European Journal of Operational Research 43 (1989) 136–142
North-Holland

Case Study

Measuring the efficiency of maintenance units in the Israeli Air Force

Y. ROLL, B. GOLANY and D. SEROUSSY
Faculty of Industrial Engineering and Management, The Technion—Israel Institute of Technology, Technion City, Haifa, 32000 Israel

Abstract: This paper presents an application of DEA methodology to maintenance units in the Israeli Air Force. Some aspects of using the technique for measuring efficiency within a complex public organization are demonstrated and discussed. Emphasis is put on the choice of factors to enter the analysis and on assigning numerical values to qualitative factors. Relative efficiency ratings, using different reference sets, are used to construct a hierarchical efficiency monitoring system, by which performance of various levels of the organization is evaluated.

Keywords: Data envelopment analysis

1. Introduction

The task of Air-Force maintenance units (MUs) is to provide technical and logistical support to the respective units in which they are assigned. Controlling the performance of these MUs is of crucial importance, and monitoring their efficiency is a vital part of the control system. The maintenance division of the Israeli Air Force (IAF) is a fairly complex organization, structured in several hierarchical levels. MUs at the squadron level were chosen as the lowest 'centers' for efficiency control, since they are the smallest units exercising judgement on maintenance methods and practice. Managing such MUs is considered a difficult task, involving real-time multiple decisions along several dimensions. Severe budgeting constraints intensify the dilemma of allocating scarce resources to the different units, thus making the monitoring of efficiency ever more important.

Received March 1988; revised December 1988

Assessing the efficiency of the MUs by conventional techniques (see, e.g. [9]) proved difficult due, partly, to the following reasons:

– Performance of the MUs is characterized by a large number of 'inputs' and 'outputs' which cannot be readily weighted and compared. Some of these factors are qualitative in nature.

– Although work routines at the MU level are generally well defined, some maintenance tasks can be carried out by different combinations of such routines.

– Explicit functional relationships between inputs and outputs are not known.

Thus, although the work done by MUs is industrial in nature, the environment in which they operate imposes on the situation typical characteristics of not-for-profit public organizations.

A suitable approach for assessing relative efficiencies in organizations possessing such characteristics is Data Envelopment Analysis (DEA). This approach, initially developed by Charnes, Cooper and Rhodes [6], has attracted a growing number of researchers and practitioners for the analysis of performance in situations where conventional techniques are not applicable (see the extensive bibliography in [10]). DEA has been successfully applied in areas such as schools, courts, hospitals, etc. as well as in cases similar to the one discussed here ([1,3]).

This paper follows an extensive field study carried out in MUs of the IAF [11]. However, rather than reporting one more application, attention is focused on the methodical selection of input and output factors which are to enter the analysis, and on the presentation of outcomes so that they may serve as an effective managerial control tool. Repetition of a presentation of the DEA methodology is avoided and interested readers are referred to a recent formal presentation in [4]. The next two sections describe the approach taken for selecting pertinent input and output factors, followed by the presentation of the way qualitative factors are treated. Section 5 gives an outline of a hierarchical efficiency monitoring system, based on DEA outcomes, with some concluding comments offered in the last section.

2. Factors considered

2.1. Classification of factors

Following accepted practice in DEA applications (see [8]), a rather long initial array of input and output factors to be considered was prepared. Listed below are some of these factors, with possible units by which they may be measured, as well as potential difficulties associated with them. For qualitative factors, it was assumed that quantitative data can be obtained from some surrogate measure, linked in a suitable way to each such factor. Most of the factors could be clearly defined as outputs or inputs. Still, as has happened in other DEA applications, the proper placement of some of the factors was not self evident. Such factors were put through further analyses before deciding on their direction.

2.2. Output factors

– Number of sorties performed by the unit served by the respective MU. (Sorties were categorized into two types to which we shall refer as type I and II.)

– Number of flying hours.

– Performance flexibility. A qualitative factor indicating the ability of the MU to operate efficiently under varying work loads. (In particular, the ability to perform satisfactorily at extremely high work loads.) Two measures were considered for this factor:

(a) The ratio of the maximal number of sorties performed in any one day to the average number of sorties per day in a given period.

(b) The standard deviation of the number of daily sorties in a given period.

– Effect on the operational flexibility of the unit served by the MU. Again, two options were suggested to represent this factor:

(a) Number of aircraft grounded due to maintenance problems.

(b) Number of flight assignment cancelled due to maintenance problems.

This is a ‘negative’ output, so that reciprocals of the numerical values were entered into the model.

