' own job SDy values. Table 5 shows the means for the three SDy values of the three conditions.

Table 5 Means of the SDy estimates for the three positions: the respondent's own, the position of professor, and the position of MP.



	SDy1	SDy2	SDy3
Self			
Mean	25810	25703	19879
Professor			
Mean	11177	13256	10161
MP			
Mean	15113	28568	21132

The SDy values within each position were then compared using the Wilcoxon Matched-Pairs Signed-Ranks test. The results of the tests are given in Table 6.

Table 6. Results of the Wilcoxon test for the SDy estimates.

	Comparison	
Position	SDy1 with SDy2	SDy2 with SDy3
Self	-0.654	-2.6**
Professor	-2.186*	-4.332**
MP	-3.864**	-3.071**

*$p < .05$
**$p < .01$

The results provided partial support for Hypothesis 2. Table 6 provides evidence for a non-significant difference between the SDy1 and SDy2 values in condition 1: This result directly supports Hypothesis 2. Table 4 also shows results that are contradictory to the hypothesis. Specifically, a significant difference between SDy2 and SDy3 of condition 1. All other tests revealed significant differences between SDy values: thus, supporting the hypothesis that there would be significant differences in SDy estimates as the relative distance increased.

In addition to the apriori plans of comparing SDy1 with SDy2 and SDy2 with SDy3 a post hoc comparison of SDy1 with SDy3 was performed. The results of this analysis did not change the lack of support for the hypothesis, although it did provide some interesting contrasts. In the self rating condition this comparison was found to be significant. In both of the other conditions this comparison was found to be non significant.

Performance Distribution

When asked to select a distribution that best represented their idea of the real distribution of performance, only 42% of all respondents chose the normal curve while 58% chose a non-normal distribution: 22% chose a positive skewed graph; 15%, negative skewed; 11%, bimodal and; 10% rectangular. Less than one percent drew their own graph. Figure 5 shows the frequency and percentage of each distribution in the responses. A chi square test was used to check the hypothesis that this distribution of scores did not depart significantly from a theoretical 50/50 distribution. The results supported the null hypothesis. That is, the distribution did not differ significantly from the theoretical distribution.

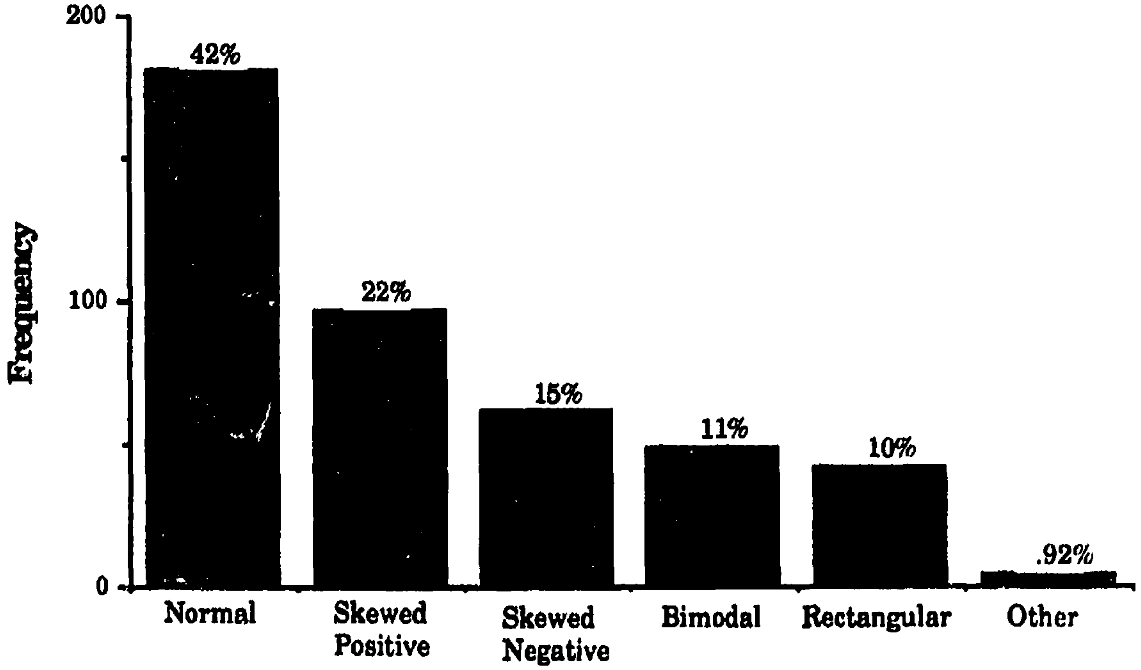


Figure 5. Frequency histogram for each of the graphical choices made by the respondents.

The performance distribution chosen by a respondent was related to the position rated There was a significant relationship between the position the participants rated and the graph they chose as best representing the performance distribution for that position ($\chi2 = 22.055$, $p<.05$). Figure 6 shows the breakdown of the distribution frequency by condition. The relationship runs contrary to the earlier findings that provided partial support for the normality of the self estimates. In this case participants in the self estimating group chose the highest percentage of non-normal distributions (64%); followed by participants in the professor group (57%) with those estimating the performance of MPs choosing the lowest percentage (55%). The steady decrease in participants choosing non-normal distributions as the relative distance increased is exactly the opposite of what was predicted in Hypothesis 2. There was no relationship between the distribution the respondent chose and whether the respondent had taken a statistics course ($\chi2 = 6.33$, ns).

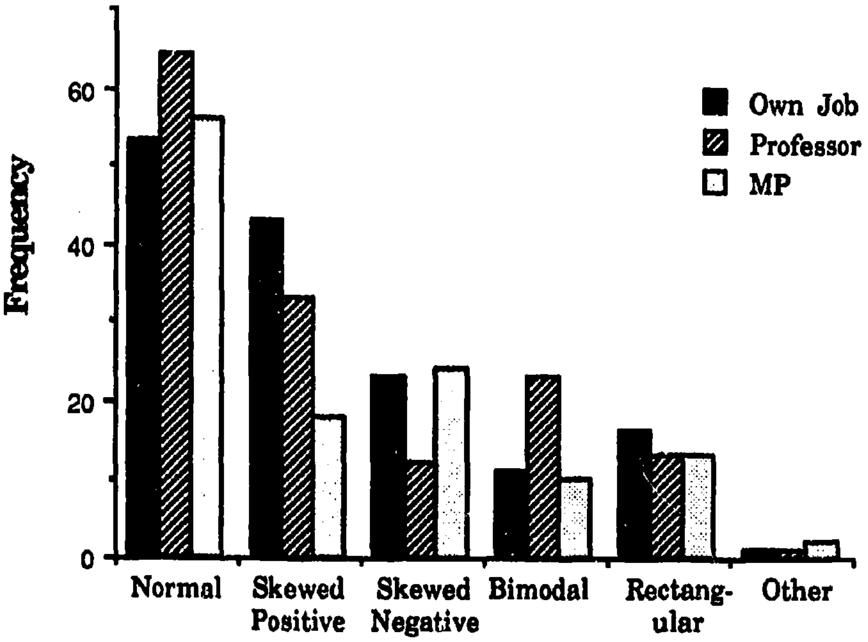


Figure 6. The breakdown of distribution choice by condition.

