Relations industrielles
Industrial Relations

Identifying Female Officer Potential: An Exploration in Predictors' Payoff

Aharon Tziner and Shimon L. Dolan

See table of contents

Explore this journal

Cite this article

Tziner, A. & Dolan, S. L. (1985). Identifying Female Officer Potential: An Exploration in Predictors' Payoff. *Relations industrielles / Industrial Relations*, *40*(1), 87–98. https://doi.org/10.7202/050111ar

Article abstract

This study examines the effectiveness of multi-predictor selection System as compared with any other alternative selection system which relies on only one predictor.

Identifying Female Officer Potential
An Exploration in Predictors' Payoff

Aharon Tziner
and
Shimon Dolan

This study examines the effectiveness of multi-predictor selection system as compared with any other alternative selection system which relies on only one predictor.

As good management is vital to all, substantial resources are invested in order to locate, train and maintain successful managers (Heneman, Schwab, Fossum and Dyer, 1980; McFarland, 1979). Nevertheless, such managers are generally considered to be scarce and thus organizations are constantly searching for candidates who possess the appropriate qualities. (Burack and Mathys, 1980; Reichley, 1971).

Although the problem of establishing criteria for successful managerial performance has been a controversial issue for quite some time (Fotter, 1981), three types of personal characteristics have always been considered as vital for successful management: (1) intellectual abilities; (2) personality attributes and (3) behavioral characteristics (Campbell, Dunnette, Lawler and Weick, 1970; Dolan, 1981; Fiedler and Chambers, 1974; Ghiselli, 1971; Harrell, 1961; Hunt and Larson, 1979). This fact is not disputed by us — on the contrary — we have used it as the basis for our study, in attempting to find whether such qualities can be effectively predicted by a combination of improved traditional selection methods.

General intelligence scores are commonly used as indicators of intellectual abilities (Reeb, 1976; Schneider, 1976). The validity of these predictors is demonstrated in two important works: The Ghiselli's studies that reported validity coefficients within the range of .20 to .30 (Ghiselli, 1973), and in a synthesis of literature that reported coefficients ranging from .14 to .53 (Campbell et al., 1970). The variations in these validities is attributed to

   • TZINER, Aharon, Professor, Department of Economics-Management, Bar-Ilam University, Israël and Faculty of Social Sciences, Tel Aviv University, Israël.

   DOLAN, Shimon, Professor, École de relations industrielles, Université de Montréal.

the differential criteria against which the predictors were validated (Schneider, 1976). Although the empirical findings categorically indicate that intelligence is an important trait of a successful manager (Dolan and Roy, 1982), the use of intelligence scores exclusively in predicting managerial success is insufficient.

Personality traits have also been suggested as an important factor in managerial success, as these appear to affect the manager's ability to cope with the social and situational demands of his environment (Wanous, 1980). Although there is some empirical support for this claim, (Grimsley and Jarrett, 1975), the reports are disappointing. Guion and Gottier (1965) reviewed 10 years of results that dealt with personality measures and concluded that the various studies yielded only a few validities of any real effectiveness in predicting managerial success (also see Ghiselli, 1973; Tziner, in press). In spite of this, we do not feel that personality traits should be entirely disregarded as a predicting measure. The disappointing results could be due to the approach used in the personality assessment of managerial candidates. Instead of measuring only a single personality trait for prediction, it would be preferable to adopt a holistic approach to personality. This view was already expressed by Argyris (1957, 1964), when he concluded that the holistic approach to personality is more reliable than the single trait approach in predicting future performance.

In addition to intellectual and personality assessments, behavioral measures are being increasingly advocated as an important aid in prediction (Hinrichs, 1978; Norton, 1977), since behavioral manifestations reflect the way a person adjusts to environmental and situational demands. Their relative importance in the selection process has considerably increased, specifically when used in classic managerial situations (Shaham, 1981).

The source of information about the canditate's behavior, however, is a much discussed point. One proposal is to use the assistance of the subject's peers (Love, 1981). The global evaluation of a subject by his peers over an extended period of time and in a variety of situations, may represent the sum total of one's familiarity with his behaviour (Cederblom and Lounsbury, 1980). Peer nominations has been shown to be highly predicting managerial potential (Amir, Kovarsky and Sharan, 1970; Downey, Medland and Yates, 1976; Kane and Lawler, 1978; Moses and Byham, 1977). The correlations reported for this index ranged from .43 to .70. Nevertheless, several recent studies warned against the use of peer nominations as the sole predictor (Barkan and Shirom, 1980).

