



## A procedure of verification of comprehensive tests for selection of candidates for operators of mobile robot

Krzysztof Sapiecha<sup>1\*</sup>, Barbara Łukawska<sup>2</sup>, Mariusz Bedla<sup>2</sup>

<sup>1</sup>*Department of Computer Science, Cracow University of Technology,  
ul. Warszawska 24, 31-155 Kraków*

<sup>2</sup>*Department of Computer Science, Kielce University of Technology,  
al. 1000-lecia Państwa Polskiego 7, 25-314 Kielce*

### Abstract

This paper presents a part of research concerning an application of virtual reality for training operators of mobile robots (mobots). Mobots are often used for exploring areas dangerous or hardly accessible. It is obvious that the operator should not be a common person. In the paper a procedure of verification of comprehensive tests such as IQ test for initial selection of candidates for mobot operators is given and evaluated.

### 1. Introduction

Mobile robots (mobots) are often used for exploring areas dangerous or hardly accessible [1, 2]. A mobot is usually controlled remotely by an operator. It is obvious that the operator should not be a common person. The operator should detect all objects in the area fast and precisely. Moreover, the perfect operator should make wise decisions and be a good strategist. The question arises how to select such a person from a group of candidates applying for this position.

---

\*Corresponding author: *e-mail address*: [k.sapiecha@tu.kielce.pl](mailto:k.sapiecha@tu.kielce.pl)

Our previous research proved that exercises on a simulator of mobot touring in virtual reality are suitable for this task [2]. However, such a method is still expensive, time-consuming and difficult to apply when a group of candidates is large.

An objective of this research is a cost-efficient method of training candidates for operators of mobots or other complex machines or vehicles. In this scope there are many unsolved problems. One of them consists in developing a procedure of initial selection of the candidates who should not be trained. As not everybody can be a musician so not everybody can be an operator.

One way to solve the problem could be initial evaluation of candidates using a comprehensive test, IQ test for example. Then a small group of the best candidates would be selected for further training. This could work correctly if the comprehensive test was properly chosen.

In the paper a procedure of verification of comprehensive tests for initial choice of candidates for mobot operators is given. Section 2 contains an outline of our methodology. Section 3 describes the comprehensive tests used in the research to select best candidates for training. In Section 4 the procedure is described. The paper ends with conclusions.

## 2. Methodology

In our research a simulator similar to the flight simulator is used [2, 4, 5, 6]. The candidate can drive virtual mobot on a board (in a simulated room). The simulator uses computer generated graphics instead of real images that would be taken by the mobot in real room. The operator can move the mobot forward, backward, and turn it left or right. The mobot takes images only on demand of the operator. The main task of the operator is to discover all changes in the room in limited time. The simulator collects data, evaluates every operator move and calculates quality of monitoring ( $Q$ ).  $Q$  depends on the number of moves, the number of correct and incorrect discoveries, the number of photos taken, etc [7].

In our previous experiments the candidates showed different predispositions to be mobot operators [2]. It was also proved that mobot operators can be successfully trained in virtual reality [4]. Hence, the straight approach to the problem is to train all candidates on the simulator and then to select the best one. However, such a method is expensive, time consuming and hard to organize. Initial selection of candidates should be applied. It would cut the costs and shorten the time of the training. The selection should be quick, easy and reliable. Only the candidates who are chosen during the selection would be trained.

It was noticed in [7] that there is a relationship between the results of exercises on the simulator and those of comprehensive tests such as IQ test and average mark if the candidate is a student. Therefore, validity of these two tests for initial selection of the candidates is firstly checked. To this end a threshold  $T$  for each of the tests is defined. The candidates having results lower than  $T$  are rejected from the training. However, this may not lead to satisfactory selection of the candidates. Hence, next a procedure of two-way evaluation of the candidates is proposed.

At the beginning of the procedure a comprehensive test is applied to a group of candidates. On the basis of the test results the candidates are ranked. According to a cost factor some percentage of the candidates lowest in the ranking are marked as 'rejected' from the training. The rest of the candidates are marked as 'chosen' for further training. That ends the first way of evaluation of candidates. Next, the other one begins. The initial partition of candidates is verified with the help of exercises on the simulator. The training of the candidates on the simulator starts. After every step of the training all candidates are evaluated (their  $Q$ s are calculated and ranked). The worst candidates are marked as 'not go' ones. The rest of the candidates are marked as 'go' ones. Finally it is checked to what extent the 'rejected' candidates fit the 'not go' ones and the 'chosen' candidates fit the 'go' ones. The test may be accepted if it rejects almost all 'not go' type candidates.

