Data Integration and Classification for an Officer Selection System.

François J. LESCREVE
Technical Director
Belgian Armed Forces' Center for Recruitment and Selection
Bruynstraat
B-1120 BRUSSELS (N-O-H)
Belgium
e-mail: Lescreve.F@itc.mil.be

Summary

This paper focuses on the integration of different selection data in order to select and assign officer applicants. First the problem is defined. Three topics are discussed in more detail: the heterogeneousness of the selection data on hand and the problems this can cause, the integration of selection data in order to estimate the suitability of an individual for a specific officer training and the problem of the allocation of candidates to different vacancies. Next, possible approaches are discussed and finally, the paper comes to some conclusions. These advocate the use of modern multi-criteria and multidimensional classification methods to capitalize on the applicant population to optimize the officer corps quality.

Introduction

In most if not any officer selection settings, one is confronted with the situation that a number of candidates apply for a number of vacancies. In many countries, the applicants are youngsters that are totally unknown by the military at the time they apply and usually they can apply for different kinds of officer vacancies depending on the different Services or on different training courses. That is the situation that we'll use as the context for this paper.

The methodology used in officer selection systems is often based on the collection of information from the individual applicants. Once the information is available for all the applicants, decisions are made concerning which applicants will be accepted and for what entry.

This approach has two major aspects. The first question that has to be solved is the one dealing with what information is needed about an applicant. The purpose of that information is to help us to estimate how appropriate it is to accept a particular applicant. It therefore makes sense to look for information that can help us predict the performance of a person if he or she would be accepted as officer cadet. This brings us in the domain of job-analysis. One of the basic rules in selection is to start with a good job-analysis enabling the definition of the job requirements. Once the job-requirements are known, it will be the joined task of selection and training to guarantee that the selected and trained cadets meet them. Although we recognize that the officer corps encompasses a wide variety of jobs and that defining the required aptitudes, skills and personality characteristics is a challenging task, we're still convinced that no selection or training can be conceived without knowing these requirements. The importance of the job analysis for the outcome of quality of the officer corps is paramount (yet frequently overlooked or heavily based on tradition and obsolete scientific points of view), but is not the topic of this paper.

The second aspect in the methodology described above, is the problem of how you integrate and use the information that was gathered to produce the selection and allocation decisions. That is the issue this paper will deal with.

Problem definition

Let us now start analyzing the problem. First, we'll recognize that the selection-data are very heterogeneous and this has tremendous bearing on the way we can integrate them. Secondly, we'll describe how data can be combined to quantify the appropriateness of enlisting an individual applicant for a particular vacancy and finally, we'll discuss the issue of selection and allocation at the group level: how can we capitalize on the applicant population to optimize the overall quality of the enlisted for the different vacancies?

Heterogeneous information

Once the job-analysis issue is solved in one way or another, it appears that different kinds of information are considered relevant to help estimating the suitability of an applicant. There is nothing new about that: Wang (1993) for instance reports that multi-faceted models of personnel assessment were already used by the ancient Chinese 300 years BC. For officer-candidates, the relevant information often includes a medical profile, physical fitness scores, psychometric test results, assessment center ratings, academic examination scores, academic transcripts, biodata, interview scores, the applicant's preference towards the different entries, etc. These data vary significantly in different ways: their relevance towards the officer-job, the degree to which

Paper presented at the RTO HFM Workshop on "Officer Selection", held in Monterey, USA, 9-11 November 1999, and published in RTO MP-55.

poor results can be remedied through training, their measurement quality and related to that, the way you can process them mathematically. Let us expand a little bit on these topics.

Relevance

The first question to be addressed when selection data are discussed is their relevance. Selection data evaluate particular abilities, skills, attitudes etc, or put in a more generic term, person-attributes. Relevance deals with the importance of having that attribute or showing a certain amount of it to perform well, e.g. as an officer. General intelligence for instance will be considered relevant for officer selection because brighter officers tend to be better officers. It is important to emphasize that this relevance is quite independent of the measurability of the attributes. Biometric attributes such as height or weight are very easy to measure yet not very important (with some exceptions). Leadership on the contrary is very relevant for officers but quite difficult to measure. Less precision at the measurement side bears the risk of concealment of the real relationship between the attribute and performance appraisal on the job. Statistical methods will then usually underestimate the importance or weight that has to be given to this attribute. The conclusion that we can draw from this in the light of this paper's topic is that the assessment of the relevance of selection data is paramount for data-integration but that relevance of an attribute is not depending on the precision of its measurement.

