Presentation of a Swedish Study Program Concerning Recruitment, Selection and Training of Student Air Traffic Controllers: The MRU Project Phase 1

Rune Haglund

Civil Aviation Administration, Sweden

Introduction

Background Phase 1

The Director of the ANS Department has set up an objective for the efficiency of screening and training procedures for air traffic controller students which implies that all students admitted "shall be considered to have the qualification for – and be given the means of – completing the training".

As a consequence, a study project has been established. It is run by the ANS Department with members from the Swedish CAA, in close cooperation with Uppsala University.

The task force of the MRU project consists of following members:

- Mr. Rune Haglund, Project Manager, Senior ATS Specialist, Swedish CAA
- Mr. Bertil Andersson, Air Traffic Controller, Swedish CAA
- Mr. Björn Backman, Industrial Psychologist, Swedish CAA
- Mr. Olle Sundin, Manager Arlanda ATS, Swedish CAA External expert
- Professor Berndt Brehmer, Ph.D., Department of Psychology, Uppsala University.

Graduation Rate

On the first of January, 1978 the military and civil ATS systems in Sweden were totally integrated into the Swedish Civil Aviation Administration. As a preparation for this alteration a new ATS Academy was created and a new integrated air traffic controller training programme implemented in 1974. One of the aims for this training programme was to decrease the failure rate to a maximum of 20 percent.

This objective has not been reached. However, since the start 1974, the average failure rate has been reduced by almost 20 percent. This improvement cannot be described as a steady curve. Instead, there is a great deal of unpredictible fluctuation around an average figure for successful training results:

• During the 1970's	Average 54%	Range 27% - 71%
• During the 1980's	Average 66%	Range 57% - 86%
• During the 1990's	Average 74%	Range 63% - 90%

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland

Conclusions about the success rate and trends regarding the present recruitment, selection and training procedures are based on simple Analysis of Variance. Each decade was considered a group and the success rate of every class in that decade is the dependent variable. McNemar's (1969) formula for groups of unequal size has been used to test the significance of differences between the means.

$$F=s^2_b/s^2_w = .082/.015 = 5.467 \; p < 0.05$$

The conclusion of the task force is that there are systematic differences between the decades and that they are due to the greater experience of the people involved, instead of the systematic changes in recruitment, selection and training procedures. Interestingly enough, the failure rate has decreased by one percentage unit per year since the start of the new integrated ATS Academy. The mean for the years 1990–1993 has been calculated on the basis of 8 completed classes with a total of 190 accepted students and from which 140 graduated (74 percent). This outcome can be compared with the rate numbers of 80 percent graduating from the FAA Academy that FAA reports (MRU Delrapport 3, 1993).

Economic Review

As a key figure for reviewing the costs of the recruitment, selection and training system, one can calculate the total costs per graduated student. The total cost to the Swedish CAA for providing a new licenced TWR/TMC air traffic controller is 205,000 USD. For an ACC controller, the cost increases to a total of 255,000 USD.

This total cost can be divided with the total amount of weeks in training as shown below (currency in SEK). Figure 1 shows the costs accumulated over weeks.

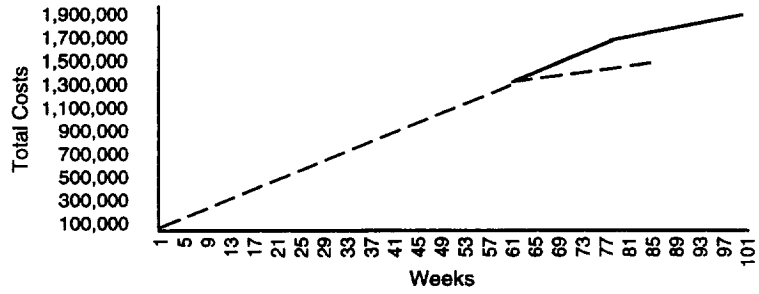

Figure 1. Total Costs over Weeks.

Today, the Swedish CAA has achieved a balance between supply and demand with respect to air traffic controllers. This leads to an acute planning problem that can be described as follows.

Today CAA has to employ 27 students to be able to deliver 20 air traffic controllers into the ATS system. This is due to the unpredictible span between accepted and graduated students in the present screening and training system: the current system will provide an outcome of qualified licenced air traffic controllers that varies, by chance, from 17 up to 24. This

uncertainty has great negative effects on both planning and economics (lack or surplus of personnel).

The outcome of today's system of recruitment, selection and training of controller students is not satisfactory because it generates both a costly fluctuation around the mean and allows students who do not have the necessary abilities to remain in training for too long before they are expelled. An efficient system with a more predictable outcome and a higher success rate, i.e., a deviation of not more than 10 percent around the mean and a 90 percent success rate, would save the Swedish CAA at least 520,000 USD per class of 20 students.

Goal Setting

It is the opinion of the MRU task force that money and other resources invested in developing procedures for recruitment, selection and training of controller students, so that the outcome is less affected by random errors, will be a good investment.

The Research

In order to pursue the causes of today's random errors, the MRU task force issues followed a two step procedure. Step one involved a job analysis, and step two a study of the correlations between tests and training results. These two steps were taken in order to validate current recruitment, selection and training methods in use by the Swedish CAA.