Comparison of Methods

A convergence study compared estimates of SDy from three different methods; the global method, the superior equivalents technique and the Schmidt et al. (1982) 40% method. The computations for these methods are given in Appendix D. A comparison of the results revealed little or no convergence among the different methods. The results are given in Table 7 and are presented for each of the three positions used in this study.

Table 7 Values of the SDy estimates for the different methods

	Global				
	SDy1	SDy2	SDy3	40%	Superior Equivalents
Self	25810	25703	19879	4454	11134
Professor	11177	13256	10161	20004	50010
MP	15113	28568	21132	30600	88170

Note: All figures are in dollars

Since both the Superior Equivalent technique and the 40% method are salary based representative salary information for the three conditions was needed. The salary data for the familiar job was obtained by averaging the income that Group 1 respondents reported earning for that position. The salary data for the professor position was obtained from the administration of Saint Mary's University. The salary information for the MP condition was obtained from the Public Information Office of the House of Commons.

Inspection of Table 7 reveals little convergence among the methods. Widely divergent estimates were obtained from the three methods. Where participants rated their own position, the global method produced the largest estimate of SDy, the 40% technique, the smallest. On the other hand, where participants rated a position other then their own, the Superior Equivalents technique resulted in the largest estimate of SDy and the global method, the smallest (Figure 7 shows the interaction between method and position).

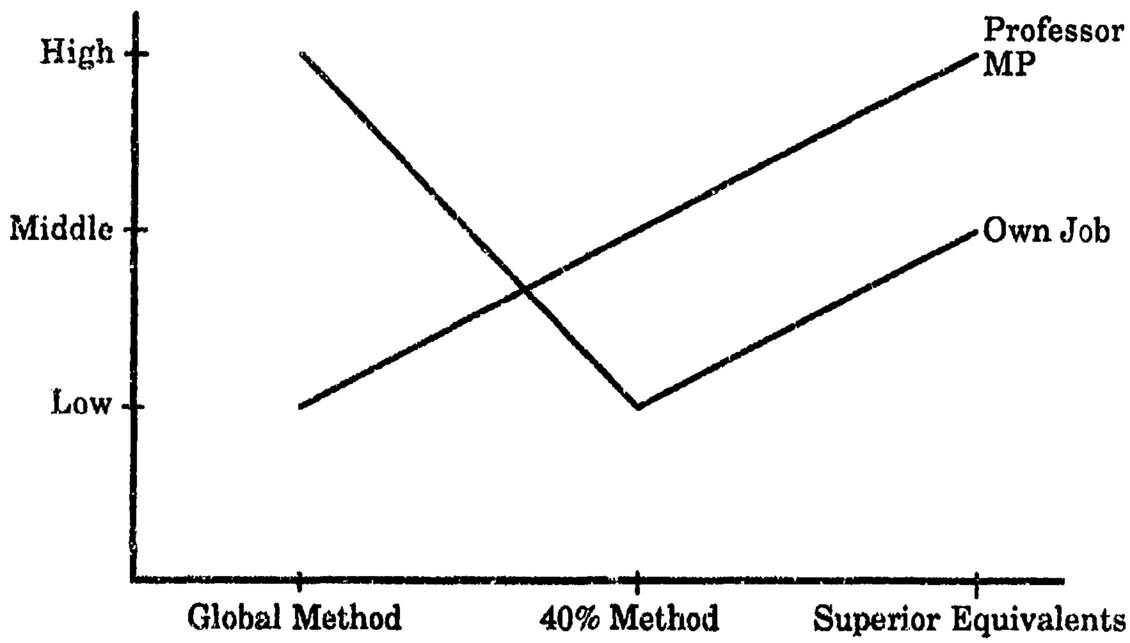


Figure 7. The interaction between Estimation method and condition.

Discussion

The goals of this study were primarily concerned with two aspects of SDy estimation: (1), the variance of the percentile point estimates across jobs that differed in familiarity to the rater and, (2) the normality of the resulting performance distribution generated by the percentile estimates. As well, the respondents' belief in the underlying shape of performance distributions was checked in relation to the similarity of the position they rated and their statistical knowledge. Finally, the convergence between SDy estimates resulting from different estimation methods/techniques was investigated.

Variance in Percentile Points

The analysis of the data supported the hypothesis that the percentile point variance increases as the distance between the rater and the rated position increases. The variance in the estimates of those respondents who estimated the performance levels of the MP position was three to six times as large as the variance in the estimates made by participants rating the professor position. These findings agree with those reported by Johnson (1989) in his study of the Canadian military. Johnson concluded that the percentile point estimate variability formed a direct relationship with the rank and experience of the judges. That is, as rank and experience increased so did the estimate variance. Johnson explained this result by postulating that the effect was due to the more recent experience of the lower rank judges in the position estimated. This explanation, although logical for Johnson's study does not fit here. Since the judges were students and the two estimated positions of interest in this hypothesis were professor and MP it is highly unlikely that any of the respondents had functioned recently, or ever, in the position they estimated. A more likely explanation would appear to be based on the effect proximity has on the estimates without any reference to the rater's prior experience in the position rated. A similar effect has been found by Landy and Farr regarding performance appraisals (landy & Farr, 1983). They found that daily and direct contact by the raters with the ratees is an important factor in rating variance. This explanation would support the present finding that the variance of MP's estimates was significantly higher than the variance of professor's estimates. The students have daily contact with their professors but are unlikely to have direct contact with their MP.

The increase in the percentile point variance with the relative distance of the position of the rater also supported Mayer's (1982) findings. He reported standard deviations between one and twenty times larger for the estimates of bank teller's performance made by branch mangers compared with the same estimates made by head tellers. An interesting aside, however, is the fact that the branch manager's estimates, although the most variable, were also the closest to those computed through an accounting procedure.

The present study, along with Johnson's (1989) and Mayer's (1982) provide a consistent finding regarding SDy estimates. The amount of inter-rater variability is contingent, in part, on the relative distance between the rater and the rated position. Apparently, the more familiarity the rater has with the rated position, the lower the inter-rater variability.

The existence of this problem of inter-rater variance has been known for a number of years. Two studies undertaken in 1984 sought to reduce this variability (Burke & Frederick, 1984; Wroten, 1984). Burke and Frederick used two feedback approaches in an attempt to reduce percentile point variance. The two approaches both involved the calculation of the average 50th percentile estimate and feeding back this information to the raters before they made any additional estimates. The procedures differed in that one involved individual feedback and the other involved group feedback. Burke and Frederick found that both forms of feedback reduced the amount of percentile point estimate variance. Wroten (1984) also used feedback in an attempt to reduce variability. Although it did not reduce estimate variance, he nevertheless recommended it from a practical standpoint. Wroten felt the feedback facilitated group discussion. Recently, the delphi method of reaching consensus on decisions was employed as an aid to reducing estimate variance. Tannenbaum and Dickinson (1987) found the use of this technique to reduce estimate variance. It appears, then, from the literature that the focus of research on estimate variance has taken one of two approaches 'What is the best way to facilitate good estimates?' or "Who is best suited to make the estimates?'

