PREDICTIVE PERFORMANCE ASSESSMENT: TRAIT AND STATE DIMENSIONS SHOULD NOT BE CONFUSED.

N. Pattyn[1,2,4], P.-F. Migeotte[2], J. Morais[4], R. Cluydts[1], E. Soetens[1], R. Meeusen[3], G. de Schutter[3], E. Nederhof[3], R. Kolinsky[4]

[1]*Dept Cognitive & Biological Psychology, Vrije Universiteit Brussel ;* [2]*Royal Military Academy, Brussels;* [3]*Dept Human Physiology – Exercise and Sports Lab, Vrije Universiteit Brussel,* [4]*Unité de Neurosciences Cognitives, Université Libre de Bruxelles*

npattyn@vub.ac.be

ABSTRACT

One of the major aims of performance investigation is to obtain a measure predicting real-life performance, in order to prevent consequences of a potential decrement. Whereas the predictive validity of such assessment has been extensively described for long-term outcomes, as is the case for testing in selection context, equivalent evidence is lacking regarding the short-term predictive value of cognitive testing, i.e., whether these results reflect real-life performance on an immediately subsequent task. In this series of experiments, we investigated both medium-term and short-term predictive value of psychophysiological testing with regard to real-life performance in two operational settings: military student pilots with regard to their success on an evaluation flight, and special forces candidates with regard to their performance on their training course. Our results showed some relationships between test performance and medium-term outcomes. However, no short-term predictive value could be identified for cognitive testing, despite the fact physiological data showed interesting trends. We recommend a critical distinction between "state" and "trait" dimensions of performance with regard to the predictive value of testing.

1. INTRODUCTION

A major aim of performance investigation is to predict real-life performance, which is why both ESA [1] and NASA [2] have described the need to validly and reliably detect potential performance decrement as absolute requirements to manned long-duration missions. Whereas the predictive validity of such assessment has been extensively described for medium-term to long-term outcomes, as is the case for cognitive performance selection of student pilots for example, similar evidence is lacking regarding the *immediate* predictive value of cognitive testing, i.e., whether these results reflect real-life performance on an immediately subsequent task. Furthermore, whereas selection procedures are derived from population-based approaches, real-time monitoring of performance is often meant to be individual, which is an additional call for caution before concluding results from one setting can readily be applied to another. In the present series of experiment, we investigated whether various combinations of cognitive tests, associated to autonomic reactivity responses assessed through cardio-respiratory recordings, would relate to real-life performance on short and medium-term outcomes.

2. METHOD

In a first experiment, we investigated whether psychophysiological results would predict success of military student pilots (SPs) on a major evaluation flight right after the testing, and success in the rest of their flight training after a 6 months period. In a second experiment, we investigated whether extensive preliminary cognitive testing and individually tailored longitudinal monitoring of physical and cognitive performance could predict success of Special Forces trainees (SFs) during their training.

2.1 Experiment 1: SPs

Subjects :17 (14) military student pilots (SPs)

Procedure: 2 practice sessions before experimental data-collection. Subjects were tested twice: in a baseline condition (N=17) and right before the evaluation flight (N=14). Both recording sessions started with a 5 min rest recording of cardio-respiratory parameters (RR-interval [RRI], tidal volume [TV], respiratory frequency [F_resp], ratio inspiratory time over total breath cycle time [Ti/Ttot] and respiratory sinus arrhythmia [RSA] computed through the peak-valley method). Subjects then filled in the Spielberger State/Trait Anxiety Inventory, and subsequently performed the MiniCog Rapid Assessment Battery (MRAB), which was specifically designed to investigate a wide range of cognitive modalities and predict potential performance decrements in astronauts [3]. Upon completion, subjects performed a range of Stroop tasks (colour-word, numerical and emotional Stroop). All testing occurred with simultaneous recording of cardio-respiratory data.

Right after the second data-collection, subjects performed their Progress Test General Flying (PTGF).

Trait predictive value of the testing was evaluated by comparing the results of the baseline recording session with the fact that SPs were still in the training program 6 month after the recordings (dichotomized variable: In vs Out). State predictive value was evaluated by comparing the results on the pre-PTGF recording session and the evaluation of the PTGF (dichotomized variable: Pass vs Fail)

2.2 Experiment 2: SFs

Subjects: 7 SF candidates.