Job Analysis in Order to Determine the Job Criteria

A number of acceptable procedures exist for conducting a job analysis. One way is to interview observers who are aware of the aims and objectives of the air traffic controller's profession and who see the controllers perform their profession on a duties frequent basis. Thus supervisors, peers and instructors may be interviewed about their observations of the critical requirements of the air traffic controller profession.

Current international research and analyses of the controller's job show that the air traffic controller profession is a very complex occupation where the tasks are performed in a very special work environment.

The Selection Procedure

Brehmer (1993) notes that the current selection procedure is based on a series of tests and interviews. The tests have been chosen by ABAR, a consulting company specialising in psychology of work and organisation. The choice of tests seems to have been made on the basis of a general analysis of the air traffic controller's job. But there has been no standardisation or statistical evaluation of the effectiveness of the selection procedure, except for later addition: the use of percept-genetic techniques.

The paper-and-pencil tests are described in terms of four factors (ABAR, 1978):

- *Flexibility and ability to find new solutions.* The aspect is measured by means of three tests: "Skeppsdestination" (Ship's destination), "Instruktionsprov II" (Instruction test II) and "Kravatt" (Neck tie) and concerns the ability to improvise and make decisions in unexpected situations.

- *Logical ability.* Logical ability is measured by means of two tests: Raven's matrices and Number series, which are designed to measure logical ability.
- *Spatial ability.* The aim is to measure the ability to construct a three dimensional picture of the air space from two dimensional information. Three tests are used for this: "Klossar" (Blocks), "Pl–tmodeller" (Metal Sheet Models) and "WIT Puzzles."
- *Attention to detail, carefulness, and short term memory.* This factor is measured with five tests: "Korrektur ABAR" ("Proof-reading ABAR"), "Sifferkorrektur" (Proof-reading of numbers), "Namnminne" (Memory for names), "Sifferminne" (Memory for numbers) and "Figuridentifikation" (Identification of figures).

In the final test battery, memory is treated as a separate factor and in addition to the tests mentioned above, two additional tests are used: "Uppskattning" (Estimation) and "Felletning" (Error search). The motives for including these are not given, and it is not clear what they are supposed to measure. In addition, a percept-genetic (PG) test is included together with an interview which aims to assess the applicant's motivation for the job. A test of capacity to process different information simultaneously is also included, and an interview by personnel from ABAR which assesses ability to cope with stress, ability to cooperate, ability to take initiative, and motivation for the air traffic controller job, i.e., many of the factors also covered by the interview by the consultants in charge of the PG test. Finally, there is an interview by personnel from the Civil Aviation Administration.

The MRU Hypothesis 1. International research and job analyses regarding the air traffic controller's profession show that a majority of the work behaviours can be described in terms of cognitive skills. The first hypothesis is that a job analysis in the Swedish work environment will replicate these results.

The MRU Hypothesis 2. The second hypothesis is that self-confidence plays an important role in coping with the critical job factors.

The MRU Hypothesis 3. The third hypothesis is that interpersonal skills play a significant part in being a skilled air traffic controller or student.

The MRU Hypothesis 4. The fourth hypothesis is that there is a significant difference in test results between those who successfully complete their controller training and those who fail.

The MRU Hypothesis 5. The fifth and final hypothesis is that training based on cognitive skills training, coaching and mentoring the students will be more effective than traditional training methods (e.g. on the job training, OJT).

Methods and Research Procedures

Job Analysis Procedures

The interviews were conducted as focused group interviews with a representative sample of ATS units. A total of 11 ATS units and 2 training units were visited. 127 air traffic controllers

participated in the focused group interviews. The interviewers who were experienced in using this method worked in pairs. Each ATS unit had been contacted in advance about the purpose of the interviews. The interviewers used interview guides prepared in advance.

The interviewees were asked about how skillful air traffic controllers coped with stressful situations or events. Both the stressful situations, or events, and the effective work behaviours were recorded. The interviewers compared notes afterwards and only notes which agreed were accepted. Almost 400 different measures to cope with stressful situations were recorded.

All responses noted were thereafter recorded and tabulated according to frequency. Thus the content of the job analysis was a frequency table of stressful events and corresponding key behaviours (the effective way how to cope with stressful event).

Control of Validity

The second step of the job analysis was to transform the responses of the interviewees into different questionnaires for different types of ATS units (i.e. TWR, TMC and ACC). The same procedures as that described above were applied to the students' working situation.

The different questionnaires were distributed to a representative sample of 158 air traffic controllers and instructors working at TWR, TMC or ACC (radar and non radar). Their task were to list, on a 7-point scale, the importance of the behaviours and how often the related situations occurred in the daily work life. Step number two was taken as a measure of the relevance or content validity of the results.

The recorded events and behaviours were compared with the causes of failures for students undergoing training. This as a test of the predictive validity of the job analysis.

The final step was to compare the results of the job analysis with data from the air traffic incident and information report system which exists at the ANS department. This step was taken to check the construct validity of the job analysis.

Analysis of the Test Battery Material

The material (Brehmer, 1993, MRU Report 7) consists of 145 students who have been admitted to air traffic controller training 1990-91. The students come from the courses starting 9007 (26 students), 9008 (24 students), 9009 (1 student), 9011 (26 students), 9107 (28 students), 9108 (24 students), and 9111 (16 students). There were 58 women and 87 men. Thirty-seven had failed and 104 succeeded, or at least not failed at the time when the evaluation of the test battery was done. Data with respect to success was missing for four students, who had taken a leave of absence from the training.

