THE ESTIMATION OF THE STANDARD DEVIATION OF THE DOLLAR VALUE OF PERFORMANCE IN A COMPLEX MILITARY OCCUPATION: A COMPARISON OF METHODS

V.W. Johnston

November 1988

SAINT MARY'S UNIVERSITY
HALIFAX, NOVA SCOTIA

ISBN 0-315-50918-X

Canada

THE ESTIMATION OF THE STANDARD DEVIATION OF THE DOLLAR VALUE OF PERFORMANCE IN A COMPLEX MILITARY OCCUPATION: A COMPARISON OF METHODS

SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR A DEGREE IN
MASTERS OF APPLIED SCIENCE (I/O PSYCHOLOGY)
SAINT MARY'S UNIVERSITY
HALIFAX, NOVA SCOTIA

APPROVED:

DR. V.M. CATANO

DR. J.K. CHADWICK-JONES

LCDR D.R. LAIRD

March 2, 1989
DATE

Abstract

THE ESTIMATION OF THE STANDARD DEVIATION OF THE DOLLAR VALUE OF PERFORMANCE IN A COMPLEX MILITARY OCCUPATION: A COMPARISON OF METHODS

V.W. Johnston

November 1988

A major impediment to the application of the Cronbach-Gleser model for estimating the utility of personnel selection programs, has been the difficulty encountered in accurately estimating the standard deviation of the dollar value of performance (SDy).

In this study, 206 Canadian naval officers estimated SDy for junior officers in a complex naval occupation using the procedure proposed by Schmidt, Hunter McKenzie and Muldrow (1979) or a modified procedure. In the modified procedure, the Schmidt et al. instructions were changed to provide judges with additional information regarding percentile point estimates, the order in which estimates were to be made, the context of the performance, and the dimensions being assessed. SDy estimates from the two procedures were compared and the results indicated that the modified procedure did not reduce between-judge

variance as predicted but did significantly affect the judges' perception of the underlying distribution of performance. It was also found that supervisory rank and experience significantly affecced the between-judge variance of the SDy estimates.

The estimates made using the Schmidt et al. and modified procedure showed little convergence with estimates made using the Superior Equivalence and 40% procedures. The latter procedures produced similar estimates which were higher than those made using the first two procedures.

The assumption of normally distributed performance was not supported in the study. This and other findings, indicate the need for further research in several areas before widespread use of any of the estimation procedures is adopted.

Acknowledgements

This study would not have been possible without the support of the officers of Maritime Command and the technical and professional support of the Commanding Officer and staff at the Canadian Forces Personnel Applied Research Unit, Willowdale. The following members of these groups deserve particular mention for their assistance, advice and encouragement:

Lieutenant Commander D.R. Laird

Captain J.A. James

Major C.D. Lamerson

Captain R.J. Angus

Lieutenant Commander M.N. Rodgers

I am especially grateful to Dr. Vic Catano who, throughout the study, provided academic support, excellent advice and timely and encouraging feedback.

Table of Contents

List of Tables

Page

Page

Introduction

Decision makers in the military, like those in business, have become increasingly concerned with the relationship between the costs of developing, conducting and evaluating personnel selection programs and the dollar-valued outcome of the programs. Although equations for calculating the utility of personnel selection programs have been available for some time (Brogden, 1949; Crobach & Gleser, 1965), there has, until recently, been little work in this area. Reviews of the history of the development of utility analysis (Catano, 1988; Cascio, 1982; Schmidt, Hunter, Mckenzie, & Muldow, 1979) suggest that a major obstacle to more widespread use of these equations has been the difficulty in estimating one of the key parameters: the standard deviation of job performance in dollars (SDy).

SDy

Early techniques for assessing the utility of selection tests (Taylor and Russell, 1939) compared the proportion of successful applicants selected using the new test to the proportion of applicants who would have been successful if the test had not been used. The Taylor-Russell model did not take into consideration the costs involved in developing and conducting the

test, nor did it consider the variation in the value of "successful" applicants. All successful applicants were assumed to be of equal value, whether their performance was superior or just met the criterion cut-off.

In 1949, Brogden developed an equation for assessing utility that expressed output in dollars. The Brogden model recognized that the performance of each of the applicants selected for a particular job was not of equal value to the organization. The utility of a selection procedure was, therefore, dependent not only on the cost of the selection procedure and its validity, but also on the variability of the dollar value of job performance or "SDy".

The Brogden model was extended by Cronbach and Gleser (1965). Their equation, which has been one of the most widely used methods for assessing utility analysis of selection programs, indicates that the utility of a selection procedure is a direct multiplicative function of SDy, the validity coefficient of the procedure and the mean standardized test score for those who are selected, or:

$$\Delta \bar{U} = \frac{\Delta U}{Ns} = rxy \quad SDy \; \bar{Z}x - CA \; / \varnothing$$

Where ΔU = the gain in utility per selectee over random selection.

rxy = the correlation of the test with the criterion.

Ns = the number of applicants selected using the test.

$\bar{Z}x$ = the mean standardized predictor score.

CA = the selection cost per applicant.

ϕ = the proportion of applicants above the predictor cutting point (selection ratio).

Given this equation, SDy directly affects the size of the potential benefit that could be derived from a new selection program. A large SDy can justify using a test of low validity. On the other hand, little variability in dollar-valued job performance among the applicant population (a small SDy) would not justify the costs of developing and using incrementally valid selection procedures.

Schmidt et al.(1979) offer the following example of how a test of low validity can have a higher utility than a test of high validity:

	rxy	$\bar{Z}x$	SDy	$\bar{U}$/ selectee
Mid level job (e.g., systems analyst)	.20	1.00	25,000	$5,000
Lower-level job (e.g., janitor)	.60	1.00	2,000	$1,200

Despite the low validity (.20) of the selection test for the mid level job, large variation in the

dollar value of performance (25,000) results in considerably larger savings per selectee than the test of high validity (.60) where variation in the value of job performance is low. This illustrates the critical role of SDy in utility analysis and the need for an accurate estimate of this parameter.

The Global Estimation Model

Until recently, it was generally accepted that the only way to estimate SDy was through costly and complicated cost accounting procedures in which the dollar value of the job performance of each employee was costed out and the standard deviation computed. These procedures entailed tremendous time and effort, while unclear and questionable methods involved "many estimates and arbitrary allocations" (Roche, 1965, p.263).

In 1979, Schmidt et al., proposed a procedure for obtaining a rational estimate of SDy. They 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". They argued that supervisors, who had the best opportunity to observe output differences on a

day-to-day basis, could be used to estimate the value of products and services produced by employees at different performance levels.

In a study of the utility of a computer programmer aptitude test, Schmidt et al. (1979) used a carefully developed questionnaire to ask 105 supervisors to estimate the yearly value to the organization of products and services produced by the low-performing (15th percentile), average performing (50th percentile) and the superior (85th percentile) computer programmer. In making their estimates, the supervisors were asked to consider what the cost would be of having an outside firm provide the same products and services. Estimates of SDy were calculated by finding the mean differences between estimates at the 15th and 50th percentile and estimates at 50th and 85th.

While recognizing that the procedure was subject to error, Schmidt et al. (1979) suggested that it was not critical that estimates of utility be accurate down to the last dollar. They pointed out that utility estimates are typically used for decisions about selection programs where only errors large enough to lead to incorrect decisions are of any consequence. They maintained that jobs at the higher levels of the

occupational hierarchy, where SDy values were largest and the utility of selection procedures potentially greatest, are handled least well by accounting methods. They also felt that, by referring dollar estimates to the cost of services by an outside consulting firm, they had provided a "relatively concrete standard" and, by averaging estimates across a large number of expert judges, they could control idiosyncratic tendencies, biases and random error.

The Schmidt et al. 'global estimation model' has since been used in at least seventeen utility studies involving various occupations, such as sales (Burke and Frederick, 1984; Burke and Frederick, 1986; Cascio and Silbey, 1979; Weekly, Frank, O'Connor and Peters, 1985; Reilly and Smithers, 1985, Greer and Cascio, 1987), financial services (Bobko, Karren and Parkington, 1983; DeSimone, Alexander and Cronshaw, 1986; Hunter and Schmidt, 1982; Mayer, 1982; Mathieu and Tannenbaum, 1985), law enforcement (Karren and Bobko, 1983; Schmidt, Mack and Hunter, 1984), nursing (Tannenbaum and Dickinson, 1987) and military occupations (Eaton, Wing and Lau, 1985; Eaton, Wing and Mitchell, 1985; Rossmeissel, 1984).

The results of the studies, most of which were aimed at evaluating the estimation procedures, can best

be described as equivocal. In those studies where comparisons could be made with objective measures of the value of job performance, such as sales performance, the global procedure produced good estimates in two studies (Bobko et al., 1983; Greer and Cascio, 1987) and poor estimates in two others (Mayer, 1982; Weekley et al., 1985).

In most of the global estimation studies, there was substantial variation across judges within each set of estimates for a particular percentile (Mayer, 1982; Schmidt, Hunter & Pearlman, 1982; Bobko, Karren & Parkington, 1983; Burke and Frederick 1984; Eaton, Wing & Mitchell, 1985; Reilly and Smithers, 1985; Weekly et al., 1985; Greer and Cascio, 1987). In at least six of the studies, the standard deviation of the estimates was found to be as large or greater than SDy (Schmidt et al., 1979; Bobko et al, 1983; Burke and Frederick 1984; Reilly and Smithers, 1985; Weekly et al., 1985; Greer and Cascio, 1987).

Many of the researchers concluded that the extreme variability between judges demonstrates the difficulty in making judgements about employee worth, particularly in situations, such as the military, where the cost of contracting for services is unknown and where the criteria of successful performance is

subjective and poorly defined (Bobko et al., 1983; Burke and Frederick, 1984; Eaton et al, 1983; Mayer, 1982; Reilly & Smithers, 1983). It appeared that judges using global estimation techniques were using very different scales and/or referring to different dimensions in making their estimates.

In an effort to reduce the variability between judges, Burke and Frederick (1984) modified the Schmidt et al. (1979) Global Estimation procedure by feeding back to manager/judges the mean estimated value for the 50th percentile before asking them to make the other percentile judgements. This sequential procedure, which was originally proposed by Bobko et al.(1983), effectively reduced the percentile point variation in the Burke and Frederick study, but did not reduce variation in similar studies by Karren and Bobko (1983), and Wroten (1984)

The 40% method

In their research with SDy estimation, Schmidt Hunter and Pearlman (1982) found that resulting estimates, when expressed as a percentage of salary, typically fell between 40 and 70%. They recommended that, as a rule of thumb, the round lower bound estimate of 40% could be used as a conservative estimate of SDy when time or resources did not permit

the global estimation of SDy. There appears, however, to be little theoretical or empirical support for using the 40% estimate (Catano, 1988). Subsequent global estimates of SDy have ranged from 19% (DeSimone et al 1986) to 133% (Reilly and Smithers, 1985) of salary.

The CREPID procedure

An alternative method for estimating SDy was proposed by Cascio and Ramos (1986). The Cascio-Ramos Estimate of Performance In Dollars (CREPID) procedure relies directly on salary and is based on the assumption that the value of a commodity is its market price. The value of an employee's labour, therefore, is equal to what an organization is willing to pay for it.

The CREPID procedure involves a job analysis phase and a performance appraisal phase and is carried out in eight separate steps:

1. Based on the job analysis, jobs are broken down into principal activities that encompass at least 10% of total performance over a one year period. The supervisor verifies the accuracy of these prinicipal activities but is not involved in providing data for the job analysis;

2. The supervisor rates each principal activity in terms of time/frequency, importance, consequence of error and level of difficulty. The "time/frequency" dimension is rated on a 0-100 scale with each principle activity rated as a

percentage of the total so that rating for all principal activities will equal 100% . The other three dimensions are rated on a 0-7 scale;

3. Each principal activity is assigned a relative weight. This is achieved by multiplying together the numerical rating for time/frequency, importance, consequence of error and level of difficulty for each principal activity and dividing the overall rating for each activity by the grand total;

4. A dollar value is assigned to each principal activity by allocating a proportional share of the employee's salary to each activity based on the relative weight calculated above;

5. The supervisor rates the performance of each employee on each principal activity using a 0-200 point scale. This modified magnitude estimation procedure results in a rectangular distribution of ratings;

6. The performance rating for each activity is multiplied by the dollar value of the activity, thus weighting the economic value of each activity by the individuals performance score;

7. The overall dollar value of each employee's job performance is computed by adding the weighted activity values from step 6 for each individual; and

8. The mean and standard deviation of dollar-valued job performance is computed.

Unlike the global estimation model, supervisors using the CREPID procedure do not have to estimate job performance in dollars, but simply judge performance, which is a typical part of their duties.