Level of remediation through training

The answer to the question on how important it is for a candidate to demonstrate a certain level of an attribute during the selection process, is not only depending on how much the attribute bears on officer proficiency, but also on how well the attribute can be acquired or developed trough training. From the knowledge that the level of physical fitness is easier to improve than general intelligence for instance, it makes sense to treat both attributes in a different way during the selection process. When a certain level of intelligence is required to become an officer, that level will have to be present at the time of selection because it is known that general intelligence is an attribute that doesn't improve through training. During the selection process an adequate measurement of the intelligence has to be performed and applicants with intelligence scores below the required level will have to be rejected. For attributes that can improve a lot by training, things are quite different. Physical fitness is a typical example. Usually, progressively more demanding standards have to be met during the officer training. Through providing adequate physical training to officer cadets and allowing sufficient time between the successive hurdles, the selection standards related to physical fitness shift from the assessment of the level required to be an officer to the assessment of the *potential* needed to reach the set standards after proper training. In practice, this could imply that a selection system would rather use a measure of oxygen absorption capacity than a Cooper test since the first measures potential and the second tests achievement. In conclusion, the weight that is given to a particular selection data or the minimum that is required during selection is partly depending on the trainability of the considered attribute.

Measurement quality

In contrast to some selection settings in civilian life, the officer selection always deals with relatively large numbers of applicants and selection-data. As a consequence, it becomes prohibitive to try to solve the selection and allocation problem in an artisan's way by persons who would know all data on the jobs and the applicants. To solve the equation it will prove to be necessary to code or quantify all selection-data. Quantification is achieved through measurement. Yet, the measurement of the different relevant person-attributes cannot be realized in a unique way. Most books on statistics explain the different measurement scales and the mathematical treatments that are or aren't allowed to perform (Stevens, 1995). For the purpose of data-integration in selection, it is essential to make at least the distinction between metric versus categorical data. Metric data can include test scores, physical fitness scores, biometric data and so on whereas categorical data may encompass previous studies, biodata or medical data. Metric and categorical data are equally important in estimating the suitability of officer applicants. The integration of both kinds of data in a sensible way however is not an easy task.

A last point to make about the heterogeneousness of the selection data concerns the job-preference expressed by the applicant. There is an essential difference between the traditional selection-data and the expressed job-preference in that selection-data are independent of the considered entry. An applicant's verbal ability for instance does not vary whether the candidate applies for the Air Force or the Army. The job-preference does. This means that the job-preference data need to be treated in a particular way. We'll come back to this later.

Individual's assessment

At the individual level, it will be necessary to express the appropriateness of accepting a particular person for a specific entry. When this expression is quantified, we'll call it the *payoff* for individual x and entry y. To establish a payoff, we will have to integrate the available information for each applicant.

But before going into more details, let's open a parenthesis on 'early rejection'. Early rejection occurs when an applicant is screened out before the end of the full selection process. This happens when he or she doesn't meet certain requirements. For instance, one can be rejected for being older than the maximum age allowed by law or because he's considered medically unfit. Screening applicants out is an attractive tool for selection centers because this diminishes the workload. But we should be very cautious when doing so for two major reasons. The one deals with the error of measurement of our selection tools (Lescrève, 1997) and the other with the power of classification methodology that can guarantee that applicants with low scores on specific attributes will not be assigned to entries for which these attributes are important.

Let's now take a closer look at the decision-making for those applicants who complete the full selection-process. Two data-integration systems are widely used: One includes some kind of selection board that reviews the applicant's data and decides on a final score, the other approach does so in a more mathematical way, without human intervention at the individual level.

Officer selection systems often rely on humans for making the final decisions concerning the suitability of an applicant to become an officer (whether or not specified for a particular entry). In many countries, there are selection boards consisting of experienced (senior) officers, eventually assisted by psychologists or other human resource management or training specialists who have at there disposal all available information concerning an applicant. Their task is to integrate that information into a final score or judgment of the applicant. Having extensive experience of military life as an officer is often required to serve in such a board. The idea beyond this is that the board member has to evaluate if the applicant will fit in the officer corps and therefore needs to rely on a personal knowledge acquired through experience. There are a number of things to say to that.