Complete data are available for only 134 of the 145 students. The number of students for which data are available varies from 134 for the selection variable with the lowest number of students to 141 for that with the highest number. It is not likely that this will have had any important effect on the conclusions.

Analysis of Relations Among Variables

Two different kinds of analyses have been performed: regression analyses and discriminant analyses. Both of these aim at assessing the extent to which it is possible to predict success in training from the various predictor variables.

Regression analysis, which shows how well success can be predicted from the predictor variables, as well as the relative importance of different predictor variables, is the standard method for this purpose. However, some objections can be directed at this method in the present case where the outcome variable is binary. Therefore, we also made discriminant analyses which show the extent to which it is possible classify the students into two groups: those who pass and those who fail the training. As we shall see, these two methods give the same results.

In these analysis, the sex of the applicant has been entered as a predictor variable in addition to the test and interview variables. The analyses are based on the 134 cases for which complete data were available.

Results

Introduction

The demanding tasks accounted for in the charts below have been compiled into five categories describing the nature of those stressful tasks. The terms used can be explained as follows.

Traffic Processing

This term is used to describe the actions taken, and the decisions made, to establish a safe and well organized flow of traffic by use of clearances, separations, applicable working methods and planning.

Coordination

This term describes the communication between air traffic controllers used to exchange information, obtain clearances, revise previous information or hand over the control of an aircraft to establish a safe and well organized flow of traffic.

Disturbances and Irregularities

This term is used to describe situations and duties when normal working methods cannot be used (e.g. technical malfunction, irregular behaviour of an aircraft etc.).

Fluctuating Workload

Description of the events and situations connected with uneven flow of traffic (e.g. high traffic intensity with a variety of performance characteristics, followed by low traffic intensity, different flight status and a mix of military and civil aviation).

Personalities and Social Skills

These terms describe how the persons interviewed perceive air traffic controllers, their personalities and social behaviours.

The results of the studies show that the reported actions and behaviours are involved with Information Gathering, Decision Making and Communication in connection with traffic processing and coordination. Actions and behaviours caused by (high) level of Ambition and (high) demands on Performance account for a large portion of the strain appearing with irregular flow and varying traffic.

ACC

The most significant behaviours in the categories "Information Gathering and Decision Making are found in the ACC function. Five of the ten most common behaviours involve Information Gathering. The problem area "Social Relations" was not awarded the same significance as behaviours more closely connected to traffic processing, which is shown by the number of behaviours that came up in the group interviews. In the ACC function the highest importance was given to behaviours dealing with coordination and traffic processing:

- Accurate, short and precise coordination with proper prioritization
- Identifying conflicts early and following up on traffic.

TMC

In the TMC function, the majority of the behaviours that are ranked high in importance or frequency appear in the areas "Decision Making and Communication." The reasons for these behaviours can mostly be found in the straining tasks in connection with coordination and traffic processing.

A similar division of work behaviours as in the ACC function appears also in the TMC function, with the difference that by comparison, it is more important to:

- Dare to say "no"
- To be, and be perceived as being, determined.

TWR

In the TWR function, the most significant behaviours are found in the category "Decision Making, followed by "Information Gathering" and demands on "Ambition and Performance." The category "Communication" was awarded lower significance than in the ACC and TMC functions.

In the TWR function, the importance of behaviours categorized under coordination decreases in favor of behaviours in the area of traffic processing. In the TWR function, the highest importance is awarded to:

- Making decisions, looking out and following up on traffic
- Working with confidence in one's ability and maintaining concentration also during periods of low traffic intensity.

Conclusions

As a description of the air traffic controller's profession and of air traffic control services, the survey largely corresponds with what the persons interviewed reported as significant behaviours in maintaining a safe and well organized flow of traffic. We can therefore conclude that the job analysis is valid as well as reliable.

Movements and changes occur in the air and/or on the ground, and the air traffic controller is expected to handle these processes in a safe and orderly manner from her or his position in the tower or control central. The air traffic controller is thus not physically situated in the surroundings where these changes occur and cannot experience the movements and changes with her/his senses. Instead, the air traffic controller must create a mental picture of the present situation, or of what the situation will be like within a limited time frame using the fragmentary information provided. As a support in constructing this mental picture, the air traffic controller has a number of technical aids: radio systems, direction finder, radar systems, data displays and monitors, telephones, telefax, telex etc.

Attribution	**Ranking**
Slow starter, unprogressive learning curve.	1
Rigid and uniform working methods.	2
Passivity, lack of initiative, inactivity, late decisions.	3
Low stress tolerance, makes mistakes in complex situations.	3
Lack of theoretical knowledge.	3
Slow worker, slow in decision making.	4
Inequality of performance.	4
Tense and nervous personality.	4
Inadequate coordination.	5
Insecure when working, doubts own decisions.	5
Insufficiant understanding of the ATC system.	5
Excessively dependant on instructor.	5
Inability to switch from low to high workload.	5
Lack of concentration.	5
Constant inability to maintain separation.	5
Insufficiant planning skills.	5
Careless, not following instructions.	5
Lack of motivation, discontinuance due to other education.	5
Total number of attributions	**33**
Average number of attributions per student	**5**

Table 1. Attribution and rank, i.e., reasons for failures during basic training between 1990 and 1993. The attributions come from 17 randomly selected students who failed to complete the training program.