Three studies have compared the CREPID procedure to other estimation methods (Weekly et al., 1985; Reilly and Smithers, 1985; Greer and Cascio, 1987).

In all three studies, the CREPID procedure resulted in smaller estimates than the global procedure. Greer and Cascio (1987), also found that the CREPID procedure resulted in much smaller estimates than a cost-accounting method, while Reilly and Smithers (1985) found that the CREPID procedure provided conservative estimates compared to objective sales data. Global estimates in the Reilly and Smithers study were more consistent with the objective sales data; however, as performance information became more difficult to convert to dollar terms, global estimates became less accurate and more variable.

In a study of store managers, Weekly, et al. (1985) found that the CREPID method and the 40% method produced comparable results that differed "markedly" from those produced by the global estimation model. Similar results were found in a recent study by Edwards, Frederick and Burke (1988). The more conservative estimates of SDy using the CREPID procedure may result from the fact that the estimates are directly linked to salaries which, in the US, have been calculated at about 57% of output.

The Edwards et al. (1988) study examined the use of organizational archival data in place of the data specifically obtained for CREPID. They compared the normal CREPID method with three modified methods using

archival performance evaluations, job analysis ratings or both evaluations and ratings. The estimates obtained from the modified methods converged with those of the normal CREPID and the 40% method at a level nearly one-fifth of that for the global procedure. Despite this convergence, Edwards et al. expressed concern about the use of archival performance evaluations as they were contaminated by their use in merit compensation decisions, the requirement of feeding back the evaluations to subordinates, and an emphasis on employee development.

Non-dollar Estimation Techniques

While conducting research with Army Tank Commanders, Eaton et al. (1985) found that 12% of their sample refused to provide dollar estimates for average and superior performers. They objected on the grounds that soldiers' lives and combat activities were not describable in dollars. Eaton et al. proposed two methods for obtaining non-dollar estimates of SDy in situations where contracting out is not possible and where supervisors are far more accustomed to thinking about the value of operational output rather than dollar value. The first method, the Superior Equivalents Technique, requires the supervisors to estimate how many superior (85th percentile) performers

would be needed to produce the output of a fixed number of average (50th percentile) performers. The performance estimates are then transformed to SDy estimates, based on the value of average performance.

The second method, the Systems Effectiveness Technique, is based on the concept of a 'system' comprised of performing units, all of which contribute to the total aggregate performance. Improved total system performance can be obtained either through improved unit performance with existing numbers of units or by increasing the number of units with the same performance. Consequently, the value of improved unit performance in obtaining higher aggregate performance is equal to the cost of the increased number of units that would be needed to obtain the same higher level of aggregate performance. The SDy in dollars then, equals the cost per unit times the ratio of the non-dollar standard deviation of performance to the initial mean level of performance.

On testing both of their techniques, Eaton et al. concluded that they would be useful in providing estimates which bracket true utility values. Catano (1988) suggested that the systems effectiveness procedure is based on two assumptions which may not hold true for many military situations. The procedure

assumes that the performance of a unit is attributed primarily to the performance of the individual in the job under investigation. It is also assumed that all units are at the same initial level of performance and that all reach a new level at the end of the treatment.

While the Systems Effectivness procedure may work reasonably well with small, well defined "systems" such as tank crews, it would not be applicable to more complex, interacting systems such as ship's departments or aircraft maintenance sections, where system effectiveness is dependent upon the performance of personnel in various occupations and upon numerous external factors. Similarly, the Superior Equivalents Technique is easily applied where individual contributions are clearly defined and superior performance can be translated into numbers of average performers. At higher levels of the organizational hierarchy, however, this translation is much more difficult, as jobs become more complex and poor performance cannot be offset by increasing the number of managers/supervisors.

In discussing their research, Eaton et al. point out the need to question whether, and how, qualitative variables and multidimensional constructs are being transformed into unitary quantitative indices. For

many occupations, such as military ones, performance is not easily translated into dollar terms. As this translation becomes more difficult SDy estimates appear to be less accurate and more variable (Reilly and Smithers, 1985).

Estimating SDy in a Complex Naval Officer Occupation

Estimating the dollar value of job performance in the military presents a special challenge. The global estimation model is likely to result in unusually high between-judge variation in the estimates because the worth of military members is not easily assessed in terms of output, and the cost of contracting out of services is generally not available. Procedures such as the CREPID and the 40% method may provide misleading results because military salaries do not normally reflect the market value of a particular military occupation but are based on occupational groupings and rank levels. The non-dollar estimation techniques proposed by Eaton et al. (1985) may provide accurate SDy estimates in selected military situations but are not likely to be useful for complex military occupations that are part of an interacting system such as a ships crew. Contextual factors, such as a wartime vs peacetime scenario, can also have a significant

effect on both the perceived (and real) worth of the servicemember (Sadaaca and Campbell, 1985).

Until researchers can understand, and control, the underlying components which lead to human judgements about worth, the estimation of SDy in military occupations will be difficult and global estimates of the dollar value of performance will be subject to a relatively high degree of variation across judges. It should, however, be possible to reduce some of the variation by controlling those factors which past research has already led us to suspect are contributing to it. One aim of this study is to investigate methods of reducing the between-judge variance in the estimation of the dollar value of job performance in a complex military job, specifically the job of a Sub-Lieutenant in the Maritime Surface and Sub-surface (MARS) occupation in the Canadian Armed Forces.

The Sub-Lieutenant MARS Officer

Canadian Forces MARS officers are carefully selected and highly trained. Officer candidates must have at least a high school graduation diploma and achieve a score at the 80th percentile (of the military applicant population) in a test of general learning ability. They enter the MARS occupation through

several different programs; some through subsidized university or military college programs; some enter directly with or without a college degree and some are commissioned from the ranks.

All of the MARS officer applicant entering from outside the military are interviewed by a recruiting officer to assess their military and leadership potential. If they meet officer entrance requirements, they attend a multiple assessment board which includes: leadership tasks in a group context; an in-basket exercise; a file review of biographical and test information; an interview by a board of senior officers; and two leaderless group discussion exercises.

Successful MARS applicants complete a 13 week basic officer training course (BOTC), where they are assessed on leadership ability, communication, decision-making and presence of command. Following BOTC they undergo six months of extensive academic and practical training in the primary aspects of seamanship and navigation. This is followed by a six month Naval Operations Course (NOC) which introduces them to the operational and administrative aspects of the Naval environment.

On successful completion of the NOC training, officers who will serve on surface ships, proceed to operational Destroyers for a 10 month on-the-job training period where they qualify in bridge watchkeeping. After receiving their watchkeeping certificate, they will normally attend a Destroyer 'D' level course in a specialty area of Weapons/Electronic Warfare, Navigation or Anti-submarine Warfare Air Control. It is not until they have completed the 'D' level training, that the officers are fully trained to perform their duties on a ship and, it is only then, that the variance in the value of performance between officers can best be assessed.

The annual salary of a Sub-Lieutenant will vary depending on length of service and method of entry into the occupation. The rounded average salary of a fully trained Sub-Lieutenant collecting a sea duty allowance is $30,000.

The MARS Sub-Lieutenant (SLt) is typical of the groups for which within-cell variation of dollar value estimates of job performance is high. Their job performance is very difficult to convert into dollar value. They are employed in a wide spectrum of activities related to the operation of naval ships, naval weapons systems and combat information systems,

and their failure to perform could predjudice the safety or success of naval operations and possibly lead to the loss of life or damage to valuable equipment.

Other factors which research has indicated may contribute to a high degree of between-judge variation in global estimates of the dollar value of a MARS SLt's job performance include the following:

1. The supervisor's rank and experience - The Sub-Lieutenant works within a clearly defined rank structure. He is generally supervised directly by a senior Lieutenant or Lieutenant Commander with a Commander at the second level of supervision.

Mayer (1982) found substantial differences in the standard deviation of point estimates by supervisor/judges at different organizational levels when estimating the worth of bank tellers. The SDy estimates by branch/district managers came closer to accounting estimates than did estimates by the tellers' immediate supervisors. Unlike military officers though, a bank teller's performance can be relatively easily assessed against rigid control systems. Reilly and Smithers (1985) suggested that, "it may be that

experienced supervisors can more accurately translate the performance of their employees into dollars even under complex conditions" (p.660). Additional research is required to determine whether military rank or experience is a factor in determining who will be the most appropriate judges.

2. Contextual factors - wartime vs peacetime scenarios -Sadaaca and Campbell (1985) found that the judged worth of military occupations changed when a wartime or peacetime scenario was used. Bobko et al.(1983) suggest that SDy may not be a static parameter and that judgements may well be different in a hostile as opposed to a benign organizational environment. In the Eaton et al. (1985) study, where Tank Commander's were relunctant to estimate worth and estimates were highly varied, supervisors were asked to estimate the value of performance "in combat".

Although military personnel are ultimately selected for their ability to perform in wartime, the dollar utility of selection programs is generally not an issue during wartime. It is during the periods of budgetary restraint in peacetime that the cost/benefit of personnel

programs is of most concern to funding authorities and it is against peacetime manning levels that benefits are being measured. Regardless of which scenario is used, it should be specified in the instructions to judges.

Although very few Canadian Naval Officer's supervisors have witnessed performance during combat, they may base their estimates of the value of job performance on a hypothetical combat scenario. They will need to be reminded that they are making estimates based on their own experience in a peacetime force.

3. Supervisors's interpretation of percentile points and perception of underlying distributions - In reviewing the research on SDy estimation, Bobko et al. (1987), stated that there is little understanding of how judges cognitively process the meaning of the 15th percentile or the 85th percentile. At least two studies (Bobko et al. 1983; Kerren and Bobko 1983), reported that over 20% of the supervisors provided inconsistent judgements in the percentile estimates. Bobko et al.(1983) suggest that judges may be using a uniform (rectangular) rather than normal (bell shaped) distribution. Schmidt, Mack and Hunter

(1984) suggested that although job performance may be normally distributed, supervisors may be more cognizant of variation at the lower end of the spectrum because thay have developed a mental set focused on avoidance of errors on the part of low performers rather than attending to outstanding performance.

In their study of U.S. Marshalls, Karren and Bobko (1984) found that because of high selection ratios and intensive post-selection training, supervisors felt that "If they got this far, they're all outstanding", and they tended to equate 50th and 85th percentile performance, while giving very low values to the 50th percentile performance (Bobko et al. 1987).

Like the U.S. Marshalls, Sub-Lieutenant MARS officers are carefully selected and extensively trained. While supervisors may experience difficulty in making percentile point estimates of their value, skewed distribution and within-cell variation in estimates might be reduced by providing clearer instructions regarding percentile estimates and a diagram of percentile points in a normal distribution.

4. Order effect - The order in which percentiles are estimated may also contribute to between judge variance in estimates. In their research, Burke and Frederick (1984) discovered that, despite instructions on the order in which to estimate percentiles, several judges used a different order (eg. estimated the 15th percentile before the 50th). They suggested that the different ordering of percentile estimates may have accounted for some of the large within-column variances in their study. In a more recent study, Schetzner and Bobko (1986) presented subjects with different orderings and found significant differences in SDy estimates. Clearer and more specific instruction on the ordering of estimates may further reduce between-judge variance.