The present finding that familiarity can impact on the variability of estimates suggests alternative ways of handling the issue. First, it suggests the best raters may be those who function closest to the rated position. It may, however, not always be possible to employ raters that satisfy this criterion. Therefore, the maximum about of information and training should be provided to the raters about the position be rated. This may act to combat the effect of unfamiliarity. Second, it appears that feedback may be useful in reducing variance. Perhaps what is needed at this point is an investigation of the two variables combined. That is, a research study that involves raters from varying relative distances and the presence and absence of feedback. Such a study might illuminate some clues to the interaction effects, if any, of these variables.

An argument might be made for standardizing the estimates made for the Professor and MP positions in order to control for a perceived larger salary for MPs over Professors. This larger salary might be construed as the reason for the larger variance between the two estimates. A review of the statistical literature, however, reveals no instructions indicating the need for such an operation. Neither Bruning and Kintz, (1988); Glenberg, (1988); Hinkle, Wiersma, and Jurs, (1988); nor Keppel, (1982) advocate such an operation. Instead they propose a simple ratio of the variances as the test of homogeneity, with no equating of scores.

Performance Distribution Shape

The second hypothesis dealt with the normality of the estimated performance distributions. Specifically, it predicted a move to non-normality as the relative distance increased between the rater and the rated position. The results are equivocal. If the Schmidt et al, (1979) test of normality is utilized then the hypothesis is supported. The established test requires only that SDy1 and SDy2 estimates be compared. A non significant difference between these two is taken as support for the normality assumption. According to this accepted procedure the hypothesis is supported. The data show that there is a non significant difference between the SDy1 and SDy2 for the self estimates; but, this difference became significant for the professor and MP estimates. Based on the Schmidt et al. (1979) procedure, it would appear that the self estimates are normal and the professor's and MP's are not.

However, this procedure is not rigorous enough. Although the Schmidt et al. requirement of equivalent SDy estimates is necessary, it is not sufficient; any symmetrical distribution will satisfy the test. This study employed a test which only normal distributions satisfy. Under this criterion three, not two, estimates of SDy must be non significant. This test requires that three estimates of SDy be obtained. If there is an underlying normal distribution, then the SDy estimate based on the difference between the performance at the first and second standard deviations should be the same as that between the mean and the first standard deviation above or below the mean. Using this more stringent test, the second hypothesis cannot be supported. That is, under this test all of the estimated performance distributions appear to be non-normal. However, there is a greater implication for this study: the assumption of normality which is crucial to the utility model, does not appear to hold for SDy estimates obtained through the global procedure.

Other authors have applied this type of rigorous testing of the normality assumption with similar results. Bobko, Karren, and Parkington (1983) found a significant difference between the upper SDy estimate (analogous to

SDy3 in this study) and two middle estimates (SDy1 and SDy2 in this study). Burke and Frederick (1984) have also employed this test and obtained similar results — a non-normal distribution for the performance estimates.

If the results do not support normality what is the shape of the performance distributions? Bobko et al. (1983) have pointed out that respondents may operate under a rectangular distribution since it is easier to process cognitively. If the respondents were operating under a rectangular or linear function then the results would be expected to follow a straight, linear path. That is, the hypothesis would predict that the two middle estimates of SDy would be the same and the upper estimate would be approximately one third of the middle estimates. The reasoning for this prediction is simply based on the fact that subjects would interpret the percentages in a linear fashion. The 85th percentile would be perceived as 35% higher than the average and the 97th percentile estimate would be twelve percent higher than the 85th. An inspection of Table 5 does not support this interpretation. Table 5 shows the mean values for the three SDy estimates across the three positions. The self estimates come closest to following a rectangular distribution. The middle two estimates are virtually equal with the upper estimate being smaller; however, the upper estimate was approximately 80 percent of the lower estimates rather than 33 percent. This larger than expected upper estimate in relation to the lower estimates comes closer to supporting the normal curve hypothesis than the rectangular distribution hypothesis. The results for the other two position estimates are not so clear. For the professor estimates all three

SDy estimates are largely different. The upper estimate is the smallest, followed by the lower (SDy1), then the middle (SDy2). It is difficult to attach a label to describe the shape of the underlying distribution. The lack of a clear distribution shape also exists for the MP estimates: SDy2 is the largest, followed by SDy3, and finally SDy1. Although this result was not exactly the same as the professor estimates it is also neither rectangular nor normal.

Respondent's Perception of the Distribution Shape

The above analysis of the performance distribution based on SDy estimates was largely inconclusive, although for the most part it suggested a non-normal distribution. Therefore, a more direct procedure was used to assess the nature of the distribution. The nature of the underlying performance distribution was pursued through direct questioning of the respondents. The respondents in all groups were asked to select a graph which reflected the performance distribution for the position they rated. Overall, the majority of the respondents chose a non-normal distribution. The fact that most respondents chose a non-normal distribution supports the earlier finding in Hypothesis 2 that the performance distribution is non-normal or is at least perceived to be non-normal.

The familiarity of the respondents with the position they rated influenced the performance distribution they chose as the most representative. Those who rated the MP position chose the normal distribution more frequently (46%) compared to those rating professor (44%) and those giving self estimating (36%). The order of these results are in direct opposition to those obtained from the analysis based on the Schmidt et al. procedure. The Schmidt et al. analysis showed the self estimates to be closest to fitting the normal distribution whereas the direct questioning revealed that respondents rating their own performance chose the normal distribution less than those rating positions other than their own.

An analysis of this paradoxical finding was beyond the scope of this research. It may; however, be the result of the judges inability to comprehend the meaning of percentile points and distributions. Future research could incorporate a pretest of the judges understanding of these concepts.

Convergence

There was a complete lack of convergence across the different methods. For the respondents who rated their own position the 40% method provided the lowest estimate, the superior equivalents provided the next lowest, and the global the highest. The results for the other positions did not follow this pattern. For both the professor and MP positions the order was as follows: global method, 40%, and superior equivalents — lowest to highest. This lack of convergence has been reported by other authors (c.f. Johnson, 1989; Weekly et al., 1985). Bobko, et al. (1983) have proposed an explanation this lack of convergence by stating that the 40% method and the superior equivalents are directly based on salary whereas the global procedure "allows judges to estimate percentile values, using whatever cues the judges wish ..." (p. 8). Although it is impossible to state conclusively from this study the reason for such discrepant estimates the fact remains that the global procedure consistently produces results that are different than than the other methods.