### 3. Comprehensive tests

IQ test is a widely acceptable comprehensive test of human abilities. Hence, it was taken into consideration, first. Moreover, the candidates who took part in our experiments were students. Average mark of a student from a study indicates a lot about his diligence, brightness, resistance to stress, flexibility etc. It is a long term test, so it does not depend on transient factors like mood, weather conditions, etc. Hence, it was taken into consideration, too.

Before the training every candidate lets us know his average mark from a study and carries out IQ test. Then, the IQ test and the average mark criteria are analyzed. Based on each of the criterion separately the candidates are divided into two subgroups. Less efficient candidates, with the results less than  $T\%$  of maximum, are assigned as 'rejected'. The rest of the candidates are assigned as 'chosen' for further training. Next, all candidates are trained in the same way. After the training the 'rejected' candidates and the 'chosen' ones are separately categorized as 'poor', 'mediocre', 'good' and 'very good'. The categories cover equally the whole range of  $Q$  (here from -100 to 300). Numbers of the candidates in each of the category according to results of the IQ test and

their average marks from a study are given in Figs 1 and 2 respectively (where the threshold level  $T$  equals 80 or 78, 84 or 82, 88 or 86 and 92 or 90).

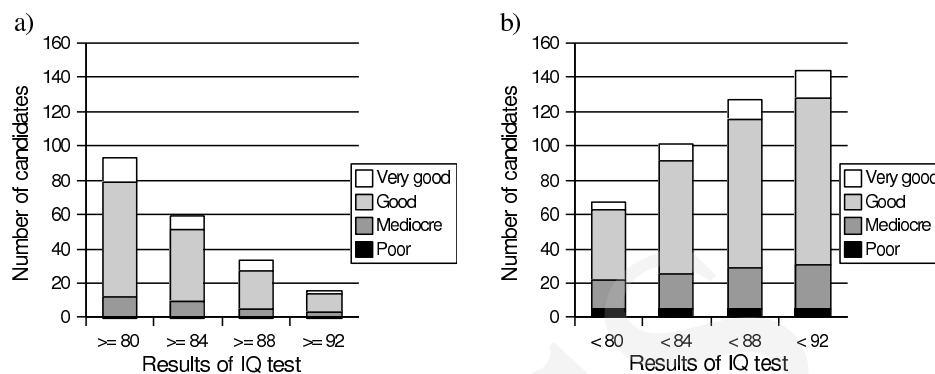


Fig. 1. Categorization according to the results of IQ test: a) 'chosen', b) 'rejected'

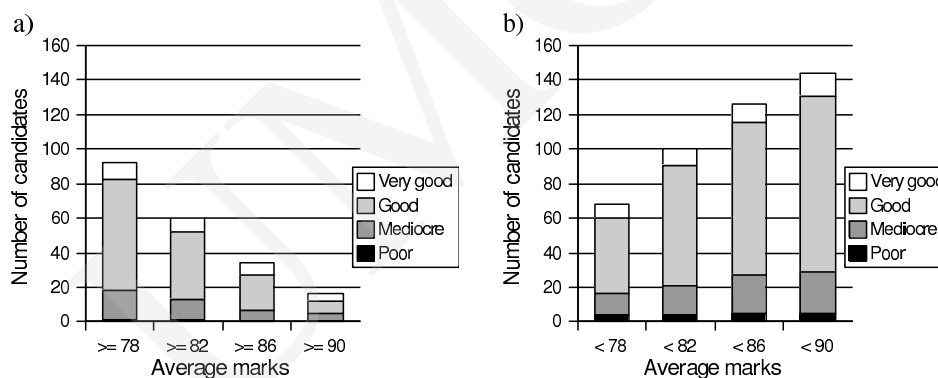


Fig. 2. Categorization according to the average marks: a) 'chosen', b) 'rejected'

Table 1 summarizes the results of the experiment. It is seen that the tests reject all or almost all 'poor' and many 'mediocre' candidates but also some 'good' and 'very good' ones. A simple method how to retain all 'good' and 'very good' candidates and reject all 'poor' and 'mediocre' ones is not known. In this case rejecting all 'poor' candidates might be satisfactory, in part.

Table 1.