The final part of this study deals with the methods of collecting and combining the predictive measures. Two methods are often cited: the clinical combination of predicting measures and the mechanical-statistical

combination (Dolan, 1981; Dolan and Roy, 1983; Meehl, 1954; Sawyer, 1966). Campbell et al., (1970) in a review of the relevant literature, concluded that the best approach would be the combination of data into a predictive composite according to a set of mechanically developed common rules (also see Dorans and Drasgow, 1978). This would require the collection of data of both a mechanical and clinical type. However, their recommendation has not yet been substantiated by unequivocal empirical evidence. It was decided to set up the present study, therefore, in order to test the predictive effectiveness of a multiple predictor selection system, in which part of the data are collected mechanically (i.e., intelligence measures and peer nominations score) and part collected clinically (i.e., the global evaluation of personality fitness for a command [managerial] position). In addition to this, a comparison was also made between the predictive validity of the mechanical-statistical method and the clinical method. This particular research project is based on a sample of male applicants for officer-training in the military, and deals with the above-mentioned issues which have, as yet, been little investigated in similar settings.

METHOD

Sample

The subjects were selected from among candidates for officer training in the military. Two samples were chosen in order to enable cross-validation of the findings at a later stage. The first sample (henceforth, Sample A) consisted of 481 subjects participating in one officer training program. The second sample (henceforth, Sample B), consisted of 396 subjects participating in the same training a year later. All subjects were females between the ages of 19 and 21 years, and the program took place during their compulsory service period.

Measures

Several measures were used to assess the potential success of the cadets. They included the following:

(1) General Intelligence Score (GIS)

This measure has been described in detail by Gal (1981). Both the measure and its validity and reliability are described in Reeb (1976). It is worth noting that the measure is an overall composite obtained by a version of Raven's Matrix and an Otis-type test. Although the score ranges from 1 (low) to 7 (high), all subjects in both groups ranked 5 or above, since applicants of lower intelligence were not accepted for training.

(2) Peer Nomination Score (PNS)

This sociometric measure was collected while the research subjects were on a NCO (non-commissioned) course. There they were faced partially with missions and partially with training-test situations resembling those in officer training. This provided opportunities to observe each other's behavior and performance within a framework which resembled an officers training course, and this presumably increased the accuracy of the evaluations. The «sociometric score» was derived from peer nominations using the answers to the instruction: «Nominate who you think may best fit an officer position among your peers.» The number of times a subject was nominated was divided by the total number of possible nominations he could have obtained. For example: if, in a group of 13 subjects, he received 4 out of a possible 12, the result would be $4/12 = 0.3$. This would then be multiplied by 100, resulting in 30 — which would be his peer nomination score. It should be noted that in a group of 13, only 12 choices were possible, as self-nomination was not allowed. The range of this measure is normally 0 to 100, but in the present study, in both sample groups A and B, it was restricted to the upper end, 40 to 100. To ensure a sufficient number of cases in each sociometric score value, for analysis purposes, the 40-100 values were grouped into six distinct categories which would still maintain the upward-skewed shape of the original frequency distribution (i.e. that which ranged from 40 to 100).

(3) General Assessment Score of
Personality Fitness to Command Positions (GASPF)

This score represents the global view of fitness for a command (officer) position formulated clinically by a qualified military psychologist. All psychologists held graduate degrees, were members of the Israel Psychological Association, and had a number of years of experience in clinical assessment. The evaluation was done near the start of the training and was arrived at as follows: each candidate completed a number of personality inventories — Thematic Apperception Test (TAT), Machover D w-a-Person Test (DAP), Biographical Inventory, and Rotter Incomplete Sentences Blank. (A description of them can be found in Anastasi, 1976.) These tests were not administered by psychologists, but by highly skilled staff specifically trained to administer such tests. The military psychologists subsequently examined the responses to the tests. Finally, based on personality tests, the psychologist arrived at an evaluation of the officer potential of each candidate. This evaluation measure normally ranges from 1 (low) to 9 (high). However, in the present study, the range was narrowed

down to a scale of 3-8. Several attempts were made in previous (unpublished) studies in the army, to evaluate the inter-rater reliability of this measure, by having two psychologists evaluating the candidates independently. The results show moderate inter-psychologist reliability as compared with similar clinical-psychological evaluations (Tziner and Dolan, 1982) ranging from .50 to .65.