First, we have to admit that the assessors' experience is personal indeed and not necessarily representative for the jobs the applicants are likely to enter. This becomes even more salient in countries having or moving to tri-service selection boards (Belgium, The Netherlands...) where Army officers have e.g. to select officers for the Navy or the Air Force. But even within a Service, the organizational culture of the different branches can be quite different. Often, 20 to 30 years have elapsed since board members joined up. So, their experience of junior officer training is quite old, not to say probably obsolete. If their personal life is mainly confined to contacts with fellow officers they can loose track of the current culture among civilian youngsters. These things bear the risk that the board members rely on irrational beliefs or actual but not representative cases and emphasize the wrong aspects during the assessment of applicants.

A second comment about the members of selection boards concerns their training as assessors. Although efforts are undoubtedly made to train the board members, training is often short and focuses on the specific skills required for the job but not including a more thorough theoretical background that is provided through behavioral sciences education. This opens the risk of having the board members' assessment influenced by affiliation tendencies, transfer, pre-scientific beliefs and so on in an unnoticed way.

The third remark concerning the way selection boards work, has to do with the desirability of assessing the person-organization fit. Keenan (1997) puts it in this way: "The method for developing criteria ... is based on the premise that selectors are primarily attempting to match abilities to job performance requirements. However, in practice, many selectors may be looking for person-organization fit as much as person-job fit, i.e. does the person share the values and 'modal personality' of the organization? While few would seriously question the logic of seeking to maximize person-job fit, the wisdom of focusing excessively on person-organization fit is much more questionable in many circumstances. For example, person-organization fit could actually be counter-productive if innovation is a key organizational requirement." Innovation and flexibility are certainly key elements for modern armies, yet it is surprising to see how often tradition and obsolete scientific principles bear on the officer selection practice.

For the sake of fairness and to help the boards to reach decisions in an economic way - meaning avoiding time-consuming arguments when discussing modal applicants- the boards tend to develop more or less formal decision rules based on the selection data. These rules are derived from practice and are usually unsophisticated.

When confronted with the proposition to use these rules automatically, say by a computer, without further intervention of the board, many selection boards would object vigorously. Their members probably will argue that their job is highly sophisticated and requires a lot of military experience, that each individual must be assessed in a particular way, and so on. They probably will tell you in the end that you can't have a computer make a decision about future officers, without recognizing that computers only apply rules that were given by decision makers in the first place.

The reliability and fairness of selection boards depend on the consistency with which they derive final conclusions from the selection data. This consistency can be altered by the changing composition of the boards over time or geographically and the evolution of actual data processing over time, for instance between the first and last applicants for an annual recruitment. In any case, the reliability of selection boards is quite low compared to the reliability achieved by an algorithm on a computer that equals about one.

Another criticism concerns the way decision rules are elaborated. As mentioned earlier, selection boards often try to develop rules in seeking objectivity and efficiency. In normal circumstances, the quality of those rules depends on the empirical data and the methodology used to develop them. By consequence, the development of such data-integration rules is a specialist's work requiring proper behavioral sciences background. Selection boards not relying on behavioral scientists to develop decision rules are at risk of using totally inadequate rules that have negative impact on the quality of selected officer cadets.

What is the alternative to these selection boards? In theory, the answer is quite simple. Once selection-data are available for an officer-candidate, these data have to be integrated into a payoff value in an objective way. Since it is necessary to quantify all selection data, the objectivity can be achieved by means of having the payoff calculated by computer software. The next question of course is what rules the computer has to apply. That will be discussed in the section on possible solutions further in this paper. For the moment, we only want to state that existing data-integration rules applied in practice are often too unsophisticated to take all relevant selection-data into account and lack specificity for the different entries as officer-cadets.