One interesting note concerning the SDy values is the fact that the 40% method failed to provide the most conservative estimates in two of the groups. This is in direct contradiction to the idea proposed by Schmidt et al. (1982) who suggested that 40% of the salary for a position be employed as a conservative estimate when other methods were not possible. In both the professor and MP groups the global method produced the most conservative; thus, suggesting a relationship between the relative distance and the conservativeness of the estimates.

This interaction between the position rated and the estimation procedure may provide some interesting advice to future pratitioners involved with utility analysis. The results of this study show the global procedure providing the largest estimates for jobs the respondents were most familiar with, their own. In contrast, positions in which respondents were less familiar with resulted in the lowest estimates being generated by the global method. Therefore, researchers and practitioners that are looking for conservative estimates of percentile point worth may employ different procedures depending on the familiarity of the raters with the rated position.

Limitations of the Study

The most limiting feature of the study is the homogeneity of the respondent group. All respondents in the study were students at Saint Mary's University in Halifax, Nova Scotia. One tact for future research might be a replication of the study on a different respondent group, perhaps a field study. This study could involve workers from many levels in an organization making estimates about their own performance as well as the performance of others at different organizational levels. It would be interesting to see if the results hold up.

Another limitation of the study is the number of positions for which respondents made estimates. The gap between the position could be considered great. Positions that do not differ as much in familiarity to the raters would help to isolate the points at which the distributions change shape.

A more powerful experimental design could be used in the study if the robustness of the ANOVA was assumed to be large. A mixed design in which the percentile point estimates act the as within subject variable and the position rated acts as the between subject variable could be utilized. In the case of a mixed model, however, it would be necessary to use a parametric statistical test in the analyses. Previous researchers have been reluctant to do this however. A thorough review of the literature shows all authors employing a non parametric test, the Wilcoxon Signed Ranks test, as the test of the normality assumption. That is, they have all used this non parametric test to evaluate the differences between SDy1 and SDy2, and in some cases SDy3 (cf. Schmidt et al., (1979); Bobko et al., (1983);Burke & Frederick, (1984); Reilly & Smither, (1985)).

Conclusion

The results of the study raise some disturbing questions about the state of SDy estimation. First, there seems to be an inordinate amount of variance within the judges estimates of a percentile point value. In the case of large samples of judges this variance would likely be cancelled out; however, in many situations just the opposite is likely to be true. That is, there is a good chance that the number of judges will be small and therefore this large variability will have a severe impact on the estimates. One promising avenue of research in this area is the use of feedback to reduce the variance. Group consensus techniques, such as the Delphi technique, have been shown to reduce variance. Future research should focus on ways to apply these techniques in organizational settings to produce the least variance results.

Second, one of the fundamental assumptions of the utility equation is not supported by the data in this study. The assumption of normality of the estimated performance distribution does not stand up under the scrutiny of a more rigorous test. The results do show however, that self estimates may produce the closest approximation to the normal curve. This result questions the validity of employing the global procedure to obtain SDy estimates. Future research may want to investigate the amalgamation of different methods. For example, a hybrid of the global method and superior equivalents technique may produce estimates that better fit the normal distribution.

It appears that judges do not believe the distribution of performance to be normal. This belief may influence their estimates away from one of the fundamental assumptions of utility analysis — that the underlying distributions are normal.

The was a complete lack of convergence found across methods. This has been the case in most studies that have compared the global method to those that are salary based. The results raise questions about the role of salary in making judgements and the validity of the global procedure.

References

Alexander, R.A. & Barrock, M.R. (1987). Estimating the standard error of projected dollar gains in utility analysis. *Journal of Applied Psychology, 72*, 475-479.

Alexander, R.A., Cronshaw, S.F., & Barrick, M.R. (1985). *Extending the managerial finance model of utility analysis to deal with uncertainty in parameter estimates.* Paper presented at the 1st Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, Ill.

Blum, M. & Naylor, J. (1968). *Industrial psychology: Its theoretical and social foundations.* New York: Harper & Row.

Bobko, P., Karren, R., & Kerkar, S.P. (1987). Systematic research needs for understanding supervisory-based estimates of SDy in utility analysis. *Organizational Behavior and Human Decision Making Processes, 40,* 69-95.

Bobko, P., Karren, R., & Parkington, J. (1983). Estimation of standard deviations in utility analyses: An empirical test. *Journal of Applied Psychology, 68,* 170-176.

Brogden, H.E. (1949). When testing pays off. *Personnel Psychology, 2,* 171-183.

Bruning, J.L. & Kintz, B.L. (1988). *Computational handbook of statistics,* Glenview, Illinois: Scott, Foresman and Company.

Burke, M.J. & Frederick, J.T. (1984). Two modified procedures for estimating standard deviations in utility analyses. *Journal of Applied Psychology, 69,* 482-489.

Cascio, W.F. (1982). *Costing human resiources: The financial impact of behavior in organizations.* Boston, MA: Kent.

Cascio, W.F. & Silbey, V. (1979). Utility of the selection centre as an assessment device. *Journal of Applied Psychology, 64,* 107-118.

Catano, V.M. (1988). *Utility analysis: A comprehensive review of methods for assessing the productivity improvement from human resource programs and their applicability in the Canadian forces* (Technical Supplement 88-3). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Cronbach, L. & Gleser, G. (1965). *Psychological tests and employment decisions.* Urbana, IL: University of Illinois Press.

Desimone, R.L., Alexander, R.A., & Cronshaw, S.F. (1986). Accuracy and reliability of SDy estimates in utility analysis. *Journal of Occupational Psychology, 59,* 93-102.

Eaton, N.K., Wing, H., & Lau, A. (1985). *Utility estimation in five enlisted occupations.* Proceedings of the 27th Annual Conference of the Military Testing Association (pp. 769-774). San Diego: Navy Personnel Research abd Development Centre.

Eaton, N.K., Wing, H., & Mitchell, K.J. (1985). Alternative methods of estimating the dollar value of performance. *Personnel Psychology, 38,* 27-40

Glenberg, A.M. (1988). *Learning from Data,* New York: Harcourt, Brace, Jovanovich

Greer, O.L. & Cascio, W.F. (1987). Is cost accounting the answer? Comparison of two behaviorally based methods for estimating the standard deviation of job performance in dollars with a cost-accounting-based approach. *Journal of Applied Psychology, 72,* 588-595.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1988). *Appled statistics for the behavioral sciences (2nd ed.),* Boston: Houghton Mifflin Company

Hull, C.L. (1928). *Aptitude testing.* Yonkers, NY: World Book.

Hunter, J.E. & Schmidt, F.L. (1982). *Fitting people to jobs: The impact of personnel selection on national productivity.* In M.D. Dunette & E.A. Fleischman (Eds.), Human performance and productivity: Vol 1: Human capability assessment (pp. 232-284). Hillsdale, N.J.: Erlbaum.