Selection method	Thresh hold (%)	Number of candidates (%)		Quality of monitoring		Number of candidates in 'chosen' group (%)				Number of candidates in 'rejected' group (%)			
		chosen	rejected	chosen	rejected	poor	mediocre	good	very good	poor	mediocre	good	very good
IQ test	80	58.13	41.88	153.49	113.59	0	12.9	72.04	15.05	7.46	25.37	61.19	5.97
	84	36.88	63.13	151.64	128.10	0	15.25	71.19	13.56	4.95	19.80	65.35	9.90
	88	20.63	79.38	154.2	132.25	0	15.15	66.67	18.18	3.94	18.9	67.72	9.45
	92	10.00	90.00	148.38	135.49	0	18.75	68.75	12.50	3.47	18.06	67.36	11.11
Average mark	78	57.5	4.5	141.53	130.36	1.09	18.48	69.57	10.87	5.88	17.65	64.71	11.76
	82	37.5	62.5	142.47	133.37	1.67	20.00	65.00	13.33	4.00	17.00	69.00	10.00
	86	21.25	78.75	148.98	133.49	0	20.59	58.82	20.59	3.97	17.46	69.84	8.73
	90	10.00	90.00	146.22	135.73	0	31.25	43.75	25.00	3.47	16.67	70.14	9.72

It seems that setting the IQ test threshold to be 88% or 92% and the average mark threshold to be 86% or 90% is justified. Setting the thresholds lower could create a too numerous subgroup of 'chosen' candidates. Setting the thresholds higher could lead to a very small subgroup which, in turn, could cause anomalies. Selected thresholds provide a suitable number of candidates (about 20% – 10% of the initial group), and among them many candidates who would be 'good' or even 'very good' operators after the training.

#### 4. Verification of comprehensive tests

The thresholds determined in the previous section select rather a small subgroup of the candidates without 'poor' ones. This sharply cuts costs of the training. However, there is a number of 'mediocre' candidates in the subgroup and a number of 'good' and 'very good' ones out of the subgroup which suggests that the tests should be applied in another way. To verify whether or not the tests described in the previous section could be used in general for the initial selection of candidates and how the following procedure is applied:

1. After every step of the training candidates are evaluated. For every candidate his  $Q$  is calculated.
2. Then the candidates are ranked according to two different criteria.
  - (a) Depending on  $Q$  the candidates are assigned to one of the four categories: 'poor', 'mediocre', 'good' and 'very good' [4].
  - (b) Depending on a cost factor  $C$ , in each step of the training  $P\%$  of the candidates (20%, for example) with the lowest  $Q$  is assigned to the group of 'not go' candidates and the rest of the candidates are assigned to the group of 'go' ones. This partition of the candidates is shown in Fig. 3.

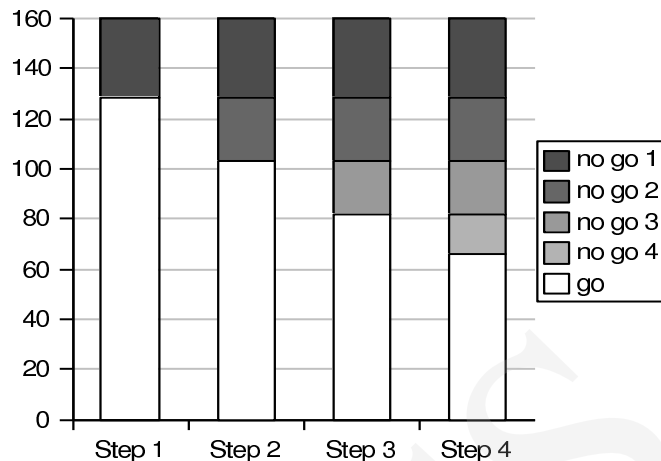


Fig. 3. 'Go' and 'not go' candidates after every step of the training ( $P = 20$ )

After four steps of the training average  $Q$  of all 'not go' candidates and the average  $Q$  of the last 'go' ones are compared. The same is done for different  $P$  (10%, 20%, 30% and 40%).