5. Cognitive dimensions - "In order to understand judges' estimates of overall worth, it is critical to unravel the dimensions from which such judgements might be derived" (Bobko et al., 1987). Studies in which supervisors were asked about the factors they included in making their estimates revealed that numerous dimensions accounted for supervisors' qualitative perceptions

(Burke and Frederick, 1984; Mathieu and Tannenbaum, 1985).

Although supervisors of Sub-Lieutenant MARS officers prepare annual performance assessment reports on those officers which they supervise and should make value estimates based on similar dimensions, they may in fact select certain critical factors on which to base their estimate. It may be possible to reduce variation between judges by ensuring that they are working within similar dimensions when making their judgements, by having judges weight the same principal job activities immediately prior to making their estimates.

Summary

One of the major obstacles to the calculation of the utility of selection devices appears to have been overcome by the development of procedures for estimating the standard deviation of the dollar value of job performance or SDy. The most widely used procedure for estimating SDy is that proposed by Schmidt et al. (1979). Evaluations of this 'global estimation model' have had mixed results and large

variation in the estimates between judges have raised concerns about its accuracy.

Two other estimation procedures, the CREPID (Casio and Ramos, 1986) and 40% method (Schmidt et al., 1982), are tied directly to salary and may be of limited use in a military context where salary is fixed regardless of performance and may not reflect the real value of employees to the organization. Non-dollar estimation methods proposed by Eaton et al. (1985) should be useful in estimating the value of performance in certain well defined military jobs but would be less useful as the jobs become more complex.

Factors which research has indicated may contribute to between-judge variation in estimates of the dollar value of job performance in a complex military occupation include :

1. the supervisor's rank and experience
2. contextual factors
3. the supervisor's perception and interpretation of the distribution of performance
4. order effect
5. cognitive dimensions

By modifying the instructions for the Schmidt et al. global estimation procedure it should be possible to

reduce the effect of some of these factors and thereby reduce the between-judge variance in estimates.

Purpose of the Thesis

The purpose of this research project is to determine if between-judge variation in estimates of the dollar value of performance in a complex military officer occupation, can be reduced by modifying the instructions to supervisors in the Schmidt et al. (1979) global estimation model and by controlling for the rank and/or experience of judges. The manner in which judges perceive the distribution of job performance will also be examined.

A subsidiary aim of the study is to compare the variance in judges' estimates made using the above methods, with estimates acquired using methods similar to CREPID and the Superior Equivalents Technique. Convergence of estimates from each of these methods and the 40% method will also be investigated.

Hypotheses

1. The between-judge variance in point estimates of the dollar value of performance can be significantly reduced by providing the judges/supervisors with more detailed information regarding the point at which the

estimate is required; the order in which estimates should be made; the context of the performance and, the performance activities being assessed.

2. The between-judge variance in point estimates of the dollar value of performance will be significantly less for the more experienced supervisors.

3. The between-judge variance in point estimates of the dollar value of performance will be significantly different for different rank groups.

4. There is no theoretical support for an hypothesis related to judges' perceptions of percentile points and underlying distributions. The intent here is to simply investigate those perceptions. It is also not intended that this project provide an empirical test of the subsidiary issues related to the CREPID and Superior Equivalents Technique. The collection of data in investigating the above hypothesis, does, however, provide a unique opportunity to informally examine the variation and convergence of estimates using various methods.

Method

Estimates of SDy were provided by MARS officers of Lieutenant (Lt), Lieutenant Commander (LCdr) and

Commander (Cdr) rank, who responded to one of two questionnaires. The questionnaires were mailed to 322 officers who had been identified by the Directorate of Personnel Information Systems (DPIS), at National Defence Headquarters, as having five or more years service and currently serving on ships or units on the East Coast.

Participants were randomly divided into two groups, each containing approximately the same number of officers at each rank level. The first group, which served as a control, received Questionnaire 'A' which asked them to estimate SDy using the Global Estimation Model. The second group received Questionnaire 'B' which contained modified instructions and additional information regarding percentile estimates. In addition, both questionnaires required the participants to make Superior Equivalents estimates. The second group also provided information that was used to approximate a CREPID procedure. The questionnaires are described in greater detail below.

The Questionnaires

Questionnaire 'A' (see Appendix A) provided the control group estimates of SDy. Apart from the addition of an introductory paragraph and changes of job titles, Part I of the questionnaire was identical to the

Schmidt et al.(1979) questionnaire. The introductory paragraph was added to explain that estimates were required in order to assess a new method of placing dollar values on performance, which would replace burdensome cost accounting procedures. Respondents were asked to make percentile point estimates of the dollar value of performance of "average" (50th percentile), "superior" (85th percentile) and "low performing" (15th percentile) Sub-Lieutenant ('D' level) MARS officers. The ('D' level) qualifier was added because some SLts would have just completed training and would be of less value to their ship than more experienced SLts. By specifying "D level", SLts of relatively equal levels of experience would be considered and estimates would be based on job performance and not the quality of training.

In order to assess the convergence of estimates acquired using different methods, Part II of questionnaire 'A' required respondents to make a "superior equivalents" estimate using the Eaton et al (1985) procedure. They were asked to estimate (without referring to their previous dollar estimates in Part I) the number of superior SLts that would equal 10 average SLts.

In an effort to determine the supervisor's perception of the distribution of performance among MARS SLts, Part III of the questionnaire asked respondents to select from five diagrams (normal, -skew, +skew, bimodal and rectangular) the one which in their opinion best reflected the distribution of job performance among SLt MARS officers. A space was also provided for respondents to illustrate their perception of the distribution of performance if it differed from the above diagrams.

Questionnaire B (see Appendix A) contained the experimental version of the Global Estimation procedure. In an attempt to reduce the between-judge variance in percentile point estimates, the Schmidt et al. procedure was modified by:

1. providing a clearer explanation of "percentile points" and a diagram of a normal curve indicating points at which estimates were to be made;

2. stating that the estimates were to be based on job performance in peacetime;

3. providing clear instruction regarding the order in which the point estimates were to be made; and

4. listing five principle activities (duty areas) of the MARS occupation, instructing respondents to rate

them according to the percentage of time that a typical SLt spends performing them and to rank them according to their individual contribution to the overall worth of a MARS officer. The respondents were then directed to consider these activities when making their estimates of the value of job performance. The principle activities (PA), which were acquired from published occupational specifications based on occupational analysis are:

1. Performs the duties of officer of the day in harbour;
2. Performs the duties of officer of the watch at sea;
3. Performs general/secondary duties as a ships officer;
4. Performs the duties assigned in the action organization of a ship;
5. Performs the duties of a divisional officer.

In Questionnaire B, subject were also required to make "Superior Equivalents" estimates and to indicate their view of the distribution of performance.

Because of the requirement for the direct assessment of individual performance, the CREPID (Cascio, 1982) procedure could not be replicated in

this study. Some of the respondents in the study were not currently employed in the direct supervision of SLts and privacy of information legislation precluded the use of archival performance evaluation data. In order to obtain a similar measure so that estimates could be compared with the other estimation methods, respondents were asked in Part III of questionnaire B, to rate the performance of a "typical" SLt MARS officers on each of the five principal activities. These ratings were converted to estimates of the dollar value of average performance using the CREPID procedure described above.

The final portion of both questionnaires A and B included questions about the rank and supervisory experience of the respondents.

A trial administration of the questionnaires, prior to the mail-out, revealed that respondents had no difficulty understanding or following the instruction. As a result of comments following the trial, two changes were made to the questionnaire. It was decided that the instruction should specify that estimates of performance be on "D' level SLts and, that in the CREPID procedure, respondents should rate "typical" rather than "average" SLts, as the latter implied an average rating.

Respondents

Two hundred and six of the officers responded to the questionnaires for a return rate of 64%. These responses represented more than 30% of all MARS officers in supervisory positions within the Canadian Forces. The ratio of Lts to LCdrs and Cdrs in the sample was 121:62:21. The ratio of these ranks in the MARS occupation was approximately 20:8:3. The apparent under-representation of the Lt rank resulted from the fact that only officers with five or more years service were targetted. The rank-to-rank ratio in the target population of MARS officers "with supervisory experience" is likely to be more similar to the respondent population.

The mean level of supervisory experience among respondents was 6.2 years, with Lieutenants averaging 3 years. There was a strong correlation (r=.74) between rank and supervisory experience. All but 14 of the respondents reported that they had experience in supervising junior MARS officers.

The data from two of the completed questionnaires, one of each type, were not used because the responses (estimates of the value of performance) were well outside the rest of the distribution (eg. $500,000) and

clearly met the accepted criterion for outliers, being more than three standard deviations from the mean.

Table 1 provides a breakdown of the respondents by rank and years of supervisory experience for each questionnaire type, while Table 2 indicates years of supervisory experience by rank.

Procedures

Mail-out questionnaires (Appendix A) were chosen as the method of collecting estimates because of the difficulty involved in obtaining the data directly from officers aboard the various ships and because the questionnaire is an accepted and effective method of gathering information within the Canadian Forces.

In order to mail the questionnaires to the military population, the research project required military sponsorship and the approval of both the National Defence Headquarters and Maritime Command Headquarters. Maritime Command Headquarters agreed to sponsor the project as part of an ongoing Naval Officer production research program.

Table 1.

Number of Respondents to Each Questionnaire Type by Rank and Years of Supervisory Experience

	Rank			Supervisory Experience			
	Cdr	LCdr	Lt	0-2	3-5	6 or more	Total
Questionnaire A	11	28	56	31	32	32	95
Questionnaire B	10	34	65	29	34	46	109
TOTAL	21	62	121	60	66	78	204

Table 2.

Years of Supervisory Experience By Rank

	0-2 years	3-5 years	6 or more years
Commander	0	1	20
Lieutenant Commander	3	18	41
Lieutenant	57	47	17

Three hundred and twenty-two questionnaires (161 of each type) were mailed individually to officers serving at Canadian Forces Base Halifax and aboard seventeen Naval ships. All of the questionnaires were mailed on the same date along with a covering letter (Appendix A) and a postage paid, self-addressed return envelope. One day prior to the mail-out a message (Appendix A) was sent to all participating ships and units outlining the purpose of the research and advising that participation was voluntary and anonymous. The covering letter, which contained similar information, was signed by the Commanding Officer of the Personnel Applied Research Unit in Willowdale, Ontario, where the author was employed.

Analysis and Design

Analysis of the data was completed at the Canadian Forces Personnel Applied Research Unit using the Statistical Package for the Social Sciences - Extended (SPSS-X) version 2.1. on the VAX system at York University in Toronto.

Comparison of the Schmidt et al. and modified method. For both the Schmidt et al. and the modified SDy estimation procedures, the means, variances and

standard deviation of estimates of the dollar value of a Sub-Lieutenant's performance were calculated at the 15th, 50th and 85th percentile point. Two estimates of SDy were computed by averaging the differences between the 15th and 50th percentile estimates (SDy1) and the differences between the 50th and 85th percentile estimates (SDy2).

Hypothesis 1. was tested by computing the significance of the differences in the variances, or F ratio, between the estimates from the Schmidt et al. and the modified method, at each percentile point (Ferguson, 1979, p.164).

Effects of rank and supervisory experience

Hypothesis 2 and 3 were tested by comparing the variances of estimates at each percentile point for each rank and level of supervisory experience using the Cochrane C and Bartlett-Box F tests (Winer, 1971, p.205). For analysis purposes, the respondents' reported years of supervisory experience were grouped into three levels, 0-2 years, 3-5 years and 6 years or more, based on their near equal distribution in the sample. Part years were rounded to the nearest year.

The effects of the estimation method, rank and supervisory experience on each of the percentile

estimate means were examined using two 2 x 3 ANOVAs (estimation method x rank; and estimation method x supervisory experience) for each set of percentile point estimates.

Perceived distribution of performance. The differences between dollar value estimates of performance at the 15th and 50th and the 50th and 85th percentile were calulated for each respondent and the Wilcoxon matched pairs signed rank test (Siegel, 1956 p.75) was used to test the assumption of normality of the distribution of the estimates for the entire sample and separately for the global and modified global procedures.