Johnson, V.W. (1989). *The estimation of the standard deviation of the dollar value of performance in a complex military occupation: A comparison of methods.* Unpublished Masters Thesis Saint Mary's University

Karren, R. & Bobko, P. (1983). *Conducting utility analysis: Some methodological concerns.* Paper presented at the 43rd annual meeting of the Academy of Management, Dallas, TX.

Katz, D. (1964). The motivational basis of organizational behavior. *Behavioral Science, 9,* 131-133.

Kelley, T.L. (1923). *Statistical method.* New York: MacMillan.

Keppel, G. (1982). *Design and analysis,* Englewood Cliffs, NJ: Prentice-Hall.

Mathieu, J.E. & Tannenbaum, S.I. (1985, March). *Supervisor's estimates of the dollar value of job performance: Some qualitative and quantitative findings.* In J.E. Mathieu (Chair) Estimating the utility of job performance: Examination and advancement. Symposium conducted at the meeting of the Southeastern Psychological Association, Atlanta, GA.

Mayer, R.S. (1982). *An evaluation of alternative methods of estimating the standard deviation of job performance to determine the utility of a test in a fixed-treatment sequential employee selection process.* Unpublished doctoral disseration, Wayne State University, Detroit, MI.

Orr, J.M., Sackett, P.R., & Mercer, M. (1989). The role of prescribed and unprescribed behavior in estimating the dollar value of performance. *Journal of Applied Psychology, 74,* 34-40.

Reilly, R.P. & Smither, J.W. (1985). An examination of two alternate techniques to estimate the standard deviation of job performance in dollars. *Journal of Applied Psychology, 40,* 651-661.

Roche, W.J. (1965). A dollar crtierion in fixed-treatment employee selection. In L. Cronbach & G. Gleser (Eds.), Psychological tests and personnel decisions (pp. 254-266). Urbana: University of Illinois Press.

Rossmeissel, P.G. (1984, June). ASVAB validity and improvements. ARI Research Highlights. Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Schmidt, F.L., Hunter, J.E., McKenzie, R., & Muldrow, T. (1979). The impact of valid selection procedures on work force productivity. *Journal of Applied Psychology, 64,* 609-626.

Schmidt, F.L., Hunter, J.E., & Pearlman, K., Hirch, H.R., Sackett, P.R., Schmitt, N., Tenopyr, M.L., Kehoe, J., & Zedeck, S. (1985). Forty questions about validity generalization and meta-analysis with commentaries. *Personnel Psychology, 38,* 697-798.

Schmidt, F.L., Hunter, J.E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on work force productivity. *Personnel Psychology, 35,* 333-347.

Schmidt, F.L., Hunter, J.E., Pearlman, K., & Shane, G.S. (1979). Further tests of the Schmidt-Hunter validity generalization model. *Personnel Psychology, 32,* 257-281.

Schmidt, F.L., Mack, M.J., & Hunter, J.E. (1984). Selection utility in the occupation of U.S. park ranger for three modes of test use. *Journal of Applied Psychology, 69,* 490-497.

Tannenbaum, S.I. & Dickinson, T.L. (1987, August). *Estimating SDy: The effectiveness of Delphi and critical incident methodologies.* American Psychological Association Annual Convention, NYC.

Taylor, H. & Russell, J. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23,* 565-578.

Weekly, J.A., Frank, B., O'Connor, E.J., & Peters, L.A. (1985), A comparison of three methods of estimating the standard deviation of performance in dollars. *Journal of Applied Psychology, 40,* 122-126.

Wroten, S.P. (1984, August). *Can supervisors really estimate SDy? In S.P. Wroten (Chair), Overcoming the fulitities of utility applications: Measures, models and management.* Panel discussion conducted at the meeting of the American Psychological Association, Toronto, On.

Appendix A

Questionnaire given to respondents in condition 1

Rating a job they had performed

Dear Participant;

My name is Mark Boyle and I am presently completing my Masters degree in psychology. The following surveys are intended to help us understand the way in which persons rate performance. They may provide information that could be useful to professors and administrators (eg. designing course evaluations); however, the study is primarily of a theoretical nature. I would very much appreciate your assistance in completing them; as your answers will also provide the data for my Master's thesis. The surveys have been reviewed by the Psychology Department's Ethical Committee and have received ethical approval.

If, after you complete the questionnaire, you have further questions concerning the study you may contact either me at 420-5862 or Dr. Grace Pretty in the Psychology department.

Your answers will be kept strictly confidential and when the data have been processed and the results are ready I will put a copy in the reserve section of the library under my name and the Psy.695.0 section.

Thank you for your cooperation and help.

I would first like you to consider a job or position you have held: one in which you are readily familiar with and have experience completing. This should be a position you have held for an extended period of time (i.e, a summer job, a part time job while attending high school or university, etc.). It is important to have a good grasp of the job as you will be asked to give some estimates of the worth of the job.

The job/position I am familiar with and will use to answer the following questions is

________________________________.

The average salary (per year) for this position is $_____________.

Note: Please be as specific as possible and write down the name of one position only.

Based on your experience with the position you have held, I would like you to estimate the yearly value to the organization of the products and services produced by the *average performer in the position you have held.* Consider the quality and quantity of output typical of the average performer.

Based on my experience, I estimate the value to the organization of the average performer at ___________ dollars per year.

I would now like for you to consider the superior performer. Let us define a *"superior" performer* as an employee who is at the 85th percentile. That is, his or her performance is better than that of 85% of his or her co-workers, and only 15% turn in better performance. Consider the quality and quantity of output typical of the superior performer.

Based on my experience, I estimate the value to the organization of the superior performer at ___________ dollars per year.

I would now like for you to consider the *"very superior" performer.* Let us define a very superior performer as an employee who is at the 97th percentile. That is, his or her performance is better than that of 97% of his or her co-workers, and only 3% turn in better performance. Consider the quality and quantity of output typical of the very superior performer.

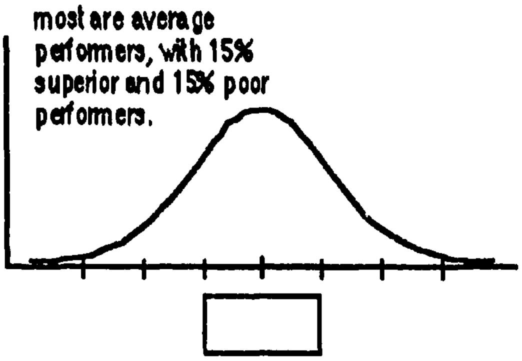
Based on my experience, I estimate the value to the organization of the very superior performer at ___________ dollars per year.

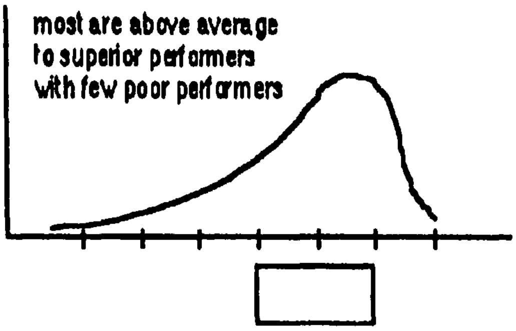
Finally, I would now like for you to consider the *"low performing"* employee. Let us define a low performing employee as an employee who is at the 15th percentile. That is, 85% of all co-workers turn in better performances than the low performing employee, and only 15% turn in worse performances. Consider the quality and quantity of output typical of the low performing employee.