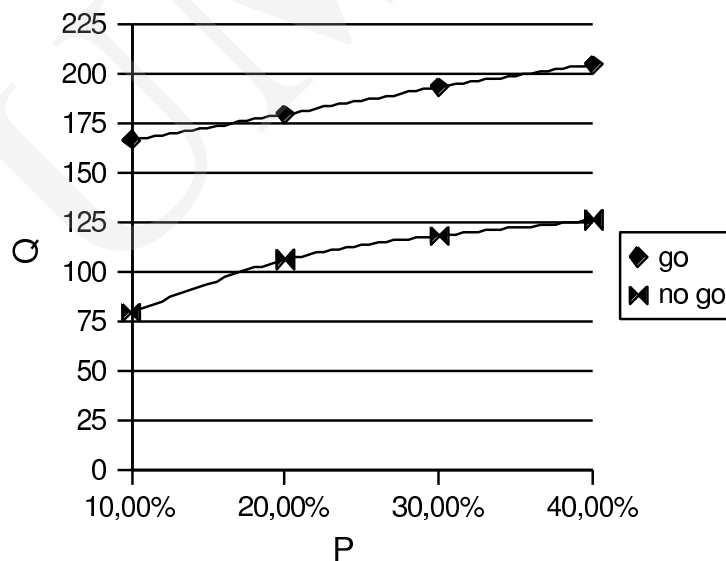


Fig. 4.  $Q$ s for 'go' and 'not go' as a function of  $P$

As it is seen in Fig. 4  $Q$ s are always considerably higher for 'go' candidates than for 'not go' ones. Moreover, the higher  $P$  the higher average  $Q$  in both 'go' and 'not go' subgroups.

Fig. 5 shows the distribution of 'poor', 'mediocre', 'good' and 'very good' candidates over the subgroups of 'go' and 'not go' candidates. Comparing Figs 1 and 2 with Fig. 5 one can observe that categorization of candidates based on their  $Q$ s (after the training) fits only to some extent their categorization based on the results of their initial tests. To estimate validity of the comprehensive tests for the initial selection of candidates average  $Q$  in 'chosen' (Figs 1a and 2a) and 'go' (Fig. 5a) subgroups are compared.  $T$  (Figs 1 and 2) and  $P$  (Fig. 5) are paired in such a way that cardinalities of these subgroups are similar. These are shown in Table 2.

Using  $T\%$  and  $P\%$  as shown in Table 2 we got subgroups including approximately 60%, 40%, 20% and 10% of the total number of candidates. The average values of  $Q$  in every subgroup are shown in Table 3.

When no selection is applied then average  $Q$  for all the candidates is about 136.78 and average  $Q$  for 10% of the best candidates is about 221.36.

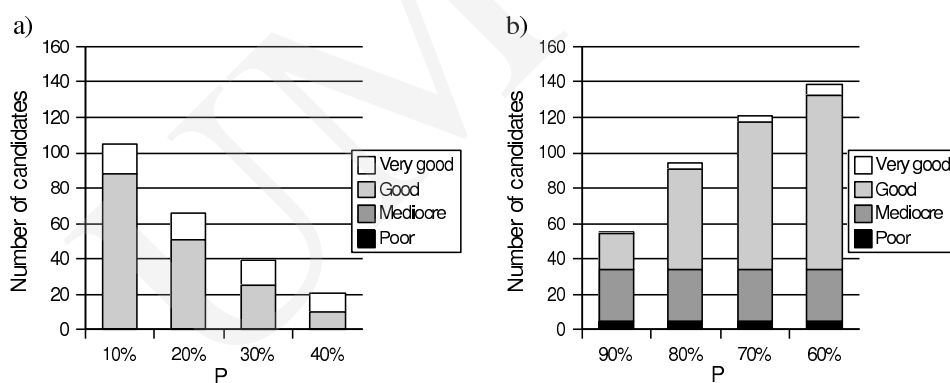


Fig. 5. Categorization according to  $P$ : a) 'go' group, b) 'not go' group

Table 2.

Rejection method							
IQ test		Average mark		Rejection after first step		Rejection after each step	
T (%)	Number of candidates (%)	T (%)	Number of candidates (%)	P (%)	Number of candidates (%)	P (%)	Number of candidates (%)
78	57.50	80	58.13	40	60.00	10	65.63
82	37.50	84	36.88	60	40.00	20	41.25
86	21.25	88	20.63	80	20.00	30	24.38
90	10.00	92	10.00	90	10.00	40	13.13

Rejection of the candidates after the first step of the training is an intermediate method between initial selection and rejection after each step of the training. The first step of the training is used to familiarize candidate with the simulator. Rejection after the first step is more expensive than the initial selection because every candidate must be partly trained. However, it is much less expensive than rejection of the candidates after each step of the training because much more candidates are rejected very early (60% compared to 20%, row 2 in Table 2).