Frequencies were calculated for respondents' selection of the diagram which best reflected the distribution of performance. Chi-square analyses were used to determine if there was a significant difference in the frequencies of responses by rank, supervisory experience or estimation method.

Superior Equivalents. Frequencies, medians and modes were calculated for the respondents' estimates of Superior Equivalents. The "representative value of central tendency" for superior equivalent estimates was selected and the value of the superior SLt was

determined by multiplying the average salary of $30,000 by the ratio of superior to average SLts. The difference in value between the average and superior SLt was the estimate of SDy (Eaton et al.,1985).

CREPID procedure. To assess the CREPID procedure, the weights that were assigned to each of the five principle activities (PA), were multiplied by the rankings of importance of the PAs to provide an overall weight. These total weightings for each PA were then divided by the grand total to obtain a relative weight. Proportional shares of the average salary were allocated to each PA according to its relative weight. The ratings of performance (expressed as a decimal) of the "typical" SLt on each PA were then multiplied by the share of salary or value of the PA. These net values for each activity were added to obtain the total value of the typical Slt. The standard deviation of these estimates were calculated for comparison with the global 50th percentile estimates.

Finally, overall SDy estimates were calculated for the Schmidt et al. and modified methods by averaging SDy1 and SDy2 from each method. These estimates of SDy were compared with the estimates from 40% and Superior Equivalents procedures.

RESULTS

The results of the analysis of questionnaire responses did not support the hypotheses that the modified estimation procedure would reduce between-judge variance in the estimates of the dollar value of performance nor was there support for the hypothesis that variance in estimates would be significantly less for more experienced supervisors. There was partial support for the hypothesis that the variance in estimates would be significantly different for each rank level.

The differences in the estimates of SDy1 and SDy2 indicated that, on average, the respondents did not perceive the value of performance among MARS SLts as normally distributed. The differences between the SDy1 and SDy2 estimates were greater for the Schmidt et al. method than for the modified method.

When asked to select a diagram which, in their opinion, best reflected the distribution, the majority of the respondents selected diagrams other than the normal distribution, however, a larger number of judges using the modified method selected the normal distribution than did those using the Schmidt et al method.

A comparison of SDy estimates made using the Schmidt et al., modified, Superior Equivalents and 40% methods, revealed that the Superior Equivalents and 40% estimates were similar to each other, but considerably larger than the global estimates. Estimates of the value of the "typical" SLt, using the CREPID type procedure, were more conservative but had the same level of between-judge variance as the global 50th percentile estimates.

Comparison of Schmidt et al. and Modified Method

The percentile point estimates of the dollar value of performance, presented in Table 3, revealed little difference between the Schmidt et al. and modified method. Respondents' estimates were slightly higher for the Schmidt et al. method but remained within 5% of estimates from the modified method. T-tests (see Appendix B) revealed no significant differences at any of the percentile points. For both methods, the mean estimates at the fiftieth percentile were within 2% of the $30,000 average salary of the SLt.

Estimates of SDy were slightly, but not significantly, lower for the Schmidt et al. procedure than for the modified method. All of the SDy estimates (SDy1 and SDy2 for both methods) were between 20 and 25

Table 3.

Mean Percentile Point Estimates
of $ Value of Performance and Estimates of SDy
using the Schmidt et. al. and Modified Method

	n	15%ile	50%ile	85%ile	SDy1	SDy2
Schmidt et. al. Method	94	22840.09	30332.36	36403.49	7492.28	6071.13
Modified Method	109	21741.65	29354.31	36274.31	7612.16	6920.00
All	203	22250.29	29807.20	36334.13	7556.92	6526.93

SDy1 = (50%ile estimate) - (15%ile estimate)
SDy2 = (85%ile estimate) - (50%ile estimate)

percent of the SLt's salary. The difference between SDy1 and SDy2 estimates was greater for the Schmidt et al. method (1421.15) than for the modified method (692.16).

Standard deviations of the estimates at each percentile point were calculated and are reported in Table 4. Again, there was only a small difference between the methods, with the SD of the Schmidt et al estimates only slightly larger than the modified method. F values (see Appendix B) revealed no significant differences in the variances of estimates at any of the three percentile points. Hypothesis 1 was not supported.

Differences in Estimates by Rank and Supervisory Experience

As no significant difference was found between the Schmidt et al. and modified methods, the data from both questionnaires was combined for the analysis of the effect by rank and experience. The means of the percentile point estimates and SDy estimates for each rank and each level of supervisory experience are reported for all respondents in Tables 5 and 6 respectively. Two 2x3 ANOVAs (see Appendix C) were used to examine the effects of estimation method and rank, and estimation method and supervisory experience

Table 4.

Standard Deviation of the Percentile Point Estimates of $ Value of Performance and Estimates of SDy using the Schmidt et. al. and Modified Method

	15%ile	50%ile	85%ile	SDy1	SDy2
Schmidt et.al. Method	7627.86	8091.03	10629.48	5605.60	4177.49
Modified Method	7322.39	7317.32	10394.18	4783.55	5358.67

SDy1 = (50%ile estimate) - (15%ile estimate)
SDy2 = (85%ile estimate) - (50%ile estimate)

Table 5.

Mean Percentile Point Estimates of the
$ Value of Performance and Estimates of SDy, by Rank

	n	15%ile	50%ile	85%ile	SDy1	SDy2
Commander	20	24650.00	33175.00	42375.00	8525.00	9200.00
Lieutenant Commander	62	21796.30	30011.97	37031.10	8215.66	7019.13
Lieutenant	121	22086.26	29145.62	34978.51	7059.36	5832.89

SDy1 = (50%ile estimate) - (15%ile estimate)
SDy2 = (85%ile estimate) - (50%ile estimate)
NOTE: the above means are calculated on the entire respondent population regardless of questionnaire type.

Table 6.

Mean Percentile Point Estimates
of the $ Value of Performance and
Estimates of SDy, by Supervisory Experience

	n *	15%ile	50%ile	85%ile	SDy1	SDy2
0 - 2 years	60	22447.33	29352.00	35323.33	6904.66	5971.33
3 - 5 years	66	20918.13	27689.39	33681.82	6771.25	5992.42
6 or more years	77	23238.58	31977.17	39395.17	8738.58	7418.00

SDy1 = (50%ile estimate) - (15%ile estimate)
SDy2 = (85%ile estimate) - (50%ile estimate)
* - includes all respondents

on each of the three sets of percentile point estimates. The analyses of variance revealed no main effect for estimation method but there was a significant main effect for rank at the 85th percentile (F (2,202) = 4.62, p = .011) and supervisory experience at the 50th percentile (F (2,202) = 6.22, p = .002) and 85th percentile (F (2,202) = 5.95, p = .009). There were no significant interactions between estimation method and rank or estimation method and supervisory experience.

Standard deviations of the percentile point estimates are reported in Table 7 for each of the three levels of supervisory experience. The standard deviations are largest among the most experienced group, opposite to that predicted by hypothesis 2.

Both Cochran C and Bartlett- Box tests (Table 8) revealed significant differences in the variances at the 50th and 85th percentiles but not at the 15th percentile. Pairwise comparisons (F-values) indicated that the largest differences in variances were between the "6 or more years" group and the other two levels (Table 9).

The standard deviation of percentile point estimates for each rank, reported in Table 10, indicate an increase in the SD by rank, especially at the 85th

Table 7.

Standard Deviation of the Percentile Point Estimates of $ Value of Performance by Supervisory Experience

	15%ile	50%ile	85%ile
0 - 2 years	7069.79	6596.94	7836.99
3 - 5 years	6920.24	5928.92	6568.36
6 or more years	8116.16	9173.48	13813.44

Table 8.

Tests of Homogeneity of Variance Among
Percentile Point Estimates By Supervisory Experience Levels

	15%ile	50%ile	85%ile
Cochran C	.4023	.5168**	.6460**
Bartlett-Box F	1.076	7.475*	21.557**

* p<.005
** P<.001

Table 9.

Pairwise Comparisons of the
Variance of Estimates (F values) by
Supervisory Experience at Each Percentile Point

comparison	15%ile	50%ile	85%ile
0-2 years 3-5 years	1.04	1.24	1.42
3-5 years 6 or more years	1.38	2.39*	4.42**
0-2 years 6 or more years	1.32	1.93*	3.11**

* p<.01
** p<.001

Table 10.

Standard Deviation of the Percentile Point Estimates of $ Value of Performance by Rank

	15%ile	50%ile	85%ile
Commander	8215.99	11649.80	17978.70
Lieutenant Commander	8046.81	7501.20	11197.97
Lieutenant	7011.38	6847.01	7854.88

percentile. A comparison of the variance in the estimates for each rank provided partial support for hypothesis 3. The Cochrane C and Bartlett-Box F tests for homogeneity of variance (Table 11) revealed significant differences in the variances by rank at the 50th and 85th percentile. Post-hoc pairwise comparisons (F-values) revealed that the most significant differences were between the Lieutenant and Commander ranks ($p<.001$, Table 12).

The Distribution of Performance

Estimates of SDy1 were larger than SDy2 for both estimation procedures, suggesting that the value of performance was negatively skewed among MARS SLts. The Wilcoxin matched pairs signed-rank test (Seigel, 1956) confirmed that a significantly larger number of respondents estimated the difference between the 85th and 50th percentile as being smaller than the difference between 50th and 15th percentile (Wicoxin signed-rank, $P<.05$). This was true for both the Schmidt et al. estimation procedure and the modified procedure. The mean difference between the two estimates, however, was larger for the Schmidt et al. method than for the modified method.

When asked to select a diagram which, in their opinion, best reflected the distribution of performance

Table 11.

Tests of Homogeneity of Variance
Among Percentile Point Estimates By Rank Levels

	15%ile	50%ile	85%ile
Cochran C	.3721	.5682**	.6443**
Bartlett-Box F	.988	5.8565*	16.546**

* p<.005
** p<.001

Table 12.

Pairwise Comparisons of the Variance of Percentile Point Estimates (F values) By Rank Level at Each Percentile Point

comparison	15%ile	50%ile	85%ile
Commander Lieutenant Commander	1.04	2.41*	2.58*
Lieutenant Commander Lieutenant	1.32	1.20	2.03*
Commander Lieutenant	1.37	2.89**	5.24**

* p<.01
** p<.001

among MARS SLts, 66.6% of the respondents using the Schmidt et al. procedure selected diagrams other than the normal distribution, Of those using the modified procedure, only 52.8% selected a non-normal distribution. A Chi-square analysis revealed that the difference between estimation methods was significant, X (1,N=80)=3.68, p=.055).

Sixty-one percent of the entire sample selected non-normal distributions; of those, 40% indicated that, in their opinion, the distribution was negatively skewed. The remainder were nearly evenly distributed between positive skew, bimodal and rectangular. The percentage of officers selecting the normal, negatively-skewed or other distribution are reported by method, rank and supervisory experience in Table 13.

A Chi-square analysis revealed no significant differences in the selection of diagrams by supervisors at different rank or experience levels.

Superior Equivalence Estimates

As indicated in Table 14, responses to the Eaton et al. (1985) Superior Equivalents procedure, were fairly consistent across methods, ranks and experience levels. The median response was 6 or 7 and the mode was 7 for all but the Cdr rank.. Given an average

Table 13.

Percentage of Supervisors Selecting Type of Distribution of Performance, by Questionnaire Type, Rank and Supervisory Experience.

	Questionnaire		Rank			Experience		
	Schmidt et. al.	Modified Method	Cdr	LCdr	Lt(n)	0-2 yrs	3-5 yrs	6 + yrs
Distribution								
Normal	33.3	47.2	50	37.3	41	48.3	38.3	36.8
Negatively Skewed	30.0	20.8	30	32.2	20.5	20.0	18.3	34.2
Other	36.7	32.1	20	30.5	38.5	31.7	43.3	28.9

* Other - includes positive skew, bimodal and rectangular

Table 14.