Procedure: Initial assessment: Wechsler Adult Intelligence Scale (WAIS); training session for Stroop and finger precueing task; exhaustive exercise test on treadmill. Weekly assessment as from the start of the training: Profile of Mood States (POMS) questionnaire, 5 min rest recordings of cardio-respiratory data (similar to Exp 1), Stroop task, finger precueing task, exercise test on the treadmill (2400 m). All testing occurred with simultaneous recording of cardio-respiratory data.

Trait predictive value was evaluated by comparing the WAIS results with a ranking of the candidates according to their course performance. State predictive value was assessed by comparing the evolution of the results from the weekly testing with the simultaneous course performance.

3. RESULTS

Considering the very large number of variables generated by the testing, only the relevant differences will be reported here.

3.1 Medium-term predictive value. Experiment 1. 11 of the 17 SPs were still in the training program 6 months after the baseline testing. The main differences between the "In" and "Out" group were the "In" group showing i) overall faster RTs and slightly higher error rates on most of the cognitive testings; ii) a higher "State" anxiety score [U=29,5; p= 0,009]; iii) a higher interference effect for reaction times of student pilot-specific emotional stimuli [63 vs 10 ms; U=12; p= 0,036]. Experiment 2. The differentiating variable was the WAIS performance IQ. Subjects performing better on the training also seemed to have a more optimal speed/accuracy trade-off on the Stroop and finger precueing tasks.

3.2 Short-term predictive value. Experiment 1. 14 SPs were still part of the program at the time of the evaluation flight. Of these 14 subjects, 9 passed their test and 5 failed. For the cognitive testing, there was no tendency of the passed group to show better results, on the contrary: for several subtests, their performance was far worse, both in terms of reaction times and in terms of error rates. However, none of the observed differences showed significance. For the cardio-respiratory data, the passed group showed a larger rest RSA (140 ms vs 103 ms) and a larger reactivity for all the recorded parameter, which approached significance only for TV reactivity [U=3;P=0,088] and RSA reactivity[U=2;p=0,059]. Experiment 2 was interrupted, but for the available data, the POMS seemd to be the most sensitive indicator from weekly performance variations.

4. DISCUSSION

These results suggest that a critical distinction could be made regarding predictive performance assessment, namely *trait* and *state* dimensions. Since one of the intended uses of operational test batteries is to provide an instantaneous measure of the cognitive status of the subject to allow the immediate execution of critical tasks, our results show this would be an inappropriate application so far. However, a dimension showing promising potential is the physiological reactivity.

A limitation of the present research is the small number of subjects. However, this might be an additional difference in the "Trait" vs "State" approach: whereas selecting for stable characteristics might imply a certain tolerance for false negative (thereby discarding people who do show the desired capacities, but who are not selected), testing in operational settings requires to deal with the present resources, i.e. subjects, or personnel. Therefore, any method to be applied should not require a large number of subjects to show validity, since the intended use will always be for people performing critical functions, i.e., small groups. A possible origin of the discrepancy between cognitive test results and immediately subsequent performance might be the motivational appraisal of performance testing. Indeed, in selection contexts, people want to performa at thei best, whereas this was probably not the case in our experiment. Manipulating that variable in future research might uncover a potential link between test results and immediately subsequent performance.

Whereas operational priorities clearly state the need for performance evaluation tools for real-time decision making, their application cannot guide operational choices before sufficient validation allows justifying such decisions.

5. REFERENCES

1. *HUMEX: Study on the Survivability and Adaptation of Humans to Long-duration Exploratory Missions*. European Space Agency, 2000.

2. *BPCR: Bioastronautics Critical Path Roadmap*. National Aeronautics and Space Administration, 2004.

3. Shephard, J. M. & Kosslyn, S. M. *Avn Space Env Medicine*, *76*, B192-B197, 2005.

6. ACKNOWLEDGEMENTS

The present work was supported by the Belgian Dept of Defense (grant ERM HF-10), by the European Space Agency through PRODEX and the Belgian Federal Office for Science Policy (grant Prodex n°90030) and by the Flemish government through a grant from the Minister of Sport. Dr Pattyn's work was supported by a Euro Space Society grant.