Table 3.

Rejection method				
Numer of candidates (%)	IQ test	Average mark	Rejection after first step	Rejection after each step
~60 %	153.49	141.53	149.16	166.71
~40 %	151.64	142.47	155.15	180.06
~20 %	154.20	148.98	164.81	193.58
~10 %	141.53	146.22	170.20	205.14

The next factor that estimates validity of the selection procedure is distribution of the best candidates among 'go' or 'chosen' subgroups. The distributions of the best one (B1), the best three (B3) and the best 10% (B10) in 'go' or 'chosen' subgroups are shown in Table 4.



Table 4.

	Rejection method											
	IQ test			Average mark			Rejection after first step			Rejection after each step		
Number of candidates (%)	B1	B3	B10	B1	B3	B10	B1	B3	B10	B1	B3	B10
~60 %	0	2	8	1	3	12	1	3	14	1	3	16
~40 %	0	1	7	0	2	7	1	2	9	1	2	14
~20 %	0	0	6	0	1	5	0	1	6	1	2	13
~10 %	0	0	4	0	0	1	0	0	2	1	2	10

The first two rows of Tables 3 and 4 seem to contain the most promising numbers. These guarantee considerable reduction of costs of the training (about 50%) and still very good candidates for training.

## 5. Conclusions

As follows from our research the comprehensive tests work quite correctly if  $P$  equals about 50% for the initial selection of candidates. Surprisingly, almost half of the candidates may be initially rejected with no risk to lower the final result of the training considerably. In such a case not much of the average monitoring quality  $Q$  of the best candidates at the end of the training is lost. Also, not so many candidates who would become very good operators in the course of training are dropped. Besides, even such simple initial selection may lead to significant reduction of costs of the training (about 50-60%). On the other hand, it should be noticed that the more candidates are dropped, the more good operators are dropped, too. Hence,  $P$  should not exceed 60%.

A big group of candidates were exercised in the research (160 persons). Hence, the conclusions drawn are valid, despite the fact that the procedure was verified only once. The selection procedure using IQ test gives results better correlated with those of the training. Correlation factors prove so (0.14 between the average mark and the results of IQ test; 0.42 between the results of IQ test and the value of  $Q$  after the training; 0.13 between the value of  $Q$  after the training and the average mark).

Rejecting candidates after the first step of the training is an intermediate method between the initial selection and rejection after each step. In this case, all candidates undertake one step of training but many of them are dropped

just after that. It works better than comprehensive tests do. This might be a good approach when a group of candidates are not so large.

Rejection after each step of the training gives the best results. However, even using this, a very expensive method, some 'very good' operators are dropped. This could be explained in two ways: some of the candidates get poor results initially, but if they get a chance, they can learn much and catch up with the best ones. On the other hand, some candidates get worse results in the consecutive training stages, and fall to lower categories. This is a proof that candidates move from one subgroup to another.

The selection procedure may be applied to candidates for operators of mobots or other complex machines or vehicles. After verification comprehensive tests may be applied to candidates of any kind.

### References

- [1] NASA, <http://www.nasa.gov>
- [2] Polski Wortal Robotyki, <http://www.asimo.pl>
- [3] Sapiecha K., Lukawska B., Paduch P., *Experimental Data Driven Robot for Pattern Classification*, Annales UMCS – Informatica AI 3 (2005).
- [4] Lukawska B., Paduch P., Sapiecha K., *An application of virtual reality for training and ranking operators of mobile robot*, Annales UMCS – Informatica AI 5 (2006).
- [5] Sapiecha K., Bedla M., Lukawska B., Paduch P., *Computer-based system for training and selecting mobile robot operators - evolving software tools*, Konferencja 'Informatyka - badania i zastosowania', Kazimierz Dolny (2007).
- [6] Bedla M., Lukawska B., Sapiecha K., *Software architecture for a system of remote training mobot operator*, Konferencja 'Advanced Computer Systems and Networks: Design and Application' (ACSN 2007), Lwów, Ukraina (2007).
- [7] Sapiecha K., Lukawska B., Bedla M., *Computer-based system for training and ranking mobot operators - selection procedure*, Konferencja 'Informatyka - badania i zastosowania', Kazimierz Dolny (2007).