Superior Equivalents:
Estimates of the Number of Superior (85th Percentile)
Sub-Lieutenants Required to Perform the Same Duties as an Average SLt

	Questionnaire		Rank			Experience			
	Schmidt et. al.	Modified Method	Cdr	LCdr	Lt	0-2 yrs	3-5 yrs	6 + yrs	All
Mean	6.15	5.91	6.26	6.20	5.89	5.96	5.56	6.45	6.02
Median	7*	6	6	7	7	7	6	7	7

* SDy 7=($12857.14) 6=($20000.00)

salary of $30,000 dollars, using the Eaton et al. procedure, the superior SLt would be worth 10/7 times $30,000 or $42,857.14. The SDy, therefore, would be $12,857.14. Comments during the trial administration and written comments on two of the questionnaires suggested that some respondents believed that a set number of officers are required to run a ship and, regardless of his performance level, an officer "can only be in one place at a time".

CREPID procedure

Using the CREPID procedure, the mean estimate of value of the "typical" job performer was $25,845.81, considerably lower than the mean 50th percentile global estimate of $29807.20. There was no significant difference in variances between the CREPID estimates and the 50th percentile Schmidt et al. estimates; the standard deviation of the CREPID estimates was 8075.07 compared to 8091.03 for the Schmidt et al. 50th percentile estimates.

Comparison of Methods

A comparison of the SDy estimates using the Schmidt et al., modified Schmidt et al., Superior Equivalents and 40% methods (Table 15), revealed little convergence between the estimates made using the global procedures and the other two methods. The Superior Equivalents

Table 15.

SDY Estimates by Estimation Method

Schmidt et. al. (SDy1)	Schmidt et. al. (SDy2)	Modified Method (SDy)	Modified Method (SDy2)	40% Method	Superior Equivalence
7492.28	6071.13	7612.66	6920.00	12,000.00	12,857.14

SDy1 = (50%ile estimate) - (15%ile estimate)
SDy2 = (85%ile estimate) - (50%ile estimate)

estimate was approximately 49% of salary, while the two global estimation procedures resulted in lower estimates of 20-25% of salary.

Discussion

SDy Estimates Using Schmidt et al. and Modified Methods

The primary aim of this study was to determine if the between-judge variance in SDy estimates using the global estimation procedure could be reduced by modifying the Schmidt et al. (1979) instructions to address problems related to: the judges perceptions of distribution and percentiles; the order and context of the estimates; and finally, the cognitive dimensions used by judges. The changes to the instructions did not reduce between-judge variance as was predicted but did affect the judges' reported perception of the distribution of performance and produced SDy estimates which were more consistent with assumption of normally distributed performance than were those produced by the Schmidt et al method.

The SDy estimates using both the Schmidt et al. and the modified instructions were not significantly different and were relatively conservative at 20-25% of salary. The variance in estimates for both methods was relatively small compared to previous studies, suggesting that some of the problems which the modified

instructions were supposed to address may not have been relevant to the naval officer population in this study. For example, the peacetime vs wartime context of the estimates may not have been as significant in this study as it was in U.S. Army studies (Sadaaca and Campbell, 1985 Eaton et al., 1985), where some of the supervisors had been exposed to combat situations. Very few, if any, of the officers in this study had experienced combat and most may have used a peacetime scenario in making their judgements whether instructed to or not.

The effect of having supervisors rate principal activities prior to making estimates of the value of job performance may also have had less impact on this population than on others. All of the supervisors had once performed the SLt's job and had received similar training. Perhaps they were more aware of the relative importance and value of various facets of the jobs under study than supervisors in previous studies who may not have worked at the job being assessed. In other words, the naval officer supervisors may have already been working with similar dimensions.

Rank and Experience of Judges

As predicted there were significant differences in the variance of percentile estimates made at different rank levels. The differences were not as large as those found between Head Tellers and Branch/District Managers in the Mayer (1982) study but do indicate that judges at different levels in the organization may be using different processes or different populations in making their estimates. In this study it may well have been the latter, as those with increased rank were normally more experienced and would have been exposed to a broader range of performance among Sub-Lieutenants.

If, as suggested above, there is less variance in estimates among judges who have worked in the job under study, the recency of that experience may also be a factor in reducing variance. The Lts who more recently shared the experiences of the SLt may be using more similar dimensions in making their estimates than the Commanders whose SLt experience was many years ago.

Another possible explanation for the lower levels of variance among Lt's estimates is the fact that they have greater interaction with SLts in the work place than do LCdrs or Cdrs. In the area of performance assessment, Landy and Farr (1983, p.130) report a study

in which performance ratings made by supervisors with daily but peripheral contact with ratees had interrater reliability of .24 compared to .62 reliability for raters with more relevent contact.

The unexpected finding that between-judge variance increased rather than decreased with the amount of supervisory experience might also be explained by the earlier premise that there is less variance between those judges with more recent experience in the job under study, as those with less supervisory experience are generally those who were most recently employed as SLts.

Distribution of Performance

The assumption that job performance in dollar terms is normally distributed was not supported in this study. The differences between estimates of SDy1 and SDy2 suggest that most of the respondents were working with non-normal distributions when making their estimates. This was confirmed by responses to Part IV of the questionnaires , where 61% of the supervisors selected distributions other than normal as representative of the distribution of job performance among SLt MARS officers.

A comparison of responses from the Schmidt et al and modified procedures indicated that a greater number of respondents using the modified procedure were working with a normal distribution. A significantly larger number (47.2%) of the supervisors using the modified method chose a normal curve as representative of the performance distribution than did supervisors using the Schmidt et al. method (33.3%). Also, the difference between SDy1 and SDy2 estimates was larger for the Schmidt et al. procedure than for the modified procedure. This suggests that the explanation of percentile points and presentation of the diagram of a normal curve in the modified instructions may have had the desired effect of reducing the differences between judges in the way that they perceive the distribution of performance.

The distribution of job performance among D level trained SLts may, in fact, be skewed. Several of the respondents commented that they had supervised few "D level trained" Sub-Lieutenants because within a year of receiving D level training, most Sub-lieutenants are promoted to Lieutenant. As a result, the distribution of performance within the target population may well have been narrow and negatively skewed relative to the entire MARS

occupation. Poor performers would not have successfully completed the demanding training up to this point and would have been selected out. Meanwhile, their relatively junior rank level would not have provided superior performers with the opportunity to fully demonstrate their ability in Naval operations.

The much larger population of Lieutenants was not selected for study because the diversity of training, employment and experience made estimation of the value of job performance very difficult and restricted the number of supervisors.

The skewed distribution of estimates may also have resulted from the fact that supervisors, who are involved in the training process, may have been attending more to poor performance and its consequence of error than to outstanding performance (Schmidt et al. 1984).

These findings illustrate the need for more theoretical attention to the population frames of reference in estimating SDy. Estimates based on a population where performance is not normally distributed or is perceived by judges to be not normally distributed could be grossly inaccurate. In a

study of U.S. Marshals (Karren and Bobko, 1983), some SDy estimates were zero.

Alternative Estimation Methods

When asked to estimate the number of superior SLts that are equal to ten average SLt MARS officers, 66% of the supervisors responses were in the 6-8 range. Comments during the pilot administration and on the questionnaires indicated there were set limits on how few officers, even superior ones, were required run a ship. The SLt is a member of a ships department which is part of an interacting system and cannot perform independently in the safe and efficient operation of a ship. It is suggested that the use of the Eaton et al.(1985) Superior Equivalence procedure, is limited to those situations where individuals work independently, or are in charge, and can in fact provide the same output as a set number of average performers in the same job, such as in sales or in the management of independent departments.

The CREPID procedure estimates of the value of the "typical" MARS SLt were, on average, $4000 below the 50th percentile estimates of the Schmidt et al. procedure. Between-judge variance in the estimates was not improved over that of the Schmidt et al. procedure.

Rating and rankings of the five principle activities were fairly consistent across ranks, indicating that all rank levels had a similar appreciation of the importance and time spent at each of the activities. One advantage of this procedure is that it permits investigation of judges' perceptions of the relative importance and relative time spent by employees at various aspects of the same jobs. This offers a means of determining whether estimates are based on similar dimensions and provides a face validity of the procedure that the Schmidt et al. procedure does not possess.

The estimation of SDy as 40% of salary provided an estimate close to that of the Superior Equivalence method but, unlike previous studies (Schmidt et al., 1982 and Weekly et al., 1985), it resulted in a larger estimate of SDy than the Schmidt et al. (1979) procedure. These results support the arguments of Casio and Ramos (1986) that in certain jobs, particularly at higher levels of the organization, the distribution of performance may be over-estimated with the 40% estimate.

Because between-judge variance was relatively low, and not significantly different, for the Schmidt et al., modified and CREPID methods, it is not possible

to clearly recommend one method for use in a military population. While the 40% method could be recommended for its ease of use, results of this and other studies suggest that it may not provide an accurate estimate of SDy. Feedback from respondents and comments by senior naval officers sponsoring this study, suggest that the CREPID procedure has greater face validity, however, it is a complex estimation procedure which is tied to salary and, as a result, may result in conservative estimates of SDy. The CREPID procedure could be made less cumbersome by using archival personnel evaluation and job analysis data (Edwards et al., 1988). For some Canadian Forces selection situations, Personnel Evaluation Reports and Occupational Analysis data which include frequency and importance (eg. training emphasis) information might be used to replace the information gathered in the CREPID procedure.

What is needed is a procedure which combines the relative simplicity of the Schmidt et al. procedure while reducing between-judge variance and providing face validity by having judges rate similar dimensions which are known to contribute to the value of job performance. During the course of this research, such a procedure was introduced by Tannenbaum and Dickinson (1987). They were able to reduce the variability in

estimates by employing Delphi and Critical Incidence methodologies. In making their percentile estimates, managers were instructed to refer to a list of critical activities which had been compiled from a list of activities, previously generated by managers, which were seen to influence the total yearly value of employees. As in the Burke and Frederick (1984) study the 50th percentile estimates were fed back to the managers. The Delphi technique yielded significantly smaller variances than the normal Schmidt et al. method or the Burke and Frederick procedure.

Limitations of This Research

Because of the apparent difficulty in estimating the dollar value of performance in an occupation such as MARS SLts, where there are no dollar valued outputs, and because senior Naval Officers have traditionally provided strong support for costly and time consuming assessment programs for new officers, it was expected that SDy estimates in this study would be high. This was not the case, however, and estimates were smaller than in many of the previous studies. The reasons for these results could not be investigated in this study.

The use of a mail-out questionnaire precluded the collection of data on the number of supervisors who did

not respond because they were unable or unwilling to place a dollar value on a MARS officers employment. Nor was it possible to determine if the apparently conservative SDy resulted from percentile estimates that were based on an incumbent population that was narrowly defined. A few comments during the pilot administration and on returned questionnaires indicated that a combination of the above factors may have contributed to the small SDy estimates.

Implications for Utility Analysis in the Military

In military occupations, such as naval MARS officers, where the costs of multiple assessment selection procedures are high; where large numbers of applicants are selected annually, and where the correlation of the selection procedure with the criterion is relatively low, the estimation of SDy will have a significant impact on the estimated utility of selection programs. Inaccurate estimates of SDy could result in costly errors in decisions to accept or reject selection methods. The results of this and other studies illustrate the need for an awareness of the potential for error when using supervisor's estimates of SDy in selection utility analysis. Further research is required before any of the procedures can be used

with confidence that they will provide reasonably accurate estimates.

Without an objective measure of the distribution of performance among MARS SLts, it is not possible to determine which of the estimation procedures used in this study provides the most accurate estimates of SDy. It is likely that the Sohmidt et al. method provided an overly conservative estimate due to the narrowly defined population, and that the Superior Equivalence and 40% methods come closer to the true SDy within the MARS SLt population.

Further study, using a more broadly defined population, is required to determine if the modified global procedure can be effective in reducing between-judge variance in other military occupations.

Implications for Future Research

While SDy estimates need not be accurate to the last dollar, severe over- or under-estimation of the value of job performance could lead to decisions with costly consequences. The results of this study suggest that before any estimation procedure is adopted for widespread use, a variety of research needs still must be met.