Classification

Classification is important in settings where all selected officer applicants do not start the same training. This happens in tri-service selection systems where candidates can apply for different Services simultaneously and where an allocation decision has to be made as to what Service they will enter. Classification is also needed in the many settings where the selection is organized for one Service only but where different training courses are available. If there are different entries, it will be necessary to estimate the suitability of an applicant for each one of these entries. As long as there is only one entry, the allocation problem is simple. We can make a single ranking of the applicants based on their payoff values and admit the 'best' ones. The problem is tougher with more entries. Some officer selection systems then try to solve the problem by making a single ranking of all applicants and then assign them sequentially according to their preferences if they meet the minimal standards for the entry of their preference. Modern selection and classification evaluation methodology can demonstrate that these sequential allocation algorithms yield quite poor solutions in terms of entry related quality of the enlisted officer-cadets (Darby et al., 1996).

Possible solutions

In this section we will discuss some approaches to solve the problems identified so far. Afterwards, we'll try to illustrate our points with an example originating from the selection and classification method developed by the Belgian Armed Forces and called *The Psychometric Model.*

Data-integration at the level of the individual officer-applicant

Recognizing that different entries as officer-cadet require different (levels of) attributes, the problem of quantifying the appropriateness of enlisting a candidate has first to be considered at the level of a particular entry. How can we then, based on the metric and categorical selection data and the applicant's job-preference, compute a payoff for a particular entry, that would reflect the desirability to enlist a particular applicant compared to the others? Let's review and comment some techniques.

The *Subject Matter Experts* method consist of identifying a number of persons who can be considered experts in the field of officer training and officer performance appraisal after the training period. These experts decide what information and measures need to be obtained from the officer-candidates during the selection process and how these should be combined to yield a payoff for the entry they are experts in. This method has advantages in the sense that it can be used for new training courses and that it can give an important weight to abilities that are believed to be important but not essential in the early stage of training and officer job. That won't be possible with the next two methods we'll discuss because these rely on the statistical relation between selection and performance after enlistment. The tremendous drawback of this subject matters experts' method lays in its subjectivity. Changing the 'experts' usually will change the requested combination of selection data and this is not likely to be acceptable.

The next two methods are aimed at the computing of payoff values that correlate highly with some chosen criterion that is believed to summarize the assessment of how well an officer (cadet) functions in the Forces. In practice, these payoff values are statistical estimates of the criterion score.

Multiple Linear Regression models are designed to estimate a dependent variable, based upon a series of independent variables. These models are particularly well suited for the prediction of performance criteria based upon selection data and are in fact quite widely used. These models however cannot yield a perfect solution to our problem. First there is the problem of the integration of categorical data. Linear models use primarily metric data and although they can incorporate categorical data or even be based upon categorical data only -in which case they're called *analysis of variance models*- they tend to become fairly complex as soon as the number of categorical variables and the number of classes per variable increases. It is possible to convert each class of a categorical variable into a Boolean 'indicator' variable with value one if the person belongs to the class and zero else. Each indicator variable has then to be integrated in the model equation. Practice shows that huge amounts of data are needed in order to get significant beta-weights for these indicator variables. Another current method allocates a value to each class of the categorical variables that consequently will be treated as if it were metric. The basic difficulty with allocated values is that they define a metric for the classes of the qualitative variable which may not be reasonable (Neter, 1990).

A second problem arises as soon as a performance criterion needs to be chosen to build the linear model. That choice will affect the composition of the officer corps to a large extent. Say for instance that you take the final grades of the first year of the Military Academy as criterion. If the main topic of that academy is rather academic than military, it can be expected that selection variables such as general intelligence or mathematics knowledge will have a large weight in the regression equation compared to e.g. leadership and physical fitness. This means that if you use that linear model to generate the payoff values for your officer selection, you will end up with a hired group that probably will succeed well in the academic training but could perform quite poorly in leadership and physical fitness once they become platoon commanders.

Artificial Neural Networks offer an alternative to the linear models to predict a chosen criterion such as training results or job performance assessments. These models are more powerful to integrate metric and categorical data but require a strong mathematical background and expertise that isn't widespread up to date. They also need huge amounts of data to be developed properly and their usefulness is also conditioned by the choice of a relevant criterion.