The predictive validity of the job analysis. A useful basis for studying success and failure is Heider's (Hastorf, Schneider and Polefka, 1970) Theory of Attribution. The central issue of Heider's theory is viewing behaviours as caused either by environmental factors or by the individual herself. The conception of reasons also leads to predictions about future behaviours and concequences. The attribution itself becomes a deterministic prediction of the future: chance factors are not considered.

Problem area	Controller	Failed trainee
Decision making	1	1
Ambition	2	2
Information gathering	3	5
Relations	4	3
Communication	5	5
Irregularities	6	4
Technical environment	7	6
Theoretical facts	8	5

Table 2. A comparison between air traffic controllers and failed trainees, regarding ranking of problem areas.

Spearman's correlation of rank (Renyon-Haber, 1971) describes the statistical connection in ranking of problem areas between air traffic controllers and failed students. The correlation is .78.

The Construct Validity of the Analysis

One noticeable discrepancy between the findings of Mattson (1979) and the MRU project is that Mattson only found two separate activities in air traffic control services: Decision Making and Communication. The MRU project has found that, in traffic processing and coordination, Information Gathering and Processing are highly important as a preparatory stage and that Self Confidence is an important characteristic of the controller.

The interpretation made by ANS/HQ of the irregularity reports, taken from the ANS department's air traffic incident and information report system from 1991, is: "The air traffic controller assumes or expects, often as a result of indistinct or incomplete phraseology, that a pilot will act in a certain manner. The controller therefore neglects to take measures that would ensure the pilot to perform in the manner assumed by the controller."

The executives in charge of the ANS/HQ judged that all of those incidents could have been avoided if the controller had taken action to ensure that the pilots performed in the way intended by the controller.

The importance of following up on one's decisions and measures is regarded as a key behaviour in the annual analysis, published by the ANS/HQ, and also in the MRU project's job analysis.

One interpretation in the job analysis is that important behaviours in the TWR function are Decision Making, to look out and follow up on traffic, to work with confidence in one's ability and to maintain concentration even during periods of lesser workload.

This corresponds with the summary presented in the annual statistics, published by the ANS/HQ, concerning irregularity reports.

The interpretation of those reports show that the most serious incidents occur in the immediate vicinity of the airport. The final evaluation of the incidents is that operators and supervisors have not sufficiently emphazised methods of working and phraseology.

The importance of a distinct and fixed phraseology, in order to verbally express one's decisions and measures to the party or parties concerned, constitutes a key behaviour in all sources accounted for.

Relationship Between Test-Results and Training Outcome

Table 3 shows means and standard deviations for the different predictor variables. For each variable, these computations are based on data for those students for whom the result in that variable was known; i.e., the number varies from 134 to 141. In these analyses, the results for the PG test, which are reported only in terms of two categories, + (a positive value) and +/- (doubtful) have been dummy coded with 1 for the +/- category and 2 for the + category. No students with a pure value on PG (i.e., students for which the prognosis according to this test was clearly negative) had been admitted.

It is important to note that the means in the predictor variables in terms of standard scores (stanine scores in this case, with a mean of 5 and a standard deviation of 2) are generally about one unit above the mean, and that the standard deviations do not differ very much from 2. This is likely to result from the compensatory effect of summing the scores for the individual tests into scales, as is done in the selection procedure. A result of such a summation is that high scores in one variable will compensate for low scores in another variable in the scale. It is therefore possible to find minimum scores of 2, and even 1, in most of the tests. This means that the scores for the tests for those who have been admitted to air traffic controller training will not deviate too much from the scores of those who apply, as is also shown by the fact that the standard deviations do not differ very much from 2. This means that correlational analyses, the results of which are affected by the standard deviations (but not the means) of the variables, will be meaningful, and that the effects of the possible restriction of range as a result of the fact that the students have been selected on the basis of the tests being evaluated, will not be too serious. We certainly do not have to expect that the restriction is so serious that it will be impossible to detect the relations that might exist between success in training and the predictor variables.

For some of the variables, the restriction is, however, considerable. This is especially true of the interview variables and the PG test. For these variables, it is clear that the whole range of scores is nor represented in the present sample. Concerning PG, only two categories are found in that no student with a clearly negative prognosis has been admitted, and the distribution of PG scores is quite skewed, with very few +/- values. In the ABAR interviews all of the scores are between 5 and 7, and in the interviews by the Civil Aviation Authority personnel all scores are between 5 and 8. This means that it is difficult to say very much about these variables from the results of the regression and discriminant analysis.

Regression Analyses

In the regression analyses, training outcome, sex and PG have been dummy coded. The results of a regression analysis with all predictor variables are shown in Table 4.

The multiple correlation, R, is 0.413. This represents an overestimation of the strength of the relations between the predictors and the outcome in that it capitalises on sampling error. Moreover, the ratio of predictors to observations is high (19 to 134). After correction for such errors, the adjusted squared multiple correlation is 0.032 ($p < 0.25$). That is, there is no significant relation between the predictors and the training outcome and the whole set of predictors explain only 3 percent of the variance in the training outcome.