Additional research is required on the effect of the experience and organizational level of judges on SDy estimates, as well as the effect of the relationship between judges and those on whom estimates are being made. Research is also required to determine if between-judge variance in SDy estimates is lower among supervisors who have previously worked at the jobs for which estimates are being made than among those who have not. In the military context, there is also a need to study the effect of previous combat experience on SDy estimates by military supervisors.

The unequal estimates of SDy between the 15th and 50th and 50th and 85th percentiles indicate that the Schmidt et al.(1979) assumption of normally distributed performance may not always be supported. Before global estimation procedures are adopted, further study is required on both the distribution of the dollar value of performance and on the judges' perceptions of that distribution.

Studies using different estimation methods across different job types are required to determine if different methods are more appropriate in different situations. There should also be further research on the processes used by judges in making estimates and the factors they consider when making them.

Summary and Conclusion

The major findings of this research were that:

a. modifying instructions of the Schmidt et al.(1979) SDy estimation procedure regarding percentile point estimates, their order and context, did not significantly reduce between-judge variance in the estimates;

b. the between-judge variation in percentile point estimates of the dollar value of performance was significantly different for different rank levels at the 50th and 85th percentile;

c. the variance in between-judge estimates of the dollar value of performance increased with level of experience;

d. the Schmidt et al. assumption that job performance in dollar terms is normally distributed was not supported;

e. modifying instructions to the Schmidt et al. procedure significantly effected the judges' reported perceptions of the distribution performance;

f. SDy estimates using the Superior Equivalence Technique were limited by the nature of employment of naval officers;

g. the Schmidt et al. procedure provided a more conservative estimate of SDy than did the 40% or Superior Equivalence methods; and

h. the between-judge variance using a CREPID type procedure was similar to that of the Schmidt et al. method at the 50th percentile.

The results of this study illustrate the need for a great deal more research before any of the SDy estimation procedures are universally adopted for use in organizations such as the military. In the meantime, methods such as the delphi procedure reported by Tannenbaum and Dickinson (1987), which combine reduced between-judge variance, ease of use and face validity may provide the best estimates of SDy for utility analysis.

References

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 Processes, 40 69-95.

Bobko, P., Karren, R., & Parkington, J.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

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

Burke, M.J., & Frederick, J.T. (1986). A comparison of economic utility estimates of alternative SDy estimation procedures. Journal of Applied Psychology, 71, 334 -339.

Cascio, W.F. (1982). Costing Human Resources: The financial impact of behavior in organizations. Boston, Mass: Kent.

Cascio, W.F. & Ramos, R.A. (1986). Development and application of a new method for assessing job performance in behavioral/economic terms. Journal of Applied Psychology, 70, 651 - 661

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) Technical Supplement 88-1, Canadian Forces Personnel Applied Research Unit, Willowdale, Ont.

Cohen, J., (1977). Statistical power analysis for the behavioral sciences New York, NY: Academic Press.

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

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

Edwards, J.E., Frederick, J.T. and Burke, M.J. (1988) Efficacy of Modified CREPID SDys on the Basis of Archival Organizational Data. Journal of Applied Psychology, 73 529-535.

Ferguson, G.A. (1971) Statistical Analysis in Psychology and Education. New York, McGraw-Hill.

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.

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.

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

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

Landy, F.J. & Farr, J.L. (1983) The measurement of work performance: Methods, theory and application. New York: Academic Press.

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 dissertation, Wayne State University, Detroit.

Mathieu, J.E., & Tannenbaum, S.I. (1985). Supervisors' estimates of the dollar value of 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 Southeastern Psychological Association, Atlanta, GA.

Reilly, R.P., & Smither J.W. (1985) An examination of two alternative techniques to estimate standard deviation of job performance in dollars. Journal of Applied Psychology, 71, 20 - 28.

Roche, W.J., Jr. (1965) A dollar criterion in fixed-treatment employee selection. In L. Cronbach & G. Gleser (Eds.), Psychological tests and employment decisions (pp. 254-266). Urbana, IL: 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.

Sadaaca, R., & Campbell, J.P. (1985). Assessing the utility of personnel/classification system. Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, GA.

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

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

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.

Siegel S. (1956). Non-parametric Statistics For The Behavioral Sciences. New York, McGraw-Hill.

Shetzer, L., Bobko, P. (1986). The effects of framing and presentation order on estimates of SDy in utility analyses. Unpublished manuscript, Department of Management, University of Kentucky.

Tannenbaum, S.I. & Dickinson T.L. (1987). Estimating SDy: The effectiveness of Delphi and Critical Incidence Methodologies. Submitted to the American Psychological Association Annual Convention, Division 14, NY.

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.H. (1985). A comparison of three methods of estimating the standard deviation of performance in dollars. Journal of Applied Psychology, 70 651-661.

Winer, B.J. (1971) Statistical Principles in Experimental Design 2nd ed. New York, McGraw-Hill.

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

APPENDIX A

QUESTIONNAIRES

RESEARCH QUESTIONNAIRE (A)

Given government budgetary restrictions, it is often necessary to justify military personnel development and selection programs by placing a dollar value on the benefits that will accrue to the CF if a program is implemented. In order to accomplish this, it sometimes becomes necessary to place a dollar value on the performance of military personnel. One way to do this is through burdensome cost accounting procedures. An alternative, but unproven method, is to have officers estimate the dollar value of performance in occupations which they have supervised. In this questionnaire, you will be asked to make dollar estimates of the value of the performance of MARS officers to the CF.

The dollar utility estimates we are asking you to make are of the type that could be critical in estimating the relative dollar value to the Canadian Forces of different selection methods. In answering these questions, you will have to make some very difficult judgements. We realize they are difficult and that they are judgements or estimates. You will have to ponder for some time before giving each estimate, and there is probably no way you can be absolutely certain your estimate is accurate when you do reach a decision. But keep in mind three things:

1. The alternative to estimates of this kind is application of cost accounting procedures to the evaluation of job performance. Such applications are usually prohibitively expensive and, in the end, they produce only imperfect estimates like this estimation procedure.

.../2

2. Your estimates will be averaged in with those of other supervisors of MARS officers. Thus errors produced by too high or too low estimates will tend to be averaged out, providing more accurate final estimates.

3. The decisions that must be made about selection methods do not require that all estimates be accurate down to the last dollar. Substantially accurate estimates will lead to the same decisions as perfectly accurate ones.

PART I

Based on your experience with MARS officers onboard ships we would like you to estimate the yearly value to your ship of the duties performed by the "average" Sub-Lieutenant ('D' level trained) MARS officer. Consider the quantity and quality of work typical of the average Sub-Lieutenant MARS officer and the value of this work. In placing an overall dollar value on his work, do not assume that an average SLt MARS officer is worth exactly what he is paid. We want your opinion of the value of his performance, which may be more or less than his salary. In making your estimates, it may help to consider what the cost would be if it were possible to contract his work outside to a civilian agency.

> Based on my experience, I estimate the value to my ship of the average Sub-Lieutenant MARS officer at _____ dollars per year.

We would now like you to consider the "superior" Sub-Lieutenant ('D' level) MARS officer. Let us define the superior performer as a MARS officer who is at the 85th percentile. That is, his performance is better than 85% of his fellow SLt MARS officers and only 15% turn in better

.../3

performances. Consider the quality and quantity of the work typical of the superior SLt. Then estimate the value of his services. In placing an overall dollar value on his work, it may again help to consider what the costs would be of having an outside civilian agency perform this work.

> Based on my experience, I estimate the value to my ship of a superior Sub-Lieutenant MARS officer at _____ dollars per year.

Finally, we would like you to consider the "low performing" Sub-Lieutenant ('D' level) MARS officer. Let us define the low performing MARS officer as one who is at the 15th percentile. That is, 85% of all SLt MARS officers turn in better performances than the low performing MARS SLt, and only 15% turn in worse performances. Consider the quality and quantity of the work typical of the low performing SLt. Then estimate the value of his services. In placing an overall dollar value on his work, it may again help to consider what the costs would be of having an outside civilian agency perform this work.

> Based on my experience, I estimate the value to my ship of a low performing Sub-Lieutenant MARS officer at _____ dollars per year.

PART II

Having responded to Part I of the questionnaire you no doubt appreciate the difficulty in trying to put a dollar value on the performance of MARS officers. An alternative method of finding this value might be to rate average and superior performers in terms of their

.../4

relative value. For example, if a superior performer completes twice as many tasks as an average performer in a day, then all else being equal, 5 superior performers are equal to 10 average performers.

Without referring to your dollar estimates above, we would like you to estimate the relative value of average vs. superior SLt MARS officers.

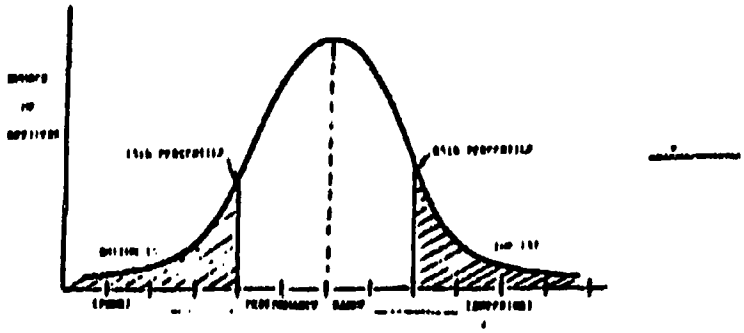
I estimate that, all else being equal, _____ (number) "superior" SLt MARS officers are equal to 10 average SLt MARS officers.

PART III

In Part 1 of this questionnaire, we assumed that the distribution of job performance among SLt MARS officers is normal as represented in the graph below. That is, most MARS SLts are average performers, with equal numbers (15%) falling into the "superior" and "poor" categories.

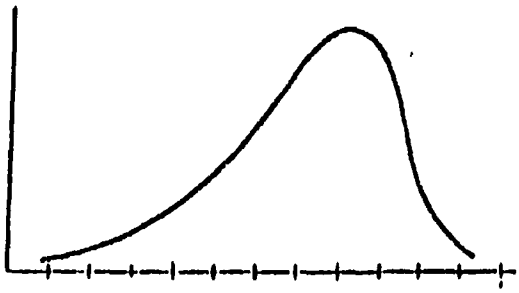
Perhaps, in your experience, you have found that this is not the case. In this last exercise, we would like for you to place a check mark ✓ in the space beside the graph which, in your opinion, best reflects the distribution of job performance among SLt MARS officers.

normal distribution: most are average performers, with 15% superior and 15% poor performers

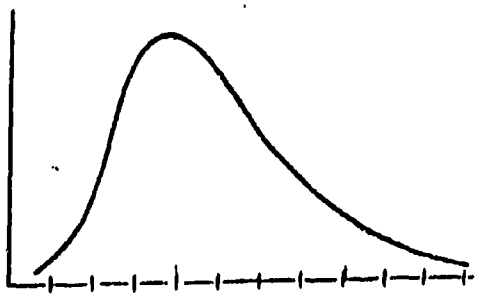


.../5

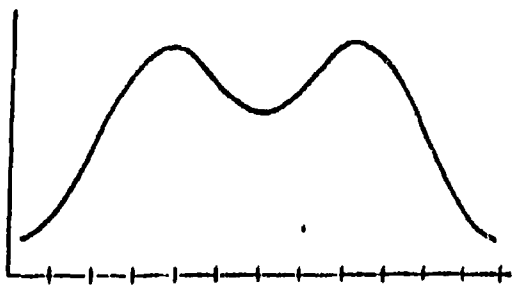
most are above average to superior performers with few poor performers



most are below average to poor performers with few superior performers



performance tends to be above average or below average, with few average performers



performance tends to be evenly distributed among above average, average and below average performers with superior and poor performance not going beyond certain limits (limited by selection, etc.)

If, in your opinion, none of the above figures reflect the distribution of job performance amongst SLt MARS officers, please indicate below with a graph or written explanation how you see the distribution.

