



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Predicting the performance of teams in systems

M.A. Sinclair¹, C.E. Siemieniuch¹, M.J.d.C. Henshaw¹, R.A. Haslam¹
and L. Evans²

¹Loughborough University, ²BAE Systems
United Kingdom
m.a.sinclair@lboro.ac.uk

Abstract

The paper describes the development of a tool to predict the success of a team executing a system process. It is expected to be used by systems engineers in initial stages of systems design, when concepts are still fluid, including the team(s) who are expected to be operators within the system. Currently, the tool is undergoing verification and validation; to date, the tool predicts well and shows promise. An unexpected finding is that the tool creates an *a priori* case for Human Factors Integration in projects, to reduce the probability of substandard operational performance.

Introduction

Slowly, we have developed a prototype tool that is able to predict the performance of teams that are executing a process. Verifications and validations of the tool indicate that it performs satisfactorily, and this paper outlines both the development and the evaluations of the tool to date, though it should be noted that further independent industrial validations are under way. These may adjust some of the statements below; in view of this, the paper is conservative in tone.

The original purpose of the tool was to provide designers of military systems with a simple tool to use in the conceptual stages of design (when variables are still variables and not parameters) to help in risk reduction exercises when considering the staffing of

processes. Some sample questions for which the tool could help in providing answers are:

- What is the likelihood that this team will be successful in executing the given process?
- By how much can the team size be reduced, before the likelihood of success becomes unacceptable?
- By how much can the attributes of the individuals in the team be reduced, before the likelihood of success becomes unacceptable?

It has subsequently become evident that these questions could also be developed within the context of Through-Life Capability Management (for systems), with regard to the delivery of Service Level Agreements (guaranteeing performance of the system).

It will also be noted that these questions are phrased in terms of success. This is a significant point; ‘success’ is defined here as executing the process correctly and attaining all of the goals of the process; no reworking, no extra resources, no extra time. ‘Likelihood’, as usual, is expressed as a probability.

The tool rejoices in the name, PEAT – ‘Performance Assessment and Evaluation for Teams’.

Development of the tool

From the outset, three constraints determined the development of the tool:

- **A user-defined constraint:** If the tool needs more than two pages of A4 to explain it, it will not be used. This was a unanimous view among a target group of engineers who were interviewed at the beginning of the project.
- **A business process constraint:** Systems designers in the UK are already familiar with, and may be using, techniques such as HEART and CREAM for assessing the reliability of individuals. The tool should incorporate these techniques, or any other in-house technique, to enhance ease of acceptance into design processes.
- **A design constraint:** At the conceptual stages of design, little will be known about the individuals in the team that executes the process, apart from generic attributes. Equally, the process will be undefined – perhaps just a single flow diagram sketched on a sheet of A4. Hence, the tool must make a minimum demand for input data.

The first constraint has been met, albeit with a back-up manual. The second constraint has been met by creating a three-stage tool; Stage 1 collects data about the attributes of the individuals in the team and the intercommunications necessary for execution of the process. Stage 2 collects an analysis of the process environment, using either HEART or CREAM, or the organisation's in-house technique. Stage 3 convolves the outputs of Stages 1 and 2, and produces a likelihood of success, depending on the binding of the team to the process.

Development of the tool was based on an iterative design process, similar to Extreme Programming. Paper-based versions were developed initially, with subsequent development in spreadsheets. At each step forwards in development, a small-scale study was conducted, usually in real-life situations (e.g. stations in London Underground, yacht crew, helicopter crew, student project groups). In all, 9 identifiable versions of the tool were

developed. The description below is for the latest version.

Starting conditions for the use of the tool assume that the systems engineer (or process designer) has a process description available, though this does not have to be in detail. There will be some team assigned to execute the process; its size, and the expected binding of individuals to the process activities that have been identified so far should be known. Finally, there needs to be some level of knowledge about the environment in which the process will be executed.