The Belgian Psychometric Model uses a combination of the first two methods. Metric data are processed as a weighted sum to generate a provisional payoff. Subject matter experts determine the weights based upon multiple regression equations, where available. The provisional payoffs are consequently altered in a multiplicative way by coefficients given by the experts to each class of the categorical data. In a last step, the modified payoff is altered again to reflect the applicant's preference towards the considered entry. The generic formula used for computing the payoff of an applicant for a specific entry is given below. More details about this formula are provided by Lescrève (1997, 2)

$$Y_{ij} = \left(\sum_{m=1}^{u} \beta_{mj}.X_{im}\right)\left(\prod_{c=1}^{v} \gamma_{cij}\right)\left(\prod_{p=1}^{w} [((X_{ijp}/X_{Maxjp}).\varphi_{pj})+(1-\varphi_{pj})]\right)$$

Y_{ij} is the payoff-value of person i for job j;
m (1 to u) represent the metric variables;
β_{mj} is the weight given to variable m for job j;
X_{im} is the score of person i on variable m;
c (1 to p) represent the categorical variables;
γ_{cij} is the coefficient given for job j to the category of variable c to which person i belongs;
p (1 to w) represent the variables concerning the preferences;
X_{ijp} is the expressed preference of person i for job j on variable p;
X_{Maxjp} is the scale maximum of $X_{ijp.}$ The reason why this is required, is to obtain a maximum value of 1 for the expression X_{ijp}/X_{Maxjp} ;
φ_{pj} is the weight given to preference variable p for job j;

The classification issue

Once a payoff is computed for each applicant-entry combination, a final answer to the question of who will be accepted for each vacancy, still has to be given. The number of possible solutions is huge. If one considers a particular vacancy, it is quite obvious that the 'best' applicant is the one with the highest payoff for that vacancy. On the other hand, when you look at a particular officer-applicant and you wonder for which vacancy the Military would prefer to enlist him or her, you should look for which vacancy he or she has the best results, compared to the other applicants. Provided that the payoff-values have been standardized per vacancy, the vacancy looked for is the one for which the applicant has his or her maximal payoff. This can be done for each vacancy or each applicant but it's trivial that this will not yield the expected solution to our classification problem. Outstanding applicants can indeed be the best for different vacancies for instance. The best possible solution is the one that maximizes the payoff-values of the applicants who are accepted for the vacancies. Finding the solution then becomes an operational research problem. Fortunately, some algorithms derived from the so-called *Hungarian method* or the *Traveling salesman problem* do the job very well. Such approaches are found in the CLASP method (Kroeker & Rafacz, 1983) or the Belgian Psychometric Model (Lescrève, 1993, 1995, 1996). Such models are by far more powerful than sequential allocation methods. As quoted by Hardinge: "By reworking ASVAB data sets, Alley and Teachout (1992) showed that using a differential placement model could result in performance gains of one-third of a standard deviation above current assignment procedures."

Quality assessment of the different solutions

Previous sections have clearly illustrated the complexity of the data-integration issue for officer selection. It is quite obvious that there is no unique solution to the problem. To reach a sound solution, it is necessary to acquire a good understanding of the different aspects of the matter, their interactions and the methods that can help us solve the question. But that is not enough. An essential task in the development and maintenance of a good selection and classification system is the assessment of its quality. This needs to be approached in two ways. The first focuses on the distribution of the attributes in the different entries. For instance, we will compute the average and standard deviation of the enlisted applicants' general intelligence scores in each entry and compare that to the general intelligence level of the applicant population and check whether the results match the different entry requirements (Lescrève, 1998). The second and more important method will look at criterion related quality indicators such as pass-fail rate in the different entries, predictive validity of the payoffs or long-term prediction of officer proficiency.

Conclusion

In this paper, we tried to demonstrate that the quality of the officer selection process, applied to a given group of applicants and a number of different vacancies, does not only depend on the choice and the quality of the selection data, but also to a large extend on the way the available information is processed, first at the individual level and secondly to solve the allocation issue.

Concerning the individual's level, we illustrated some of the methodologically less sound aspects of widespread current practice in officer-selection and tried to explain why these should be avoided. At the group-level, we advocated to get rid of overall rankings in favor of computing separate payoffs for each entry. We also emphasized the need to move from obsolete sequential allocation algorithms for they're bound to perform poorly and to replace them by better multi-dimensional classification methods.