	INST (R)	INST (F)	INST (S)	SERIER (R)	SERIER (F)
M	33.46	4.48	6.43	19.32	3.02
S	2.83	2.47	1.62	3.19	2.04

	SERIER(S)	KLOSS (R)	KLOSS (F)	KLOSS (S)	KORR (R)
M	6.39	39.41	8.11	6.51	21.49
S	1.56	3.39	3.74	1.65	3.20

	KORR (F)	KORR (S)	SKEPP (R)	SKEPP (F)	SKEPP (S)
M	2.44	6.18	41.70	5.14	6.12
S	1.76	1.70	5.21	4.47	1.61

	UPPSK (R)	UPPSK (F)	UPSK (S)	WIT (R)	WIT (F)
M	5.97	127.28	2.38	5.55	12.13
S	1.29	20.86	2.26	1.82	2.20

	KRAVAT (F)	KRAVAT (S)	PLÅTMO (R)	PLÅTMO (F)	PLÅTMO (S)
M	2.18	6.26	28.11	2.64	5.59
S	1.71	1.58	4.11	2.53	1.63

	PS IF (R)	PS IF (F)	PS IF (S)	FELLET (R)	FELLET (F)
M	51.24	1.44	6.22	16.39	1.08
S	5.61	1.44	1.74	2.92	1.20

	FELLET (S)	MATRIS (R)	MATRIS (F)	MATRIS (S)	VAR 1 (R)
M	6.77	38.13	8.09	6.31	18.86
S	1.45	4.91	3.67	1.43	3.24

	VAR 1 (S)	VAR 2 (R)	VAR 2 (S)	VAR 3 (R)	VAR 3 (S)
M	6.59	12.68	6.60	18.06	6.32
S	1.36	2.39	1.46	3.32	1.40

	VAR 4 (R)	VAR 4 (S)	VAR 5 (R)	VAR 5 (S)	SUMMA (R)
M	17.97	6.26	13.32	6.99	32.76
S	3.78	1.54	2.37	1.44	4.28

	SUMMA (S)	SIMULTAN	MINNE	ABAR	LFV
M	7.07	5.71	6.04	5.55	6.01
S	1.15	1.29	1.51	0.55	0.67

	PG
M	1.92
S	0.27

Table 3. Showing means (M) and standard deviations (S) for the different predictor variables.

Variable	Coefficient	p
Intercept	0.125	0.869
ABAR-intervju	0.061	0.430
Skeppsdestination	0.057	0.032
Felletning	0.054	0.073
Instruktionsprov	0.007	0.791
Klossar	-0.026	0.328
Korrektur ABAR	0.055	0.050
Kravatt S	0.005	0.849
Kön	-0.019	0.844
LFV-intervju	-0.121	0.073
Matriser	-0.013	0.687
Minne	0.002	0.930
PG	0.124	0.416
Plåtmodeller	-0.011	0.694
PS IF figurer	0.026	0.291
Serier	0.019	0.504
Sifferkorrektur	-0.027	0.302
Simultankapacitet	0.016	0.645
Uppskattning	-0.033	0.307
WIT Pussel	-0.009	0.806

Table 4. The results of the regression analysis with all predictor variables.

The best predictors are "Skeppsdestination" (B=0.057, p=0.032), "Korrektur ABAR" (B=0.055, p=0.05), "Felletning" (B=0.054, p=0.073) and the interview made by the personnel from the Civil Aviation Authority (B=-0.121, p=0.073). The latter variable has a negative weight; however, applicants given a high rating on the basis of this interview are less likely to succeed. Further support for the conclusion that these are the most powerful variables for predicting the training outcome is given by the results of a stepwise regression analysis which selected three of these variables ("Skeppsdestination", "Korrektur ABAR" and the interview by the Civil Aviation Authority personnel). The multiple correlation for this stepwise regression was R = 0.334, R2 adjusted = 0.091, F 3/130 = 5.437, $p < 0.01$).

Discriminant Analysis

The discriminant analysis was performed with the same predictor variables as the regression analysis above. The results from the initial F-tests for these variables agreed with those from the regression analysis (as would be expected) in that significant F-values were obtained for "Skeppsdestination" (F 1/132 = 18.885, $p < 0.01$) and "Korrektur ABAR" (F 1/132 = 4.248, $p < 0.05$). As in the regression analysis, the results for the interview by personnel from the Civil Aviation Authority (F 1/132 = 3.429, $p < 0.07$) and "Felletning" (F 1/132 = 3.586, $p < 0.07$) were close to significance. The discriminant function correctly identified 12 out of the 36 who failed and 89 out of the 98 who succeeded in the training; i.e., 101 (75 percent) applicants were correctly identified. This should be compared with the number to expected if the predictor variables are ignored and only the base rates are considered; i.e., the number of correct

classifications that would be expected randomly. This yields an expected rate of correct classifications of 62 percent or 83 students. Thus, the discriminant function improves the selection by 18 cases, compared to no selection procedure at all. Thirty-three students are incorrectly classified, compared to 47 that would be expected on a random basis. This agrees with what would be expected on the basis of the uncorrected multiple correlation between the predictors and the training outcome of R = 0.413. As noted in the discussion of the regression results, the sample estimate represents an overestimation of the possibilities of predicting the outcome; this is true also of the results of the discriminant analysis. Unfortunately, there is no procedure for estimating a discriminant function corrected for sampling errors comparable to the procedure for the multiple correlation. However, in the present case with a binary outcome variable, multiple regression and discriminant analysis are basically the same, and we should therefore expect that after correction for sampling errors, we should have the same decrease in effectiveness; i.e., we should expect that the ability to make correct classifications using the discriminant function should decrease by about 80 percent after sampling errors have been taken into account. A reasonable estimate of the improvement in the number of correct classifications of training outcomes from using the current set of tests and interviews is 3-4 cases (about 3 percent). This is the same estimate as that which we obtained from the regression equation.