(4) Predictive-Clinical Composite Score (PCCS)

The above-mentioned measures (GIS, PNS and GASPF) were clinically integrated (i.e. by a clinical judgment) into a single composite which ranged from 1 (lowest) through 12 (highest). This combination was made by unqualified but extremely skilled personnel, who have wide experience in this field. Inter-judge reliability for this clinical composite score proved most satisfactory, ranging from .73 to .84.

The Criterion

The criterion was defined in terms of the final grade obtained by the candidate upon completion of the officer training. It is a weighted composite of scores on subject-matter objective examinations and instructors' ratings of performance in field command exercises. The weights reflected relative importance of the grade components to officers' success on the job. They were agreed on through group discussions with officers who had held command positions and had performed successfully in them. The final grade ranged from 1 (highest) through 7 (lowest/failure).

Procedure

The GIS index was collected at the commencement of the subject's military service, the sociometric score near the end of a former NCO (non-commissioned) training course and the psychological evaluation, near the start of the officer training. The criterion data were obtained from the subject's personal files.

RESULTS

Table 1 displays the zero-order correlation of each predictor with the criterion in Sample A as well as in Sample B. Some scales were inversed so that only positive directions are reported in Table 1. Differences between part of the validity coefficients in the two samples are noticeable. However, these differences did not prove to be significant when tested with Fisher's Z.

Table 1

Zero Order Correlations of Predictors with Criterion Measure
(Coefficients of Predictive Validity)

Predictors	Sample A (N = 481)		Sample B (N = 396)	
	Validity Coefficients		*Validity Coefficients*	
	Uncorrected for restriction in range	*Corrected for restriction in range*	*Uncorrected for restriction in range*	*Corrected for restriction in range*
GIS	.15	.30	.21	.40
PNS	.23	.43	.23	.43
GASPF	.14	.28	.13	.26
PCCS	.24	.45	.27	.50

All correlations are statistically significant, at least p < .05

This is attributable to the fact that subjects in Sample B were distributed slightly higher or lower on the entire range of the variables compared with those in Sample A. Apart from this, the coefficients are reasonable when compared to predictive validities of traditional predictors (Dolan and Roy, 1982). However, the effective validities amongst applicants could be higher, if corrected for restriction in range using the formula proposed by Ghiselli, Campbell, and Zedeck (1981, p. 299) (Table 1). Table 1 also demonstrates that the clinical composite score, PCCS, constructed by combining the three predictors — GIS, PNS and GASPF — yields the highest predictive validity. This finding is most significant — and lends some support to the superiority of a multi-predictor approach to selection. Among the other predictors, the sociometric score (PNS) was shown to be the most valid in relative terms. This is consistent with other studies which reported similar results (Amir et al., 1970; Kane and Lawler, 1978; Lewin and Zwany, 1976; Love, 1981). The General Intelligence Score (GIS) was found to be somewhat inconsistent as a predictor of officer training performance, whereas the GASPF (the assessment of personality fitness to command), while yielding lower validity, appeared to be more stable.

The second aim of the study was to compare the predictive validity of the statistical-mechanical combination of predictors with that of the clinical combination, (GNS, GIS and GASPF). For this a linear regression was used in order to obtain statistically combined measures. The same procedure was

used independently for each of the two sample groups, and following this a cross-validation checked the stability of the measure; this was done by applying the regression weights found in Sample A to Sample B, and vice-versa.

Table 2

Cross-Validation of the Statistical Predictive Measure

	Sample A (N=481)		Sample B (N=396)	
	Uncorrected validity coefficients	Corrected** validity coefficients	Uncorrected validity coefficients	Corrected** validity coefficients
Statistical predicting measure for officer training success*	.30	.54	.33	.59
Statistical predictor score with reversed weights	.27	.50	.27	.50

All coefficients are significant, at least p < .05

* This predicting measure was obtained by the linear combination of GIS, PNS and GASPF. The weights utilized were those calculated by the multiple linear regression of the predictors on the criterion measure.