We tried to indicate ways to solve the enumerated problems and illustrated the feasibility of what we discussed by referring to the Belgian Armed Forces Psychometric Model as an example of a generic multi-criteria multi-dimensional selection and classification tool.

Selection was born from the fact that persons differ in many ways. Neither officers nor officer jobs are all the same and it is foolish to believe that or act as if all officers are equal, all-round and interchangeable. As psychologists or military officers involved in officer selection, we should respect the specificity of each applicant, evaluate his or her competency-profile and use it to match it with the profiles required for the different officer jobs. Or, to end with a quote of Hardinge (1997) who peeks in the future of the selection of military staff stating: "Hopefully as attitudes and policies mature, the focus will be more on accepting differences and managing diversity. We will have stopped trying to find better measures of ability and will be looking for a better match between the individual's unique combination of physical, cognitive and affective functions and the organization's training and job requirements. Allocation models will be a major concern."

References

- Alley, W. E., *Recent advances in classification theory and practice.* In M. G. Rumsey, C. B. Walker & J. Harris (Eds)., Personnel *selection* and classification. Hillsdale, New Jersey, Lawrence ERlbaum Associates, Publishers. 1994.
- Burke, E., Kokorian, A., Lescrève, F., Martin, C., Van Raay, P. & Weber, W., *Computer based assessment: a NATO survey*, International Journal os Selection and Assessment, 1995
- Darby, M., Grobman, J., Skinner, J. & Looper, L., *The Generic Assegnment Test and Evaluation Simulator.* Human Resources Directorate, Manpower and Personnel Research Division, Brooks AFB, 1996.
- Hardinge, N. M., *Selection of Military Staff.* In International Handbook of Selection and Assessment., Edited by Neil Anderson and Peter Herriot, Wiley, 1997. p 177-178.
- Keenan, T,. (1997) Selection for Potential: The Case of Graduate Recruitment. in International Handbook of Selection and Assessment., Edited by Neil Anderson and Peter Herriot, Wiley, 1997. p. 510.
- Kroeker, L. & Rafacz, B., *Classification and assignment within pride (CLASP): a recruit assignment model.*, US Navy Personnel Research and Development Center, San Diego, CA, 1983.
- Lawton, D., *A review of the British Army Potential Officer Selection System* In Proceedings of the 36th Annual Conference of the International Military Testing Association, 1994.
- Lescrève, F., *A Psychometric Model for Selection and Assignment of Belgian NCO's* in Proceedings of the 35th annual conference of the Military Testing Association. US Coast Guard, . 1993 p. 527-533.
- Lescrève, F., *The Selection of Belgian NCO's: The Psychometric model goes operational.* in Proceedings of the 37th annual conference of the International Military Testing Association. Canadian Forces Personnel Applied Research Unit. 1995, p. 497-502.
- Lescrève, F., *The use of neural networks as an alternative to multiple regressions and subject matter experts in the prediction of training outcomes.*, Paper presented at the International Applied Military Psychology Symposium, Lisboa, 1995.
- Lescrève, F., *The Psychometric model for the selection of N.C.O.: a statistical review.* International Study Program in Statistics, Catholic university of LEUVEN, 1996

- Lescrève, F., *The* determination *of a cut-off score for the intellectual potential.* Center for Recruitment and Selection: Technical Report 1997-3.
- Lescrève, F., *Data modeling and processing for batch classification* systems. in Proceedings of the 39th Annual Conference of the International Military Testing Association, Sidney, 1997
- Lescrève F., *Immediate assessment of batch classification quality.* In Proceedings of the 37th annual conference of the International Military Testing Association., 1998, Internet: www.internationalmta.org
- Neter, J., Wasserman, W. & Kutner, M. H., *Applied Linear Statistical Models*, Richard D. Irwin, Inc., 3rd Ed, 1990. p. 376.
- Stevens S.S., *Mathematics, measurement and psychophysics.* In S.S. Stevens (Ed.) Handbook of experimental psychology. New York: Wiley, . (1951). p. 1-49
- Wang Z.M., (1993) *Culture, economic reform ad the role ofindustrial/organizational psychology in China.* In M.D. Dunette and L.M. Houghs (eds), Handbook of Industrial and organizational Psychology, 2nd edition, pp. 689-726. Palo Alto, CA: Consulting Psychologists Press.