The following information is required for research purposes:

1. What is your present rank? Cdr _____ LCdr _____ Lt(N) _____

2. Are you currently supervising a MARS officer? YES NO

3. If no, how long has it been since you last supervised a MARS officer? _____ yrs.

4. How many years experience do you have as a supervisor of MARS officers? _____ yrs.

RESEARCH QUESTIONNAIRE (B)

Given government budgetary restrictions, it is often necessary to justify military personnel development and selection programs by placing a dollar value on the benefits that will accrue to the CF if a program is implemented. In order to accomplish this, it sometimes becomes necessary to place a dollar value on the performance of military personnel. One way to do this is through burdensome cost accounting procedures. An alternative, but unproven method, is to have officers estimate the dollar value of performance in occupations which they have supervised. In this questionnaire, you will be asked to make dollar estimates of the value of the performance of MARS officers to the CF.

The dollar utility estimates we are asking you to make are of the type that could be critical in estimating the relative dollar value to the Canadian Forces of different selection methods. In answering these questions, you will have to make some very difficult judgements. We realize they are difficult and that they are judgements or estimates. We also appreciate the difficulty of placing a dollar value on military performance where success or failure in wartime could be measured in lives. But we want you to make the estimates based on your peacetime experience. You will have to ponder for some time before giving each estimate, and there is probably no way you can be absolutely certain your estimate is accurate when you do reach a decision. But keep in mind three things:

.../2

1. The alternative to estimates of this kind is the application of cost accounting procedures to the evaluation of job performance. Such applications are usually prohibitively expensive and, in the end, they produce only imperfect estimates like this estimation procedure.

2. Your estimates will be averaged in with those of other supervisors of MARS officers. Thus errors produced by too high or too low estimates will tend to be averaged out, providing more accurate final estimates.

3. The decisions that must be made about selection methods do not require that all estimates be accurate down to the last dollar. Substantially accurate estimates will lead to the same decisions as perfectly accurate ones.

PART I

Before providing dollar estimates, we would like you to consider the duties performed by the typical SLt ('D' level trained) MARS officer. Given the performance areas listed below, what percentage of a SLt MARS officer's total work time is spent performing the duties required in each area. List the percentage beside each performance area in column 1 so that the percentages total 100%.

In column 2, rank the performance areas as to their importance to your ship, in your opinion. (From 1st = most important, to 5th = least important).

	% time	ranking
1. Performs the duties of officer of the day in harbour.	_____	_____
2. Performs the duties of officer of the watch at sea.	_____	_____

.../3

3. Performs general/secondary duties as a ships officer. _____ _____

4. Performs the duties assigned in the action organization of a ship. _____ _____

5. Performs the duties of a Divisional Officer. _____ _____

Based on your peacetime experience with MARS officers onboard ships, and keeping in mind the above duties, we would like for you to now estimate the yearly value to your ship of the duties performed by the "average" Sub-Lieutenant ('D' level) MARS officer. Consider the quantity and quality of work typical of the average Sub-Lieutenant MARS officer and the value of this work. In placing an overall dollar value on his work, do not assume that an average SLt MARS officer is worth exactly what he is paid. We want your opinion of the value of his performance, which may be more or less than his salary. In making your estimate, it may help to consider what the cost would be if it were possible to contract his work outside to a civilian agency.

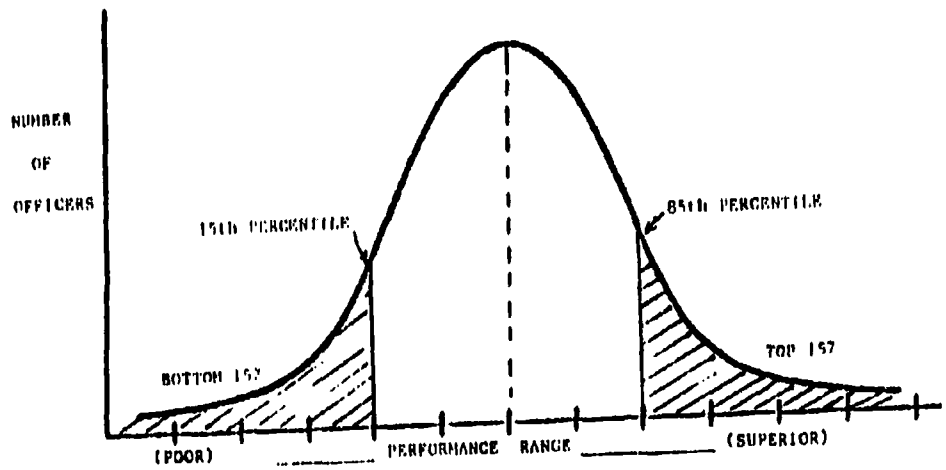
Based on my experience, I estimate the value to my ship of the average Sub-Lieutenant MARS officer at _____ dollars per year.

DO NOT MOVE ON TO THE NEXT STEP UNTIL YOU HAVE RECORDED YOUR ESTIMATE FOR THE AVERAGE SLt.

In the next two steps you will be required to estimate the performance of officers at different levels of performance. The questions will refer to "percentile" points. The percentile represents an individual's relative position in a group. A person at the 50th percent-

.../4

ile, for example, would be the middle or average individual. The performance of a person at the 60th percentile would be such that 60% of the group would fall below him in performance and 40% would perform better. The diagram below illustrates percentile ranks in a normal distribution.



We would now like you to consider the "superior" Sub-Lieutenant ('D' level) MARS officer. Let us define the superior performer as a MARS officer who is at the 85th percentile. That is, his performance is better than 85% of his fellow SLt MARS officers and only 15% turn in better performances. Consider the quality and quantity of the work typical of the superior SLt. Then estimate the value of his services. In placing an overall dollar value on his work, it may again help to consider what the costs would be of having an outside civilian agency perform this work.

Based on my experience, I estimate the value to my ship of a superior Sub-Lieutenant MARS officer at _____ dollars per year.

.../5

DO NOT MOVE ON TO THE NEXT STEP UNTIL YOU HAVE RECORDED YOUR ESTIMATE FOR THE SUPERIOR SLt.

Finally, we would like you to consider the "low performing" Sub-Lieutenant MARS officer. Let us define the low performing MARS officer as one who is at the 15th percentile. That is, 85% of all SLt MARS officers turn in better performances than the low performing MARS SLt, and only 15% turn in worse performances. Consider the quality and quantity of the work typical of the low performing SLt. Then estimate the value of his services. In placing an overall dollar value on his work, it may again help to consider what the costs would be of having an outside civilian agency perform this work.

> Based on my experience, I estimate the value to my ship of a low performing Sub-Lieutenant MARS officer at _____ dollars per year.

PART II

Having responded to Part I of the questionnaire you no doubt appreciate the difficulty in trying to put a dollar value on the performance of MARS officers. An alternative method of finding this value might be to rate average and superior performers in terms of their relative value. For example, if a superior performer completes twice as many tasks as an average performer in a day, then all else being equal, 5 superior performers are equal to 10 average performers.

Without referring to your dollar estimates above and considering the quality of their work and the amount of supervision they require, we

.../6

would like you to estimate the relative value of average vs. superior SLt MARS officers.

I estimate that, all else being equal, _____ (number) "superior" SLt MARS officers are equal to 10 average SLt MARS officers.

PART III

In this part of the questionnaire, we would like you to consider again the "typical" SLt MARS officer and rate his performance relative to the five principal activities we used in Part 1 and which are repeated below. Even though his overall performance may be average, a SLt may perform each of the principal activities at a different level. Use the rating scale below for each of the listed principal activities to rate the performance of a "typical" SLt.

```
0        25        50        75        100       125       150       175       200
+--------+--------|--------+--------|--------+--------|--------+--------|
           performs these      better than       better than     better than
         duties better than        50%               75%             99%
        25% of MARS officers
       I have seen perform them
```

In your opinion, based on the principal activities below and relative to all SLt MARS officers you have seen perform these duties, how does the job performance of the "typical" SLt compare? (Use any number of the 0-200 scale above).

	Rating 0 to 200
1. Performs the duties of officer of the day in harbour.	_____

.../7

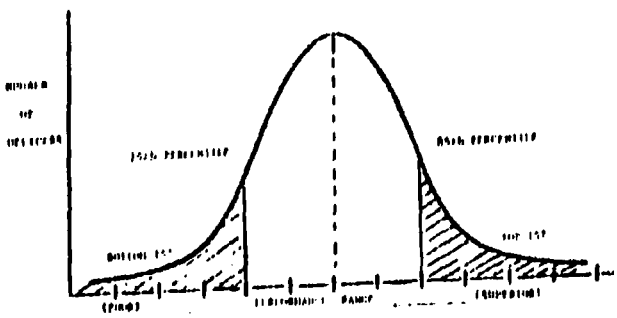
	Rating 0 to 200
2. Performs the duties of officer of the watch at sea.	_____
3. Performs general/secondary duties as a ships officer.	_____
4. Performs the duties assigned in the action organization of a ship.	_____
5. Performs the duties of a Divisional Officer.	_____

PART IV

In Part 1 of this questionnaire, we assumed that the distribution of job performance among SLt MARS officers is normal as represented in the graph below. That is, most MARS SLts are average performers, with equal numbers (15%) falling into the "superior" and "poor" categories.

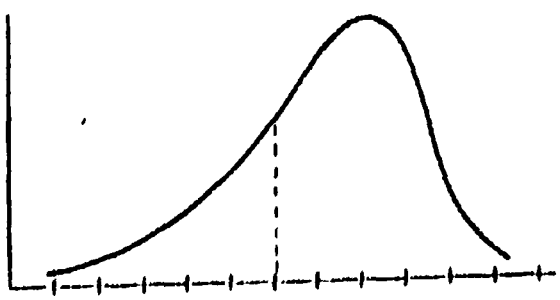
Perhaps, in your experience, you have found that this is not the case. In this last exercise, we would like for you to place a check mark √ in the space beside the graph which, in your opinion, best reflects the distribution of job performance among SLt MARS officers.

normal distribution: most are average performers, with 15% superior and 15% poor performers



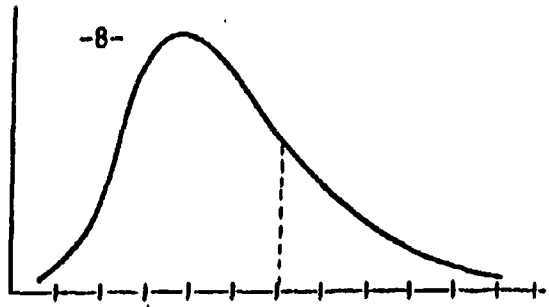
_____

most are above average to superior performers with few poor performers



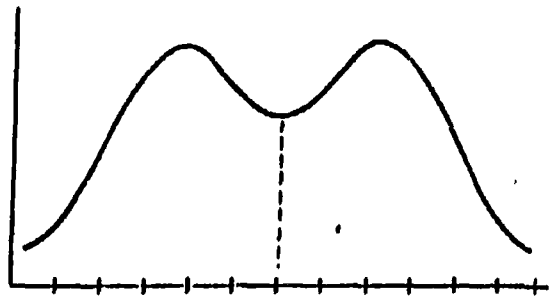
_____

most are below average to poor performers with few superior performers



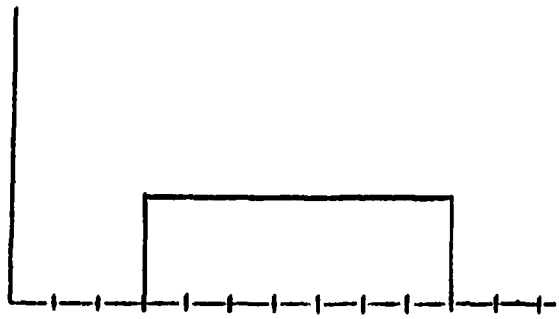
_____

performance tends to be above average or below average, with few average performers



_____

performance tends to be evenly distributed among above average, average and below average performers with superior and poor performance not going beyond certain limits (limited by selection, etc.)



_____

If, in your opinion, none of the above figures reflect the distribution of job performance amongst SLt MARS officers, please indicate below with a graph or written explanation how you see the distribution.