** They were corrected for restriction in range.

Table 2 indicates that the statistical-linear combination of the three predictors — GNS, GIS and GASPF — yields a valid predictive measure of officer's success, and proved to be superior to the clinical method: its uncorrected validity coefficient ranged from r = .30 to r = .33 as compared with r = .24 of the clinically obtained measure, PCCS. These findings support Sawyer's (1966) assertion about the superiority of the mechanical combination method of predicting data. In addition, the above statistically derived composite proved also to be stable in terms of criterion-related validity in both samples, as demonstrated by the results of the cross-validation procedure.

CONCLUSIONS

These findings suggest that a traditional selection system, when carefully designed, could be effective in predicting successful female performance in training[1].

Certainly, an important finding here is that a multi-predictor selection approach achieves a much higher validity than any other alternative selection system based on only one of these traditional predictors. This conclusion is not surprising, since others have already advocated the use of this approach for forecasting managerial success (Guion, 1976; Wanous, 1980).

However, at the same time it is important to note that the sociometric score (PNS), emerged as the most valid single predictor. This finding is consistent with results previously reported by Downey, Medland and Yates (1976). The reasoning commonly attributed to the sociometric score's predictive validity, is three-fold: (1) the exposure of the candidate to his peers takes place over an extended period of time, and in many different and varied situations; these conditions necessarily produce examples of authentic behavior; (2) the sociometric score is based on the opinions of several judges (peers), which increases its validity and reliability; (3) the sociometric score is less prone to social and/or cultural «biases» than other predictors (Cederblom and Lounsbury, 1980; Kane and Lawler, 1978). However, although the sociometric score emerges as a valid predictor of successful performance it is not always practical in actual selection systems because the judges (peers) must have prior opportunity to acquaint themselves with the subject's behavioural traits relevant to the position in question.

The intelligence score (GIS) also proved to be a reasonably valid predictor, though it is not surprising that the ability to cope with the training and intellectual demands is a pre-requisite for success in an officer training course. Although its contribution to prediction of officer training success cannot be denied, we cannot, however, overlook the fact that GIS' predictive validity was revealed as somewhat unstable. This can be attributed to the fact that the scores of Sample B subjects were slightly higher than those in Sample A[2]. However, according to the Fisher-Z-test the differences between the validity coefficients in the two samples were not significant, and can be attributed to occasional fluctuations.

Psychologist evaluation of personality fitness for a command position was found to be only marginally valid, and this finding is consistent with

---

1 This conclusion is further substantiated by the corrected coefficients of validity.

2 Security constraints prevent us from providing frequency distributions on the variables ranges.

similar findings in other vocational contexts (Anastasi, 1976; Wanous, 1980). The explanation for this could possibly be that although the psychologist is thoroughly acquainted with the requirements for successful officer performance, he is nevertheless influenced to a certain extent by his own preconceived ideas of the ideal officer. Thus, when processing ambiguous points in data, he could possibly be influenced by personal subjective biases and stereotyped models. This would lower the inter-rater reliability of the psychologists and, as a result, the predictive validity of this type of measure.

Finally, let us discuss our findings relating to the second aim of this study, namely the attempt to conclude which of the two methods of predictor combination is superior in terms of validity.

The findings in this research project show the statistically derived composite to be more effective and more stable in its validity than the clinically derived mode. This is consistent with conclusions in other studies (Borman, 1982; Zedeck, Tziner and Middlestadt, 1983), although it refutes Scott's (1970) assertion which advocates clinical combination.

The cross-validation procedure demonstrated that over an extended period of time, the statistical combination method produces a substantially higher rate of success in predictability than the clinical combination method.

The validity of the statistically evolved composite obtained in this study for both Samples A and B, is in the range reported by others, using the same method of predictor combination — between .30 to .33. For example, Norton (1977) in a review of 19 studies concerning the assessment of managerial potential, found that the average coefficient validity of a composite index, devised statistically from the combination of predictors, was .33 while validated against objective performance criteria. Consequently, our findings, although confined to the context of performance in officer-training, may suggest that the selection system we designed in the present study may be a reasonable procedure for identifying managerial potential.