Discussion

The Main Task

The main task for this phase of the MRU project was to evaluate current screening and training procedures and create recommendations aiming at a reduction of the present span between intake of students and output of examined new air traffic controllers.

The Analysis

The task force chose to conduct a job analysis based on the critical incident technique. The result shows that 300 reported key behaviours could be catalogued into 5 groups:

- Decision making
- Self confidence
- Information gathering and processing
- Social relations
- Communication

This result verifies the three hypothesis stated about the air traffic controller profession: Behaviours which are related to self confidence are mostly reported in connection with unexpected events and variabilities. The results from the job analysis have been compared to attributions for failure in the basic controller training. It has also been compared to incidents that have occured in actual operations according to the current official report system. Both students and controllers fail to perform the key behaviours at a sufficient level.

The training process requires the students to practise key behaviours from the very first day, aiming at minimizing the number of errors to reach a full performance level, and finally to reach a mastery level. Today an uneven learning curve is the most frequent cause of failure during the basic controller training.

At Stockholm ATS and Arlanda ATS units in Sweden, attempts have been made to improve on current methods of basic training. The results from this attempt to apply modern training techniques (for example to use programmed skill training and to transform the instructor into a mentor and a coach), is now the most promising measure taken to improve the outcome of basic air traffic controller training. To quote one of the members of the task force, professor Berndt Brehmer: "It is astonishing how little effort is made in general to train and develop an operator in a high tech environment by modern training technology, and how much one still relies on an old fashioned on the job training provided by a more experienced fellow-worker."

Present rate and variation in span in the outcome of the Swedish basic air traffic controller training can only partly be explained by inadequate psychological tests and screening procedures. To reduce the uncertainty in the outcome, it is important to improve training and learning of key behaviours for the air traffic controller work, as well as to develop screening methods with high reliability and validity. This will give a prompt and positive result. Efforts must also be made to create a continuing job analysis in order to keep up with a changing technology and maintain screening and training methods with the highest possible effectiveness. An important prerequisite for a successful training result is an efficient selection procedure based on a sophisticated chain consisting of information/introduction/skill tests assessing the substance of the most important groups of key behaviours.

The Relationship Between Test-Results and Training Outcome

The results of the analyses presented above show that it is not possible to predict the outcome of the training on the basis of the variables used in the selection procedure.

One possible reason for these depressing results, and this is true both for the regression results and those from the discriminant analysis, is that these results are based on data only for those who were admitted to the training; that is, we have a classical case of restriction of range. To ascertain the effects of this, we need to look at the standard deviations of the various predictors for the sample used in the calculations. The relevant results are shown in Table 3. As already noted in the discussion of these results, the restriction is not as severe as might have been expected. The standard deviations for the predictor variables are between 71 percent and 91 percent of those for the unselected sample used to determine the stanine scores, and for most of the variables, the standard deviations for our sample are about 80 percent of those in the unselected sample. Moreover, we have the full range of the predictor variables for many of the variables; the lowest values are 1 and 2 for many of the variables. There is therefore little doubt that we would have been able to detect the relations that might exist between the predictor variables and the training outcome. The fact that we find very few significant relations, and that the correlations that we have found are very low, can therefore hardly be explained in terms of restriction of range. Instead, it seems more reasonable to assume that the results express real deficiencies in the selection procedure. That is, the predictor variables are not very powerful predictors of the training outcome. This is hardly surprising in view of the fact that these variables have been selected on the basis of a very general job analysis without real standardisation and statistical evaluation; i.e., the tests have not been chosen on the basis of an

empirical evaluation of the actual predictive validity of the tests. This means that there was no reason to expect that the selection procedure would be very effective.

One could, of course, argue that the selection procedure concerns the job as an air traffic controller and not the training. It may well be that the training makes demands that differ from those of the job and that an evaluation of the selection procedure in terms of the training outcomes is not quite relevant. To answer this question, we need a more penetrating analysis of the demands that the training courses actually make compared to those that the job makes. At the present time, we do not have the data required for such an analysis.

Another objection is that the analysis may rely on the wrong model. The present analysis is based on a model where the probability of success in training is assumed to be a monotone function of performance in the selection variables (see Figure 2). That is, this model makes the reasonable assumption that if some ability is required, more of that ability leads to a higher probability of success than less of the ability.

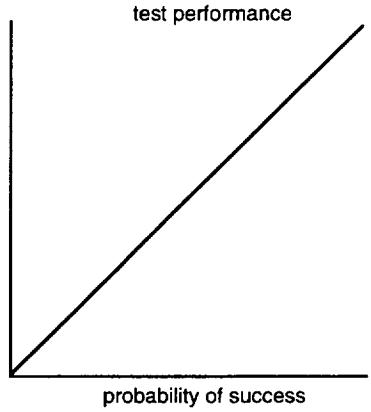

Figure 2. Basic model for the analysis in this study. This model assumes that there is a monotone relation between training outcome and test performance.

An alternative model is illustrated in Figure 3. This alternative model assumes that the training only requires some minimum ability, and that all students having at least this minimum ability will have the same probability of succeeding in the training course.