The following information is required for research purposes:

1. What is your present rank? Cdr _____ LCdr _____ Lt(N) _____

2. Are you currently supervising a MARS officer? YES NO

3. If no, how long has it been since you last supervised a MARS officer? _____ yrs.

4. How many years experience do you have as a supervisor of MARS officers? _____ yrs.

National Defence Défense nationale

5762-2-14

Canadian Forces Personnel
Applied Research Unit
Suite 600
4900 Yonge Street
Willowdale, Ontario
M2N 6B7

30 March 1987

Research Participants

PERSONNEL RESEARCH QUESTIONNAIRE

1. As part of a research project sponsored by Maritime Command Headquarters, the enclosed questionnaire is being distributed to naval officers with supervisory experience in the MARS classification. The purpose of the questionnaire is to evaluate different methods of estimating the dollar value of job performance in the MARS classification. Estimates of the dollar value of performance are required in calculating the cost-benefit or utility, of personnel selection and development programs, such as the Naval Officer Selection Boards.

2. The questionnaire will take only a few minutes to complete and your responses will help to identify a less costly and less difficult method for estimating performance value than the cumbersome and disruptive cost-accounting techniques which are currently in use.

3. You are not required to identify yourself on this questionnaire and individual responses to questions will be seen only by the research officers directly involved in the project. If you have any questions or concerns about the questionnaire, or if you would like a summary of the results, you may contact Capt V.W. Johnston, the research officer, through the MARCOM CPSO office, 427-2324.

4. Your participation in this research and early return of the questionnaire is greatly appreciated. Please return your completed questionnaire in the attached self-addressed envelope to CPSO, MARCOM.

T.J. Prociuk
Lieutenant-Colonel
Commanding Officer

Canada

R 211100 APR 87

FM MARCOMHQ HALIFAX

TO MARGEN

INFO NDHQ EXPORT OTTAWA//DPSRSC3//

CFPARU TORONTO

UNCLAS MARGEN 0051/87 DCOS P AND T 4313

SIC WAO

SUBJ: NOSB RESEARCH QUESTIONNAIRE

REFS: A. NDHQ DPSRSC 5762-2-14 8 JUN 84

B. MARCOM 3440-1 (DCOS P AND T) 4 JUL 84

1. AS PART OF THE NAVAL OFFICER PRODUCTION RESEARCH APPROVED AT REF A AND B, TWO SURVEY QUESTIONNAIRES RELATED TO THE EVALUATION OF THE EFFECTIVENESS OF NAVAL OFFICER SELECTION PROCEDURES WILL BE DISTRIBUTED TO SAMPLE GROUPS OF MARS OFFICERS DURING THE WEEK 21 APR

2. THE QUESTIONNAIRES WILL BE MAILED DIRECTLY TO INDIVIDUAL OFFICERS AND PARTICIPATION WILL BE VOLUNTARY AND ANONYMOUS. RESULTS OF THE QUESIONNAIRE WILL BE USED TO IDENTIFY AN EFFECTIVE AND EFFICIENT METHOD OF DETERMINING THE COST BENEFIT OF SELECTION PROCEDURES

3. ANY QUESTIONS REGARDING THE QUESTIONNAIRE OR REQUESTS FOR A SUMMARY OF THE RESULTS OF THE SURVEY SHOULD BE DIRECTED TO THE RESEARCH OFFICER CAPT V.W. JOHNSTON THRU CPSO MARCOM

BT

APPENDIX B

COMPARISONS OF PERCENTILE POINT ESTIMATES

T-TESTS

Differences in Percentile Points
Estimates Between Estimation Methods

	n	mean	SD	SE	T Value	2-tail prob.
15th PERCENTILE						
Schmidt et al.	94	22840.09	7627.86	786.75	1.04	0.299
Modified Method.	109	21741.65	7322.39	701.36		
50th PERCENTILE						
Schmidt et al.	94	30332.36	8091.03	834.52	0.90	0.371
Modified Method.	109	29354.31	7317.31	700.87		
85th PERCENTILE						
Schmidt et al.	94	36403.49	10629.48	1096.35	0.09	0.931
Modified Method.	109	36274.31	10394.18	995.58		

F- Values

Differences Between Variances in
Percentile Estimates by Questionnaire Type

	n	mean	SD	SE	F Value	2-tail prob.
15th PERCENTILE						
Schmidt et al.	94	22840.09	7627.86	786.75	1.09	0.680
Modified Method.	109	21741.65	7322.39	701.36		
50th PERCENTILE						
Schmidt et al.	94	30332.36	8091.03	834.52	1.22	0.313
Modified Method.	109	29354.31	7317.31	700.87		
85th PERCENTILE						
Schmidt et al.	94	36403.49	10629.48	1096.35	1.05	0.819
Modified Method.	109	36274.31	10394.18	995.58		

F-Values

Pairwise Comparison of
Percentile Point Estimates By Rank

	n	mean	SD	SE	F Value	2-tail prob.
15th PERCENTILE						
Commander	20	24650.00	8215.99	1837.15	1.04	0.680
Lt Commander	62	21796.31	8046.81	1021.95		
50th PERCENTILE						
Commander	20	33175.00	11649.79	2604.97	2.41	0.010
Lt Commander	62	30001.97	7501.20	952.65		
85th PERCENTILE						
Commander	20	42375.00	17978.69	4020.16	2.58	0.005
Lt Commander	62	37031.10	11197.97	1422.14		
15th PERCENTILE						
Lt Commander	62	21796.30	8046.81	1021.95	1.32	0.202
Lieutenant	121	22086.26	7011.38	637.40		
50th PERCENTILE						
Lt Commander	62	30001.97	7501.20	952.65	1.20	0.395
Lieutenant	121	29145.62	6847.01	622.46		
85th PERCENTILE						
Lt Commander	62	37031.10	11197.97	1422.14	2.03	0.001
Lieutenant	121	34978.51	7854.88	714.08		
15th PERCENTILE						
Commander	20	24650.00	8215.99	1837.15	1.37	0.306
Lieutenant	121	22086.26	7011.38	637.40		
50th PERCENTILE						
Commander	20	33175.00	11649.79	2604.97	2.89	0.000
Lieutenant	121	29145.62	6847.01	622.46		
85th PERCENTILE						
Commander	20	42375.00	17978.69	4020.16	5.24	0.000
Lieutenant	121	34978.51	7854.88	714.08		

F-Values

Pairwise Comparison Of Percentile Point
Estimates By Years of Supervisory Experience

	n	mean	SD	SE	F Value	2-tail prob.
15th PERCENTILE						
0-2 years	60	22447.33	7069.80	912.71	1.04	0.864
3-5 years	66	20918.14	6920.24	851.82		
50th PERCENTILE						
0-2 years	60	29352.00	6596.94	851.66	1.24	0.400
3-5 years	66	27689.39	5928.92	729.80		
85th PERCENTILE						
0-2 years	60	35323.33	7836.99	1011.75	1.42	0.165
3-5 years	66	33681.82	6568.36	808.51		
15th PERCENTILE						
3-5 years	66	20918.14	6920.24	851.82	1.38	0.188
6 or more years	77	23238.58	8116.16	924.92		
50th PERCENTILE						
3-5 years	66	27689.39	5928.92	729.80	2.39	0.001
6 or more years	77	31977.17	9173.48	1045.42		
85th PERCENTILE						
3-5 years	66	33681.82	6568.36	808.51	4.42	0.000
6 or more years	77	39395.17	13813.44	1574.19		
15th PERCENTILE						
0-2 years	60	22447.33	7069.80	912.71	1.32	0.270
6 or more years	77	23238.58	8116.16	924.92		
50th PERCENTILE						
0-2 years	60	29352.00	6596.94	851.66	1.93	0.009
6 or more years	77	31977.17	9173.48	1045.42		
85th PERCENTILE						
0-2 years	60	35323.33	7836.99	1011.75	3.11	0.000
6 or more years	77	39395.17	13813.44	1574.19		

APPENDIX C

ANALYSIS OF VARIANCE

ANOVAs for 85th Percentile Estimates

Sources of Variation	SS	df	MS	F	Signif. of F
Estimation Method	124968.424	1	124968.424	0.001	0.973
Rank	981603819.657	2	490801909.829	4.620	0.011
Method X Rank	264130816.153	2	132065408.077	1.243	0.291
Explained	1246576867.597	5	249315373.519	2.347	0.043
Residual		197	106244442.016		
Total		202	109785801.706		

Sources of Variation	SS	df	MS	F	Signif. of F
Estimation Method	15500962.399	1	15500962.399	0.146	0.703
Yrs Supervisory Experience	1261742173.326	2	630871086.663	5.948	0.003
Method X Experience	19902732.866	2	9951366.433	0.094	0.910
Explained	1282487137.979	5	256497427.596	2.418	0.037
Residual		197	106062156.379		
Total		202	109785801.706		

ANOVAs for 50th Percentile Estimates

Sources of Variation	SS	df	MS	F	Signif. of F
Estimation Method	43800878.869	1	43800878.869	0.756	0.135
Rank	277920909.517	2	138960454.759	2.400	0.093
Method X Rank	185741695.682	2	92870847.841	1.604	0.204
Explained	511944046.822	5	102388809.364	1.768	0.121
Residual		197	57904686.233		
Total		202	59005778.390		

Sources of Variation	SS	df	MS	F	Signif. of F
Estimation Method	81881745.063	1	81881745.063	1.448	0.231
Yrs Supervisory Experience	704o24375.597	2	352312187.798	6.219	0.002
Method X Experience	5965978.845	2	2982989.423	0.053	0.949
Explained	758871796.065	5	151774359.213	2.679	0.023
Residual		197	56651245.881		
Total		202	59005778.390		

ANOVAs for 15th Percentile Estimates

Sources of Variation	SS	df	MS	F	Signif. of F
Estimation Method	56411815.186	1	56411815.186	1.015	0.315
Rank	126719579.022	2	63359786.511	1.140	0.322
Method X Rank	123967611.866	2	61983805.517	1.155	0.330
Explained	311585611.414	5	62317122.283	1.121	0.350
Residual		197	55589439.198		
Total		202	55755966.077		

Sources of Variation	SS	df	MS	F	Signif. of F
Estimation Method	75231679.794	1	75231679.794	1.357	0.246
Yrs Supervisory Experience	208996512.058	2	104498256.029	1.884	0.155
Method X Experience	67o.3907.899	2	33811953.950	0.610	0.545
Explained	337518841.314	5	67503768.263	1.217	0.302
Residual		197	55457798.437		
Total		202	55755966.007		

APPENDIX D

WILCOXIN MATCHED PAIRS SIGNED-RANK TEST

Wilcoxin Matched-Pairs Signed Ranks Test

TOTAL SAMPLE

SDy1 less than SDy2	= 47		
SDy1 greater than SDY2	= 86		
ties	= 70		
Total	203	z = -3.2542	2-tailed p = .0011

SCHMIDT ET AL METHOD

SDy1 less than SDy2	= 18		
SDy1 greater than SDY2	= 39		
ties	= 37		
Total	203	z = -3.2542	2-tailed p = .0081

MODIFIED METHOD

SDy1 less than SDy2	= 29		
SDy1 greater than SDY2	= 47		
ties	= 33		
Total	203	z = -1.9933	2-tailed p = .0462

APPENDIX E

CHI-SQUARE ANALYSES

Differences in Selection of Normal
Distribution Diagram By Estimation Method

Method	Cases Observed	Expected	Residual	CHI-square	df	Significance
Scmidt et al.	30	38.58	-8.58			
Modified	50	41.41	8.58	3.684	1	0.055

Differences in Selection of
Normal Distribution Diagram By Rank

Rank	Cases Observed	Expected	Residual	CHI-square	df	Significance
Commander	10	8.24	1.76			
Lieutenant Commander	22	24.31	-2.31			
Lieutenant	48	47.45	.55	0.739	2	0.739

Differences in Selection of Normal
Distribution Diagram By Years of Supervisory Experience

Supervisory Experience	Cases Observed	Expected	Residual	CHI-square	df	Significance
0-2 yrs	29	23.53	5.47			
3-5yrs	23	25.88	-2.88			
6 or more yrs	28	30.59	-2.59	1.812	2	0.404