– Number of flight accidents. Even though this may be an important indicator of the quality of maintenance work, it was excluded from the analysis for a number of reasons. The two principal ones being:

(a) it was exceedingly difficult to isolate the effect of faulty maintenance on accidents, and

(b) there is a varying (sometime significant) time lag between a maintenance operation and the accident occurrence resulting from it.

– Number of aircraft brought back to readiness. This bears a direct relation to the amount of work done by the MU. However, again, several difficulties arise. If the quality of maintenance is poor the number of breakdowns increases and so does the number of repairs. Also, this factor is strongly linked to the type of aircraft assigned to the unit.

– Average time required to bring an aircraft back to readiness. Reservations pointed out for the former factor hold here as well.

– Savings of spare parts, measured by the difference between standard spare parts allocation and actual usage.

– Number of new personnel trained.

2.3. Input factors

– Labor resources. Total work hours spent, appropriately weighted to account for the different professional levels. Difficulties encountered here stem from the nature of aircraft maintenance, in which idle time is almost unavoidable, and where carrying out maintenance during emergencies may require more manpower than the amount needed under normal conditions.

– Operational facility, as determined by the number of planes assigned to the squadron. This is, typically, a nonlinear relationship. When the number of aircraft is relatively small, an increase in this number will generally facilitate carrying out more sorties and performing the required maintenance jobs more conveniently. However, when the number exceeds a certain level it becomes more difficult to perform the routine maintenance jobs.

– Spare parts consumption. This factor (viewed in some cases as an output factor) complements labor input, as the two may—at times—substitute each other.

– Use of equipment and special tools.

2.4. *Types of aircraft*

Although each MU works with only one type of aircraft, there are several such types in all. In order to compare the MUs in a meaningful manner, this factor should be accounted for. Four alternatives were considered:

– Seeing the aircraft type as an independent output factor, quantified in some way.

– Applying it as an 'adjusting factor' to other outputs (e.g. 'number of sorties' and 'number of flying hours').

– Taking it as an independent input factor.

– Applying it as an 'adjusting factor' to other inputs (e.g. 'number of aircraft).

Because of the many possible alternatives, it was decided to choose one of these and introduce the 'type of aircraft' into the analysis only at the final stage. Consequently, this factor was excluded from the steps taken to select the desired mix of inputs and outputs from the potential list.

2.5. *Time period*

The time period chosen for the analysis is a quarter. This is the shortest reporting period where small fluctuations in work-load cancel out each other. Thus, unless otherwise specified, all data is calculated on a quarterly basis.

3. Selection of outputs and inputs

3.1. *Judgemental approach*

This stage was based on qualitative evaluation of the various factors. The main considerations were as follows:

– The extent to which the factors seem to be correlated with overall efficiency.

– Relationships among the factors (e.g. whether they contain similar or overlapping information).

– Availability of the necessary numerical data.

At the conclusion of this stage, the remaining candidate factors for entering the model (in addition to aircraft type) were:

Inputs

X_1 Labor (average daily manpower).
X_2 Operational facility (number of planes).
X_3 Spare parts consumption (in $).

Outputs

Y_1 Sorties of type I.
Y_2 Sorties of type II.
Y_3 Flying hours.
Y_4 Performance flexibility I (ratio of max. daily sorties to average).
Y_5 Performance flexibility II (stand. dev. of number of daily sorties).
Y_6 Operational flexibility (number of cancelled flights).

When the CCR ratio model of DEA was run with all these factors (for units servicing the same type of aircraft), most MUs appeared on the efficiency frontier. This impractical outcome could be attributed to the fact that a large number of factors (particularly, on the output side) was used in the analysis of a relatively small number of units. Apparently, some of the factors tended to gloss over efficiency differences between units rather than highlight these differences. Consequently, a further factor elimination stage was necessary.

3.2. *Analytical approach*

A variety of quantitative models were tried at this stage, in order to isolate the most relevant factors for the DEA analysis.

3.2.1. Pairwise analysis of correlations between factors on the same side of the efficiency ratio

The idea here was to eliminate factors which carry the same contents of information already given by other factors. Hence, factors found to be highly correlated to some of the others were singled out as candidates for elimination.

3.2.2. Applications of the DEA model to each of the outputs separately

Following [5], the single output situation is easier to analyze since its empirical Pareto-efficient function should always be isotone. Thus, each of the output factors was run with the three inputs, using a sample of 30 MUs. The sample included 5 MUs servicing the same type of aircraft, in 6 periods. Applications reported in the literature do not suggest specific quantitative measures which can be used at this stage. In our case, we evaluate the results using the following three new measures.