If this model is true, the possibilities of detecting relations between the training outcome and the selection variables would be limited, especially if the students in the training course had been selected so that all of them had values exceeding the critical value. In the present case, this does not seem to be a very serious problem, however, since the full range of values is represented for most variables in the present sample. Thus, it should have been possible to detect whatever relations might have existed between the test variables and the training outcome, even if model 2, rather than model 1, would have been valid for the present data. Moreover, when measurement error is added, model 2 will generally be impossible to distinguish from model 1.

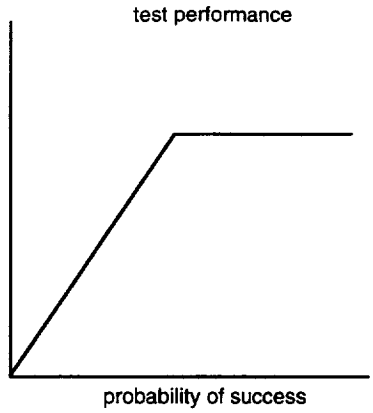

Figure 3. Alternative model. This model assumes that the probability of succeeding in training increases up to some critical values and that it then stays constant at the same level.

The analyses have been based on the individual tests rather than the scales used by ABAR in the actual selection. The reason for that is that our analyses yielded no support for these scales in that we found that these scales were intercorrelated, while the tests included in the scales were not intercorrelated as they should have been. Moreover, a principal components analysis failed to yield the scales as components. Thus, there was little support for the usefulness of these factors. Additionally, regression and discriminant analyses based on the scales used by ABAR did not give better results than the analyses based on the individual predictors.

The results with respect to PG deserves special comment. This is the only variable included in the selection procedure on the basis of an empirical evaluation procedure. In this procedure, the PG test was given to the applicants, but not used for the selection. That is, the evaluation concerned the extent to which this test could improve the selection over above what could be achieved with the original test battery. The results were quite encouraging, but we must now conclude that the conclusions from the original evaluation were overly optimistic. Thus, Svensson and Trygg (1991) concluded that it should be possible to decrease the proportion of students failing the air traffic controller training to less than 10 percent if the PG test was used. As shown in the present analysis, this has not been the case. Even when the PG procedure is used, the proportion of students who fail is 26 percent.

It was, however, not realistic to expect that one would have as good results with the present sample as with the standardisation sample used to determine what PG-variables should be used for the selection. First, the initial evaluation did not take into account the total effectiveness of the selection procedure with PG as one of many selection variables. The value of PG in such a procedure is dependent, not on the correlation between this variable and the training outcome (which is what was reported in Svensson and Trygg, 1991), but upon its unique contribution, which is dependent on the partial correlation between PG and training outcome, after its correlations with other selection variables have been taken into account. That such intercorrelations exist is demonstrated in the present sample, despite the severely restricted variation in the PG scores. Such intercorrelations decrease the weight that the PG results will receive in the final selection.

Second, one must expect a certain shrinkage in correlation when the test is used for a new sample because the values obtained for the first sample capitalised on sampling errors. In the present sample, the unique contribution from PG is far from significant. However, the extremely skewed distribution of PG values makes this correlation suspect.

A possibility of evaluating the effectiveness of the PG procedure is to compare the failure rates before and after the introduction of this test. The relevant comparison here should be with the failure rate for the 1980s, when the mean failure rate was 33.9 percent (with considerable variation among courses). In the 1990s, after the introduction of the PG procedure, the mean failure rate so far has been 26 percent, although this may well be an underestimation because not all students have completed their courses. That is, not all students have yet had a chance to fail. With this in mind, the maximum estimate of the improvement from PG would be 7.9 percentage units, but this would assume that all of the decrease in the failure rate from the 1980s to the 1990s can be attributed to the introduction of the PG procedure. This seems unlikely, especially in view of the fact that the decrease in the failure rate from the 1970s to the 1980s was about 6 percentage units (from a failure rate of 40 percent in the 1970s to a rate of about 34 percent in the 1980s) without any new selection procedures.

In the regression and discriminant analyses, three variables stand out. One of these, the results from the interview conducted by personnel from the Civil Aviation Authority, receive a negative weight. That is, they are systematically wrong: students with a high rating in these interviews perform systematically worse than those with a low rating. This suggests that this procedure must be improved.

Only two of the selection variables have systematic relations with training outcome: "Skeppsdestination" and "Korrektur ABAR". The former of these is supposed to measure flexibility, and the latter is to measure carefulness. The correlations are low, however, and they may well have been produced by chance. Therefore, one should not rely too much on these results until they have proved valid also for other samples.

Conclusions

The present evaluation of the selection procedure is clearly limited, first because it is based on a limited sample, and second because it is based on the results for a group that has been admitted to the training on the basis of the selection procedure that is being evaluated. The restriction of range problem does not seem to be as severe as one might have suspected, however. It should therefore have been possible to detect whatever relations might exist between the selection variables and the training outcome. We must therefore conclude that the fact that it has been hard to find such relations probably means that they do not exist. There are therefore good reasons to reconsider the present selection procedure. It is not possible to decide whether it is possible to design a better selection procedure on the basis of the data we have today. For this, we need a careful analysis of the air traffic controller's job to determine what demands that this job makes, and how these demands can be met by means of selection and training.

Future ATS Systems

Coming automated ATS systems cannot replace the human controller. But manual repetetive work can be eliminated and in that way facilitate information seeking and information collection.