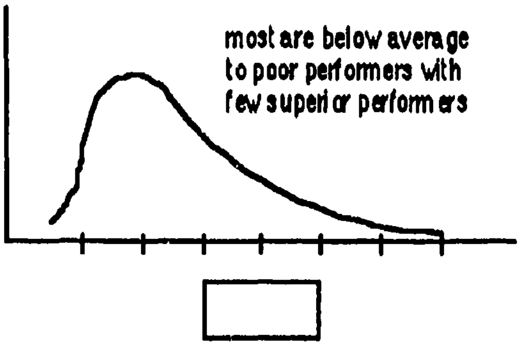
Based on my experience, I estimate the value to the organization of the low performing employee at ___________ dollars per year.

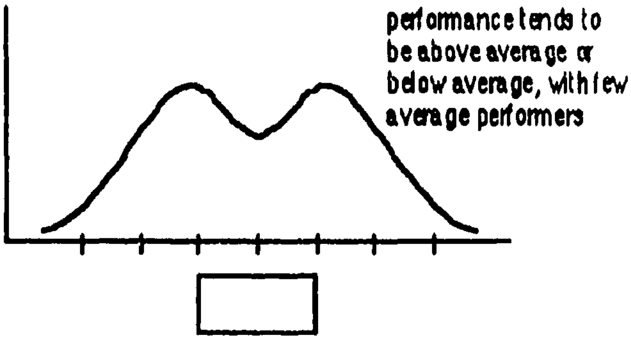
Please examine the following graphs. They represent a number of different performance distributions. In each case the horizontal axis represents performance level, moving from low performance on the left to high performance on the right. The vertical axis represents the frequency of a particular performance level.

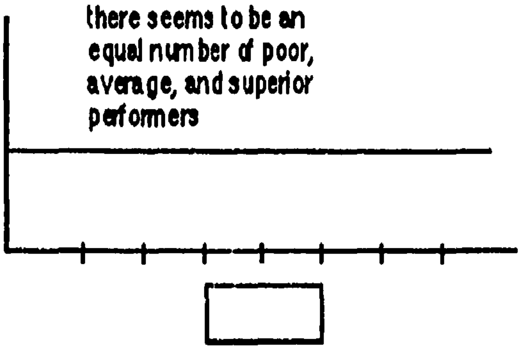
Please place a check mark in the box corresponding to the distribution you think best represents the distribution of performance in a job you have held.

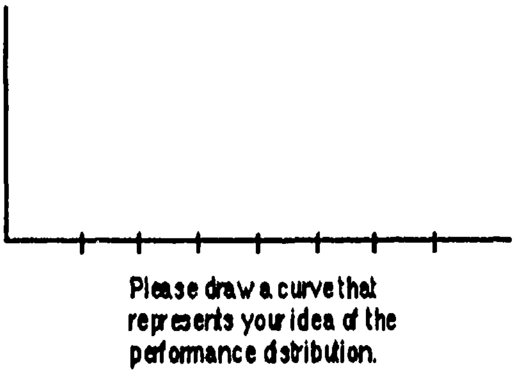












The first part of the questionnaire will have given you some insight into the difficulty involved in assigning dollar values to performance. In this part an alternative technique will be employed. This technique will allow you to rate performance from a relative frame of reference. For example, if a superior typist types 10 letters an hour and an average typist types 5 letters in the same time period then, all else being equal , 5 superior typists have the same value as 10 average typists. In this case, as before, consider superior to mean "better than 85% of co-workers".

Please answer the following question by filling in a number in the appropriate blank.

I estimate that, all else being equal, ______ superior performers are equal to 10 average performers.

For the following questions please provide the answer or check the appropriate space.

1. Please indicate your gender.
   Female _____ Male _____

2. What year of your studies are you in? ______

3. To what faculty do you belong?
   - Arts _____
   - Science _____
   - Commerce _____
   - Education _____
   - Engineering _____

4. Have you taken or are you taking a statistics course?
   yes _____ no _____

Appendix B

Questionnaire given to respondents in condition 2

Rating the position of University Professor

Dear Participant;

My name is Mark Boyle and I am presently completing my Masters degree in psychology. The following surveys are intended to help us understand the way in which persons rate performance. They may provide information that could be useful to professors and administrators (eg. designing course evaluations); however, the study is primarily of a theoretical nature. I would very much appreciate your assistance in completing them; as your answers will also provide the data for my Master's thesis. The surveys have been reviewed by the Psychology Department's Ethical Committee and have received ethical approval.

If, after you complete the questionnaire, you have further questions concerning the study you may contact either me at 420-5862 or Dr. Grace Pretty in the Psychology department.

Your answers will be kept strictly confidential and when the data have been processed and the results are ready I will put a copy in the reserve section of the library under my name and the Psy.695.0 section.

Thank you for your cooperation and help.

Below are the criteria used to evalaute the performance of a professor for the purpose of contract renewal, promotion, or tenure. Please read these criteria.

a. Quality and effectiveness as a teacher.
b. Academic credentials including degrees, special studies, and honors.
c. Scholarly and/or professional productivity or activity, including publications, other forms of reseach, consulting work, and work of a creative, cultural, or social significance.
d. Service on committees within the university; appropriate weight shall be given to such activity, although it is recognized that committee service is dependent either on appointment or election and is not necessarily under control of candidates for renewal, promotion, or tenure.
e. Other contributions to the university, including participation in its effective operation through academic advising, supervision of students, service as chairperson, director of division or program coordinator, and performance of other functions which have been traditionally accepted as part of the collegial character of the university.
f. Other contirbutions to the professional field and the community.

Considering these criteria I would now like you to complete the following questionnaire.

Based on the preceeding criteria and on your experience with university professors, I would like you to estimate the yearly value to your university of the products and services produced by the *average professor*. Consider the quality and quantity of output typical of the average professor.

Based on my experience, I estimate the value to my university of the average professor at ___________ dollars per year.

I would now like for you to consider the superior professor. Let us define a *"superior" professor* as a professor who is at the 85th percentile. That is, his or her performance is better than that of 85% of his or her collegues, and only 15% turn in better performance. Consider the quality and quantity of output typical of the superior professor.

Based on my experience, I estimate the value to my university of the superior professor at ___________ dollars per year.

I would now like for you to consider the *"very superior" professor*. Let us define a very superior professor as a professor who is at the 97th percentile. That is, his or her performance is better than that of 97% of his or her collegues, and only 3% turn in better performance. Consider the quality and quantity of output typical of the very superior professor.