– *Distinction power* (S), estimated by the standard deviation of the efficiency scores. Here we sought to find the input–output mix which will distinguish in the clearest way between the MUs. As noted in Section 3.1, using the initial list of all the inputs and outputs originally suggested, resulted in most MUs appearing as efficient.

– *Distinction sharpness* (R), measured by the maximal (ordinal) ranking range of the outcomes. This measure overlaps to a certain extent the previous one in its attempt to find the input–output mix which would bring about the largest dispersion of outcomes.

– *Conformity* (Q), the linear correlation coefficient between values of individual output factors and the efficiency scores. Poor performance here suggested a weak relationship between the output and the inputs chosen. Again, output factors performing poorly in this analysis were candidates for elimination.

3.2.3. Running alternative DEA formulations

Ten different combinations of output factors (see Table 1) were tried out with three inputs. The same sample of 30 MUs was used for these runs. In addition to the discrimination power (S) and the sharpness (R), two more measures were calculated:

– Average of the efficiency scores (P). This average usually relates to the R and S measures since the lower its value the more likely it is that the distinction measures applied earlier will have higher values.

– Grouping consistency (T). Here, the entire range of outcomes was divided into five efficiency classes, and the classification of each MU within this division noted. Then, T_j was calculated as

$$T_j = (1/n) \sum_{i=1}^{n} t_{ij},$$

where t_{ij} denotes the number of times MU_i was rated by all other formulations in the same efficiency group as in formulation j. The rationale here is that we seek to find the input–output mixes which provide robustness to our analysis in the sense that the efficiency classification which they generate is close to most other mixes.

Table 2 lists the values of the four measures, the respective ranking by each measure (in ascending order of desirability), and the overall ranking (by the product of the individual rankings). Model

Table 1
Combinations of output factors

Model	Y_1	Y_2	Y_1+Y_2	Y_3	Y_4	Y_5	Y_6
I	×	×		×	×	×	×
II	×	×		×		×	
III	×	×		×			
IV		×		×		×	
V	×	×					
VI		×		×			
VII				×		×	
VIII			×	×		×	
IX			×	×			
X			×			×	

Table 2
Evaluation and ranking of alternative model formulations

Model	P		S		R		T		Overall ranking
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	
I	1	1	0	1	1	1	7.60	8	1
II	0.98	2	0.0441	2	6	2	7.53	7	2
III	0.97	3	0.0661	4	7	3	7.60	8	4
IV	0.98	2	0.0456	3	7	3	7.37	5	3
V	0.96	4	0.0675	5	12	5	7.50	6	7
VI	0.96	4	0.078	6	10	4	7.20	4	6
VII	0.92	6	0.0937	8	14	7	6.90	3	10
VIII	0.93	5	0.0912	7	12	5	7.20	4	8
IX	0.90	7	0.1279	10	14	7	6.57	2	9
X	0.92	6	0.0984	9	13	6	6.33	1	5

VII, of the ten formulations tested, emerged as possessing the best discriminating characteristics. Its two outputs, Y_3 and Y_5, also rated favorably in the two former analyses. The final model included, thus, two outputs and three inputs, with the additional effect of aircraft type.

4. Treating qualitative factors

Of the factors selected for inclusion in the final model, there were qualitative ones. 'Type of aircraft' did not enter the previous analyses, while 'Performance flexibility II' (Y_5), and 'Operational facility' (X_2) were assigned numerical values by assuming simple linear relations to their respective surrogate measures (see [7]). This section describes the process by which suitable functional relationships, between the qualitative factors and their measurable surrogates, were identified.

4.1. The type of aircraft

Since different MUs service different types of aircraft, it was necessary to take this factor into account. As mentioned in Section 2.4, four alternative ways of incorporating this factor were considered. In addition, two possible numerical scales were tested for this factor. One was derived from an analysis of past maintenance costs. The second was a subjective scale, based on judgement by a team of experts, of the relative difficulty in maintaining the different aircraft. The eight ensuing combinations were first tested on a homogeneous sample of MUs (i.e. all operating with the same aircraft), to see whether entering the type of aircraft as an additional variable (in any one of the combinations) would alter previous outcomes. Then, the eight combinations were tested on a heterogeneous (i.e. servicing different aircraft) sample of MUs, applying the discrimination measures mentioned in Section 3.2. This procedure led to a decision to introduce the type of aircraft as a multiplier (i.e. a weighting factor) of flying hours (Y_3), using the subjective scale.