If a new technology or a new system is to be introduced, it is fundamentally important to be assured that the operators accept the new technique, and that the new technique will create opportunities for them to improve their performance. The controllers must also be informed in advance in what way they will be trained to achieve this new standard of performance.

In our view, a continued automation of the air traffic controller's work will only further emphasize the importance of adequate training to execute "new" key behaviours.

References

ABAR. (1978). Förslag till bedömningsvariabler vid psykologisk lämplighetsprövning av flygledarspiranter, våren (Proposal for assessment of variables concerning psychological aptitude tosts for air traffic controller applicants).

Andersson, B. (1993) Sammanställning av enkätsvar som inkommit från flygledare. MRU-projektet (Synthesis of questionnaire answers from operational air traffic controllers). Delrapport nr 4. Luftfartsverket.

Brehmer, B. (1993) Prestation vid urvalet och utbildningsresultat i flygledarutbildning. MRU-projektet (Performance in selection and training results in air traffic controller training). Delrapport nr 7. Psykologiska Institutionen, Uppsala universitet.

Cambell, J. P. (1988). Training design for performance improvement. In J. P. Cambell., R. J. Cambell. *Productivity In Organizations*. Jossey-Bass. USA.

Cullen, J., & Hollingum, J. (1987). *Infr total kvalitet* (Introduction of total quallity). Konsultfölaget AB.

Dolgin, D. L., & Gibb, G. D. (1989). Personality Assessment in Aviator Selection. In Richard S. Jensen (Ed.). *Aviation psychology*. Gower Publishing Company.

Flannagan, J. C. (1954). The critical incident technique. *Psychological Bulletin. Vol 51*. Sid 327- 358.

Haglund, R., Andersson, B., Backman, B., & Sundin, O. (1993). Presentation av bakgrund till MRU-Projektet samt planering för dess genomförande (Work analysis concerning air traffic controllers and air traffic contller students). MRU-projektet. Delrapport nr 1. Luftfartsverket.

Haglund, R., Andersson, B., Backman, B., Sundin, & O. Rapport. (1993). Arbetsanalys flygledare och flygledarelever. MRU-projektet. Delrapport nr 2. Luftfartsverket.

Haglund, R., & Backman, B. (1993). Studiebesök på FAA Huvudkontor och Mike Monroney Aeronautical Center (Study visits at FAA Washington Headquarters and Mike Monroney Aeronautical Center). MRU-projektet. Delrapport nr 3. Luftfartsverket.

Haglund, R., Andersson, B., Backman, B., & Sundin, O. (1993). Resultatredovisning, urval av och utbildning av flygledarelever (Results concerning selection and training procedures for air traffic controller students). MRU-projektet. Delrapport nr 5. Luftfartsverket.

Haglund, R., & Backman, B. Analys av samstämmighet mellan arbetsanalyser flygledare och operativa systemets avvikelserapportering (Analysis of concordance between work analysed of the air traffic controller's profession and data from the air traffic incident and information report system at the ANS Department). MRU-projektet. Delrapport nr 6.

Hastorf, A. H., Schneider, D. J., & Polefka, J. (1970). *Person perception.* Addison Wesley.

Hopkin, V. D. (1989). Cognitive Demands of Automation in Aviation. In Richard S. Jensen ed: *Aviation psychology.* Gower Publishing Company.

Hopkin, V. D. (1980). The measurement of the Air Traffic Controller. *Human Factors.*. 22(5), 547-560.

Josefsson, B. (1992) Information - processing in air-traffic controllers. 3-uppsats. Institutionen för pedagogik och psykologi. Universitetet i Linköping.

Latham, G. P., Wexley, K. N. (1981). *Increasing productivity through performance appraisal.* Addison-Wesley.

Manning, C. A., Kegg, P. S., & Collins, W. E. (1989). Selection and screening programs for air traffic control. In Richard S. Jensen (Ed.). *Aviation psychology.* Gower Publishing Company.

Mattson, J. (1979). *Kognitiva Funktioner vid flygtrafikledning* (Cognitive functions in air traffic control). Lund University Sweden. Department of Psychology. Dissertion Series.

McNemar, Q. (1969). *Psychological Statistics.* John Wiley and sons.

Runyon, R., & Haber, A. (1971). *Fundamentals of behavioral statistics.* Addison-Wesley.

Smith, R. C. (1973). Job attitudes of air traffic controllers: a comparison of three air traffic control specialties. Washington, DC: U.S. Department of Transportation. Federal Aviation Administration.

Sundin, O., Blyckert, J., Hansson, B., & Landerstedt, B. (1993). Utbildning av flygledare på ATS Arlanda och ATS Stockholm (Training of air traffic controller students at ATS Arlanda and ATS Stockholm). MRU-projektet. Delrapport nr 9. Luftfartsverket.

Svensson, B & Trygg L. (1991). Personlighetskarakteristika hos elever i flygledarutbildningen (Personality characteristics of candidates for air traffic controller training). Rapport från Lunds universitet, Institutionen för tillämpad psykologi.

Sörli, N. (1992). Frafall under flygelederutdanning (Loss of students during air traffic controller training). Examensarbete. Universitetet i Oslo 1992.

Wanous, J. (1980). *Organizational Entry. Recruitment, Selection and Socialization of Newcomers.* Addison-Wesley Publishing Company.

Wing, H., & Manning, C. A. (1991). Selection of air traffic controllers: complexity, requirements and public interest. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.