Finally, the present study cannot be summarized without stating an important limitation. The selection system referred to throughout this study was designed to predict potential success in female officership training and not in actual job performance; the two are not necessarily congruent. The efficiency of the procedure should therefore ultimately be evaluated in terms of predictiveness of on-the-job efficiency. Nevertheless, where training costs are high, predicting performance in training is an important intermediate step.

REFERENCES

AMIR, Y., Y. KOVARSKY, and S. SHARAN, «Peer Nominations as a Predictor of Multistage Promotions in a Ramified Organization», *Journal of Applied Psychology*, 54, 462-469, 1970.

ANASTASI, A., *Psychological Testing* (5th ed.), New York, Macmillan, 1976.

ARGYRIS, C., *Personality and Organization*, New York, Harper, 1957.

—————, *Integrating the Individual and the Organization*, New York, Wiley, 1964.

BARKAN, D. and A. SHIROM, «Constructing a Multiple Criteria Model for Manpower Analysis», (In Hebrew), *Megamot*, 26, 204-214, 1980.

BORMAN, W.C., «Validity of Behavioral Assessment for Predicting Military Recruiter Performance», *Journal of Applied Psychology*, 67, 3-9, 1982.

BURACK, E.H. and N.Y. MATHYS, *Human Resource Planning: A Pragmatic Approach for Manpower Staffing and Development*, Brace-Park Press, 1980.

CAMPBELL, J.P., M.D. DUNNETTE, E.E. LAWLER, and K.E. WEICK, *Managerial Behavior, Performance, and Effectiveness*, New York, McGraw-Hill, 1970.

CEDERBLOM, D. and J.W. LOUNSBURY, «An Investigation of User Acceptance of Peer Evaluations», *Personnel Psychology*, 33, 567-579, 1980.

DOLAN, S., *Management Selection: Issues and Viewpoints*, Unpublished manuscript, Université de Montréal, 1981.

DOLAN, S. and D. ROY, *Managerial Selection: Principles and Contemporary Problems*, Monograph No. 11, Université de Montréal, (in French), 1981.

DORANS, N.Y. and F. DRASGOW, «Alternative Weighting Schemes for Linear Prediction», *Organizational Behavior and Human Performance*, 21, 316-345, 1978.

DOWNEY, R.G., F.F. MEDLAND and L.C. YATES, «Evaluation of Peer Rating System for Predicting Subsequent Promotion of Senior Military Officers», *Journal of Applied Psychology*, 61, 206-209, 1976.

FIEDLER, F.E. and M.M. CHAMBERS, Leadership and Management, In J.W. McGuire (ed.), *Contemporary in Management: Issues and Viewpoints*, Englewood Cliffs, N.J., Prentice Hall, 1974.

FOTTER, M.D., «Is Management Really Generic?», *Acamedy of Management Review*, 6, 1-12, 1981.

GAL, R. «Characteristics of Heroism», In S. Breznitz (ed.), *Stress In Israël*, N.Y., Van Nostrand Reinhold Co., 1981.

GHISELLI, E.E., *Explorations in Managerial Talent*, Pacific Palisades, Calif. Goodyear, 1971.

—————, «The Validity of Aptitude Tests in Personnel Selection», *Personnel Psychology*, 26, 461-478, 1973.

GRIMSLEY, G. and H.F. JARRETT, «The Relation of Past Managerial Achievement to Test Measures Obtained in the Employment Situation: Methodology and Results — II», *Personnel Psychology*, 28, 215-231, 1975.

GUION, R.M. and R.F. GOTTIER, «Validity of Personality Measures in Personnel Selection», *Personnel Psychology*, 18, 49-65, 1965.

GUION, R.M., «Recruiting, Selection, and Job Placement», In M.D. Dunnette (ed.), *Handbook of Industrial and Organizational Psychology*, Chicago, Rand McNally, 1976.

HARRELL, T.W., *Manager's Performance and Personality*, Cincinnati, South Western Publishing Co., 1961.

HENEMAN, H.G., D.P. SCHWAB, J.A. FOSSUM and L.D. DYER, *Personnel/Human Resource Management*, Richard D. Irwin, Inc., 1980.