4.2. Performance flexibility II

This output factor was measured by the standard deviation (SD) of the daily number of sorties. Two general functional relationships between SD and performance flexibility II (Y_5) were considered:

$$Y_5 = (\mathrm{SD})^c, \tag{1}$$

$$Y_5 = c \cdot \exp(\mathrm{SD}). \tag{2}$$

Again, a series of tests was run with different values of the parameter c. It was found that outcomes were rather insensitive to changes in the value of c. The best results (in terms of the aforementioned discrimination measures) were, however, obtained with expression (1) and $c = 2$.

4.3. Operational facility

This input factor expresses the effect of the number of aircraft on MU performance. The relationship between operational facility and the number of aircraft is a complex one (see Section 2.3), and the following functional forms were tried out:

$$X_2 = c \cdot N^{\beta}, \quad \beta \le 1, \tag{3}$$

$$X_2 = c_1 \cdot N^{\beta} - c_2 \cdot N^{\delta}, \quad c_2 \ll c_1, \quad \beta \le 1, \quad \delta \ge 1, \tag{4}$$

where N denotes the number of aircraft and c_1, c_2, β, δ are parameters.

Expression (3) assumes that the operational facility increases with the number of aircraft, albeit at a decreasing rate. Expression (4) allows for a decrease in operational facility (from the point of view of maintenance), when N goes beyond a ceratin limiting value. Again, a team of experts was called to evaluate operational facility with different number of aircraft (all other circumstances equal), in a pairwise manner. Comparing these evaluations to results obtained from the two formulations, resulted in a decision to adopt expression (4) as representing the relations between the number of aircraft and operational facility.

5. A hierarchical efficiency monitoring system

5.1. Levels of control

For purposes of monitoring the efficiency of MUs, the IAF can be divided into four managerial levels. MUs constitute the basic lower level, responsible for carrying out the day to day maintenance operations. The upper level is that of the entire IAF which issues general policy directives. There are two parallel intermediate levels. One is the formation level, comprising generally one type of aircraft, and responsible for the engineering aspects of maintenance. The other is the base level, which looks after the various administrative aspects.

The DEA model with outputs and inputs as described earlier, was applied to the MUs across the four managerial levels. Each MU was evaluated, relatively to all other MUs within the respective level, over a series of time windows. The outcomes were used to analyze MU, formation and base performance from several points of view. In particular, the following aspects were examined:

– Base and formation management. The average efficiency of MUs within each group, at these levels, can be used to assess the overall efficiency with which these groups carry out their tasks. Differences in performance may be traced back to variations in management style, work procedures and degree of control exercised by the respective managerial levels on their MUs. General inferences can be drawn on the contribution of different policies towards higher MU efficiency.

– Individual MU management. This direct outcome of the efficiency analyses, serves as a continuous monitoring tool of MU performance. In particular, information provided by the values of the slacks in the model is used for locating inefficiencies and instituting improvement steps.

– Trends in efficiency changes. These could be followed for individual MUs as well as for the intermediate managerial levels and the IAF as a whole. Trend data can be used to evaluate the effects of past efficiency enhancement steps and in setting goals for the future.

Data obtained from the hierarchical efficiency monitoring system (HEMS) are used as additional inputs for policy setting at the various levels of the aircraft maintenance organization. They affect decisions in diverse areas such as resource allocation,

Table 3

Base	Formation		
	A	*B*	*C*
I	A_1, A_2	B_1	
II		B_2, B_3	C_1, C_2

Table 4
Outcomes of sample runs

MU	Individual MU level	Formation level			Base level		Entire sample level
		A	*B*	*C*	I	II	
A_1	1.0	1.0			1.0		0.74
A_2	0.94	0.92			0.91		0.60
B_1	0.80		0.44		0.80		0.44
B_2	1.0		0.54			0.54	0.54
B_3	1.0		1.0			1.0	1.0
C_1	1.0			1.0		1.0	1.0
C_2	1.0			1.0		1.0	0.99
Averages	0.96	0.96	0.66	1.0	0.90	0.89	0.76
Averages at the sample level					Base I		0.59
					Base II		0.88
					Formation *A*		0.67
					Formation *B*		0.66
					Formation *C*		1.0

training programs, etc., as well as overall management evaluation.