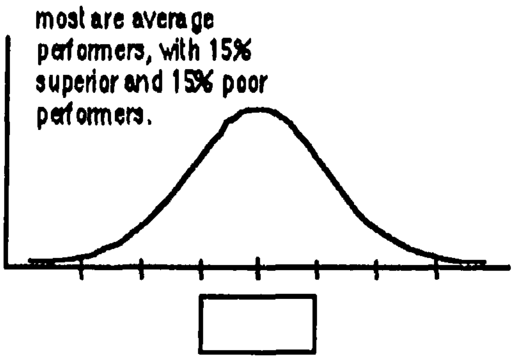
Based on my experience, I estimate the value to my university of the very superior professor at ___________ dollars per year.

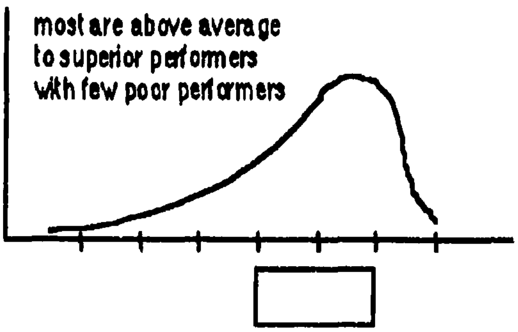
Finally, I would now like for you to consider the *"low performing"* professor. Let us define a low performing professor as a professor who is at the 15th percentile. That is, 85% of all professors turn in better performances than the low performing professor, and only 15% turn in worse performances. Consider the quality and quantity of output typical of the low performing professor.

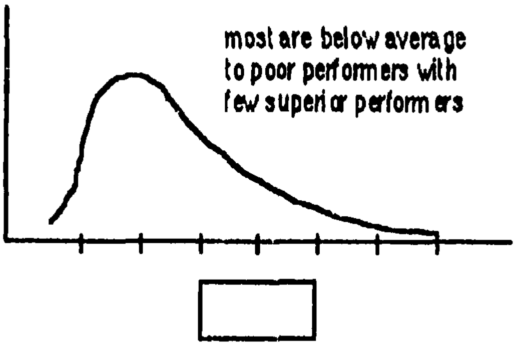
Based on my experience, I estimate the value to my university of the low performing professor at ___________ dollars per year.

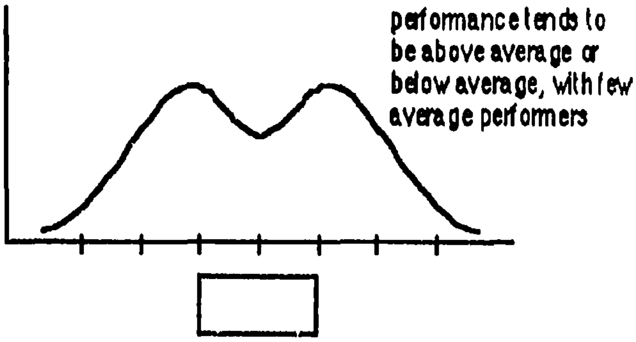
Please examine the following graphs. They represent a number of different performance distributions. In each case the horizontal axis represents performance level, moving from low performance on the left to high performance on the right. The vertical axis represents the frequency of a particular performance level.

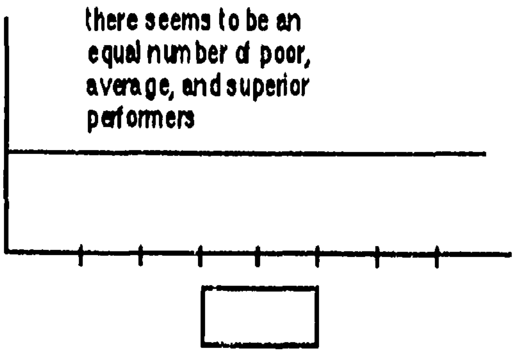
Please place a check mark in the box corresponding to the distribution you think best represents the distribution of performance in a job you have held.











Please draw a curve that represents your idea of the performance distribution.

The first part of the questionnaire will have given you some insight into the difficulty involved in assigning dollar values to performance. In this part an alternative technique will be employed. This technique will allow you to rate performance from a relative frame of reference. For example, if a superior typist types 10 letters an hour and an average typist types 5 letters in the same time period then, all else being equal , 5 superior typists have the same value as 10 average typists. In this case, as before, consider superior to mean "better than 85% of co-workers".

Please answer the following question by filling in a number in the appropriate blank.

I estimate that, all else being equal, ______ superior performers are equal to 10 average performers.

For the following questions please provide the answer or check the appropriate space.

1. Please indicate your gender.
   Female _____ Male _____

2. What year of your studies are you in? ______

3. To what faculty do you belong?
   Arts _____
   Science _____
   Commerce _____
   Education _____
   Engineering _____

4. Have you taken or are you taking a statistics course?
   yes _____ no _____

Appendix C

Questionnaire given to respondents in condition 3

Rating the position of Member of Parliament

Dear Participant;

My name is Mark Boyle and I am presently completing my Masters degree in psychology. The following surveys are intended to help us understand the way in which persons rate performance. They may provide information that could be useful to professors and administrators (eg. designing course evaluations); however, the study is primarily of a theoretical nature. I would very much appreciate your assistance in completing them; as your answers will also provide the data for my Master's thesis. The surveys have been reviewed by the Psychology Department's Ethical Committee and have received ethical approval.

If, after you complete the questionnaire, you have further questions concerning the study you may contact either me at 420-5862 or Dr. Grace Pretty in the Psychology department.

Your answers will be kept strictly confidential and when the data have been processed and the results are ready I will put a copy in the reserve section of the library under my name and the Psy.695.0 section.

Thank you for your cooperation and help.

Below is a description of the role and functions of Canadian Members of Parliament (MP). Please read this description to familiarize yourself with this position and its associated functions.

The Canadian MP can be seen as having four main functions. These functions occupy the majority of the MP's time.

The first function of an MP is that of representation. This function can be seen as involving three main parts. First, the MP represents the views of his/her constituents at the parliamentary level. Second, the MP is responsible for suggesting policy initiatives with regard to his/her constituents. Third, the MP functions in the representative role by acting as an "ombudsman" between constituents and government departments. That is, he/she will attempt to intervene on behalf of a constituent.

The second function of an MP is that of legislation. That is, it is the responsibility of the MP to help formulate and refine policy, influence departmental budgets, and actively speak out and question legislative matters.

The third function of the MP is that of surveillance. Surveillance in this sense means the MP must scrutinize both government spending and government activity. This is generally completed through questioning government decisions.

The fourth and final function of the MP is that of legitimation. Simply, it is the MP's role to act in such a manner as to legitimize the decisions of the government or opposition party. The MP carries out this role by acting in concert with the other members of his/her party.

Based on the preceeding description of an MP's role and on your knowledge of Members of Parliament, I would like you to estimate the yearly value to your area of the products and services produced by the *average MP.* Consider the quality and quantity of output typical of the average MP.