HINRICHS, J.R., «An Eight-Year Follow-up of a Management Assessment Center», *Journal of Applied Psychology*, 63, 596-601, 1978.

HUNT, J.D. and L.L. LARSON, *Cross Currents in Leadership*, Carbondale, Southern Illinois University Press, 1979.

KANE, J.S. and E.E. LAWLER, «Method of Peer Assessment», *Psychological Bulletin*, 85, 555-586, 1978.

LEWIN, A.Y. and A. ZWANY, «Peer Nominations: A Model, Literature Critique and a Paradigm for Research», *Personnel Psychology*, 29, 423-447, 1976.

LOVE, K.G., «Comparison of Peer Assessment Methods: Reliability, Validity, Friendship Bias, and User Reaction», *Journal of Applied Psychology*, 66, 451-457, 1981.

McFARLAND, D.E., *Management: Foundations and Practices* (5th ed.), New York, MacMillan Publishing Co., 1979.

MEEHL, P.E., *Clinical vs. Statistical Prediction,* Minneapolis, University of Minnesota Press, 1954.

MOSES, J.L. and W.C. BYHAM, *Applying the Assessment Center Method*, Elmsford, N.Y., Pergamon, 1977.

NORTON, S.D., «The Empirical and Content Validity of Assessment Centers Versus Traditional Methods for Predicting Managerial Success», *Academy of Management Review*, 2, 442-454, 1977.

REEB, M., «Differential Test Validity for Ethnic Groups in Israël Army and the Effects of Educational Level», *Journal of Applied Psychology*, 61, 253-261, 1976.

REICHLEY, J.A., «Our Critical Shortage of Leadership», *Fortune 84*, 88-93, 1971.

SAWYER, J., «Measurement and Prediction, Clinical and Statistical», *Psychological Bulletin*, 66, 178-200, 1966.

SCHNEIDER, B., *Staffing Organizations*, Pacific Palisades, Calif., Goodyear, 1976.

SCOTT, W.E., Jr., «The Actuarial-Clinical Controversy in Managerial Selection», In M.S. Wortman (Ed.), *Creative Personnel Management Readings in Industrial Relations* (4th. ed.), Boston, Allyn and Bacon, 1970.

SHAHAM, S., *Assessment Centers as a Means for Manpower Management*, (in Hebrew) Unpublished M.A. thesis, The Technology Institute of Haifa (Technion), 1981.

TZINER, A., «Prediction of Peer Rating in a Military Assessment Center: A Longitudinal Follow-up, *Canadian Journal of Administrative Sciences*, in press.

TZINER, A. and S. DOLAN, «Validity of an Assessment Center for Identifying Future Female Officers in the Military», *Journal of Applied Psychology*, 67, 728-736, 1982.

WANOUS, J.P., *Organizational Entry: Recruitment, Selection and Socialization of Newcomers*, Reading, Mass., Addison Wesley, 1980.

ZEDECK, S. A. TZINER, and S.E. MIDDLESTADT, «Interviewer Validity and Reliability: An Individual Analysis Approach», *Personnel Psychology*, 36, 355-370, 1983.

Identification du potentiel des officiers
Étude exploratoire de l'utilité de certains prédicteurs

Cette étude examine l'efficacité d'un système de prédiction à multiples prédicteurs, système qui est comparé à d'autres systèmes n'utilisant qu'un seul prédicteur.

Cette comparaison a été effectuée dans le but de mettre au point un système permettant d'identifier plus efficacement le potentiel des officiers. De plus, elle évalue la validité d'une combinaison clinique de prédicteurs et compare celle-ci à la validité d'une combinaison mécanique-statistique.

Deux groupes de cadets de l'armée, dont l'un de 481 sujets et l'autre de 396 sujets, ont participé à l'étude de la validité des prédicteurs et le score final de chaque cadet a été utilisé comme critère de succès.

L'une des conclusions principales de cette étude est que le système de sélection à multiples prédicteurs est significativement plus efficace que n'importe quel système à prédiction unique.

Il faut noter que, de tous les prédicteurs, la désignation par les pairs (score sociométrique) possède la valeur prédictive la plus élevée.

Enfin, les résultats démontrent que la combinaison statistique-mécanique de prédicteurs est supérieure à la combinaison clinique.