5.2. *Illustrative example*

Outcomes of the proposed HEMS and the way they can be analyzed are demonstrated with the following numerical example. A sample of seven MUs, out of three formations and two bases, was selected to represent the entire organization. The MUs were distributed as given in Table 3. DEA was run for five time windows at all levels. Thus, the sample (organization level) included 35 DMUs, base I level included 15, base II 20 DMUs, and formations *A*, *B* and *C* included 10, 15 and 10 MUs respectively. Average outcomes of the five periods are depicted in Table 4. The different outcomes for each MU, at the various levels, highlight the relative nature of efficiency ratings obtained with DEA. Figures at the individual MU level point mainly to changes in efficiency during the five periods covered by the sample, and not to the potential for further improvement. A typical example is MU B_2 which was operating at a constant efficiency during the evaluated time periods, thus attaining an efficiency score of 1.0 when compared only with respect to itself. However, with regard to the entire sample, the score of the same MU was merely 0.54. Similarly, base level averages point to efficiency variations among units in each base, thus reflecting (to some extent) the degree of administrative control over the units. However, the two bases, scoring very similar averages within their own units (0.90 and 0.89), came out significantly different when compared at the sample level (0.59 and 0.88). An analysis along similar lines of formation level outcomes may point to the degree of technical competence of the different formations, as against their ability to impose directives on their units (formations A and C seem to be similar in the latter aspect, while apparantly quite different in the first one).

Analysis of the higher level is, of course, based on the assumption that differences between the different aircraft types are captured correctly by the model. This assumption, found valid when constructing the model (see Section 4.1), is not refuted by the sample outcomes where the most marked unit efficiency differences were found within formation B operating the same aircraft.

6. Concluding comments

Measuring efficiency in a complex organization, such as the maintenance division of the IAF, usually involves taking a series of subjective decisions [2]. This is particularly prominent in the process of selecting appropriate factors to enter the analysis as well as in specifying the mechanisms for assigning these factors numerical values. Similar problems are present whenever DEA is applied. Several approaches were demonstrated for selecting from among output and input factors, with a view to obtaining a system which would discriminate clearly between efficient and inefficient units. Techniques for locating suitable functional forms, linking surrogate sets of (measurable) variables to the desired qualitative factors were presented.

The proposed HEMS makes use of the DEA characteristic by which it provides *relative* efficiency ratings. Through establishing several reference sets, different efficiency ratings are obtained. The differences between these ratings can then be used to draw general inferences on the performance of various levels within the organization.

Obviously, the stricter the standards against which efficiency is evaluated, the sharper the management control tool. Such additional sharpness can be introduced by adding to the set of observed data some reliable standard performance figures. This is one of the possible directions for further development of the HEMS. Another direction involves periodic reviews of the pertinent output and input factors, and of the way these are evaluated.

References

[1] Bowlin, W.F., "Evaluating the efficiency of US Air Force real property maintenance activities", *Journal of the Operational Research Society* 38/2 (1987) 127–135.

[2] Cameron, K.S., "Effectiveness as paradox: Consensus and conflict in conceptions of organizational effectiveness", *Management Science* 32/5 (1986).

[3] Charnes, A., Clark, T., Cooper, W.W., and Golany, B., "A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the US Air Force", *Annals of Operations Research* 2 (1985) 95–112.

[4] Charnes, A., and Cooper, W.W., "Preface to topics in data envelopment analysis", *Annals of Operations Research* 2 (1985) 59–94.

[5] Charnes, A., Cooper, W.W., Golany, B., Seiford, L., and Stutz, J., "Foundations of data envelopment analysis for Pareto–Koopmans efficient empirical production functions", *Journal of Econometrics* 30 (1985) 91–107.

[6] Charnes, A., Cooper, W.W., and Rhodes, E., "Measuring efficiency of decision making units", *European Journal of Operations Research* 2/2 (1978) 429–444.

[7] Cook, W., Roll, Y., and Seiford, L.M., "On the sensitivity of DEA outcomes to the scales on which factors are measured", Working paper, York University, 1986.

[8] Golany, B., and Roll, Y., "An application procedure for DEA", *OMEGA* 17/3 (1989) 237–250.

[9] Roll, Y., and Sachish, A., "Productivity measurement at the plant level", *OMEGA* 9/1 (1981) 37–42.

[10] Seiford, L., "A bibliography of data envelopment analysis (1978–1986)", Working paper, The University of Massachusetts, Amhurst, 1987.

[11] Seroussy, D., "Applying DEA to the I.D.F. Air-Force maintenance units", MSc Thesis, The Technion, Israel Institute of Technology, June, 1987.