Based on my experience, I estimate the value to my area of the average MP at __________ dollars per year.

I would now like for you to consider the superior MP. Let us define a *"superior" MP* as an MP who is at the 85th percentile. That is, his or her performance is better than that of 85% of his or her collegues, and only 15% perform better. Consider the quality and quantity of output typical of the superior MP.

Based on my experience, I estimate the value to my area of the superior MP at __________ dollars per year.

I would now like for you to consider the *"very superior" MP.* Let us define a very superior MP as an MP who is at the 97th percentile. That is, his or her performance is better than that of 97% of his or her collegues, and only 3% turn in better performance. Consider the quality and quantity of output typical of the very superior MP.

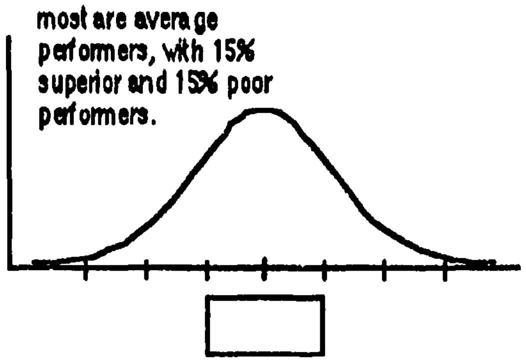
Based on my experience, I estimate the value to my area of the very superior MP at __________ dollars per year.

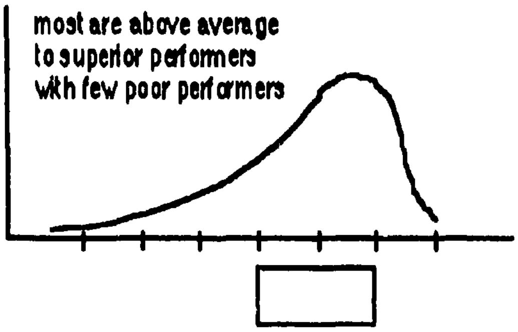
Finally, I would now like for you to consider the *"low performing"* MP. Let us define a low performing MP as an MP who is at the 15th percentile. That is, 85% of all MP turn in better performances than the low performing professor, and only 15% turn in worse performances. Consider the quality and quantity of output typical of the low performing MP.

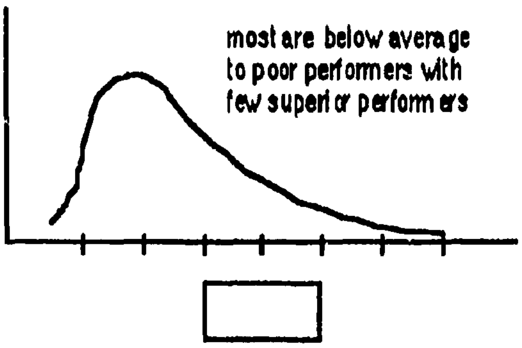
Based on my experience, I estimate the value to my area of the low performing MP at __________ dollars per year.

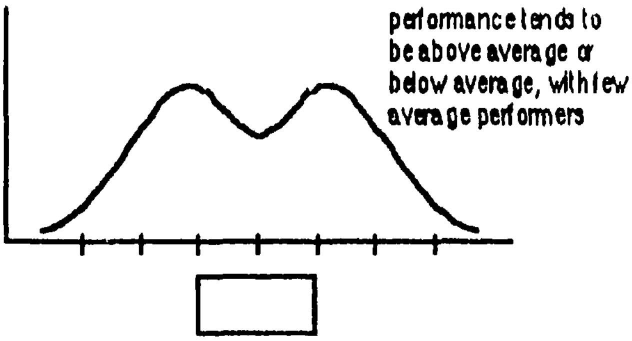
Please examine the following graphs. They represent a number of different performance distributions. In each case the horizontal axis represents performance level, moving from low performance on the left to high performance on the right. The vertical axis represents the frequency of a particular performance level.

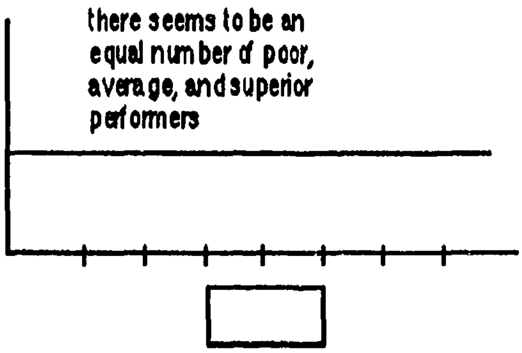
Please place a check mark in the box corresponding to the distribution you think best represents the distribution of performance in the position of MP.











Please draw a curve that represents your idea of the performance distribution.

The first part of the questionnaire will have given you some insight into the difficulty involved in assigning dollar values to performance. In this part an alternative technique will be employed. This technique will allow you to rate performance from a relative frame of reference. For example, if a superior typist types 10 letters an hour and an average typist types 5 letters in the same time period then, all else being equal , 5 superior typists have the same value as 10 average typists. In this case, as before, consider superior to mean "better than 85% of co-workers".

Please answer the following question by filling in a number in the appropriate blank.

I estimate that, all else being equal, _______ superior performers are equal to 10 average performers.

For the following questions please provide the answer or check the appropriate space.

1. Please indicate your gender.
   Female _____ Male _____

2. What year of your studies are you in? _____

3. To what faculty do you belong?
   Arts _____
   Science _____
   Commerce _____
   Education _____
   Engineering _____

4. Have you taken or are you taking a statistics course?
   yes _____ no _____

Appendix D

Computations for the Convergence Study

Superior Equivalents

$$SDy = V50[(N50/N85)-1]$$

where

V50 = the average salary for the position
N50 = the number of 50th percentile workers (fixed at 10)
N85 = the number of 85th percentile workers required to complete the work of ten 50th percentile workers

Self Estimates	Professor	Member of Parliament
V50 = 11134	V50 = 50010	V50 = 76500
N50 = 10	N50 = 10	N50 = 10
N85 = 5	N85 = 5	N85 = 4.6
SDy = V50[(N50/N85)-1]	SDy = V50[(N50/N85)-1]	SDy = V50[(N50/N85)-1]
= 11134[(10/5)-1]	= 50010[(10/5)-1]	= 76500[(10/4.6)-1]
= 11134[2-1]	= 50010[2-1]	= 76500[2.18-1]
= 11134[1]	= 50010[1]	= 76500[1.18]
= 11134	= 50010	= 88170

40% method

$$SDy = V50(.40)$$

SDy = V50(.40)	SDy = V50(.40)	SDy = V50(.40)
= 11134(.40)	= 50010(.40)	= 76500(.40)
= 4453.60	= 20004	= 30600

Global Method

SDy1 = 25810	SDy1 = 11177	SDy1 = 15113
SDy2 = 25703	SDy2 = 13256	SDy2 = 28568
SDy3 = 19179	SDy3 = 10161	SDy3 = 21132