Input data on individuals cover five variables. The first is an identifier for each person, to aid the user to understand what has happened. The second is a trustworthiness variable; the dependability of the person to deliver results. The third measures team skills; how constructive the person is in aiding the team to its goals. The fourth assesses the knowledge and skills that the person brings to this particular process. Finally, the fifth variable assesses the authority of the person within the team; rank, wisdom or experience.

In addition, the proposed communication structure (who talks to whom) is captured, as a matrix of one-way links.

From the trustworthiness, team skills, and knowledge ratings, the tool produces a Performance Shaping Factor (PSF) as an error multiplier. High-performance people will have a PSF less than 1.0; for poor performers it will be greater than 1.0

Next, the user analyses the process environment for each individual; this allows for distributed teams; if the team is co-located, only one analysis needs to be performed. This analysis may be performed using either HEART (Williams 1986) or CREAM (Hollnagel 1998), or with an in-house technique substituted for these. The probabilities of failure that emerge from this analysis are now multiplied by the PSFs from Stage 1 to produce a probability of error for each individual in the team in his/her

environment.

Stage 3 happens without necessitating user input. It now combines the values obtained above, the authority ratings, and the communications matrix to arrive firstly at what is called ‘interactive probabilities of error’ for each individual; in other words, acknowledging peer effect. The initial probabilities of error from the paragraph above are now adjusted by these extra variables. The assumption here is that a person’s performance will be influenced by the performances, knowledge, and teamworking capabilities of those who communicate with that person in the execution of the process. This might be feedforward, feedback, or the observable quality of work.

An algebraic equation performs this step, with the property that if a person communicates with very good people, then that person’s probability of error will decrease, and vice versa. Hence, we now have an ‘interactive probability of error’ for each team member.

Secondly, the team’s binding to the process is addressed. It was discovered that three classes of binding cover a wide range of real teams. The three classes are described below:

- **The ‘aircrew’ team.** Consider a helicopter, flying from A to B over hostile territory. The pilot executes the process of flying; the rest of the crew act only as advisors (e.g. ‘A SAM has been fired’), but do not play a part in flying the helicopter.
- **The ‘Boatcrew’ team.** Consider a rowing eight (nine, with the coxswain). From start to finish, each person has a specific task, and cannot perform anyone else’s task. Consequently, the absence of any crew member ensures failure.
- **The ‘Omnicompetent’ team.** Here, anyone can perform anyone else’s task, and may do so in executing the process.

Likelihoods are presented for each of these bindings; the user may choose the best-fitting example, or may calculate a specific value

from the ‘interactive probabilities of error’, if the three classes are deemed unsuitable.

Current state of PEAT

Due to the adoption of a form of Rapid Prototyping as the development methodology, PEAT is now in EXCEL spreadsheet form, in its ninth version. Functionally, PEAT is considered complete *pace* the results of further independent validation. Testing of the user manual is beginning; both the two-page version and the more detailed version.

At the moment, the tool is constrained to deal with teams of ten or less members. This is a developmental constraint; it is expected that when the tool is deemed to have been fully tested, this constraint will be removed. With cognizance of this constraint, PEAT appears to be able to deal with a number of different scenarios; as shown later in the verifications, the teams considered range from military teams to system development teams, to the Board of Trustees for a pension fund; in other words, teams of many different organisational cultures. It can accept distributed teams, and teams where some members do mainly physical tasks and others do mainly cognitive tasks.

An example of the tool in use

A verification study concerned the preparation of 26 Tornado aircraft for active service in the Middle East. A team of 5 RAF personnel prepared the radiation subsystems; radar, IFF, ECM, radio, etc. Because of the compressed timescale to deliver these aircraft, each member of the team was allocated a ‘gofer’, drawn from maintenance personnel qualified for other aircraft who fetched, carried, held, prepared paperwork, etc. for his/her team member.

The team worked as an omnicompetent team. All were rated highly on the tool’s variables, with equal authority, save for one

who exhibited symptoms of ‘burn-out’, and whose ratings were downgraded accordingly.

The process took place in a hangar, round the clock with the team and go-fers catnapping until all 26 aircraft were delivered, on time. This environment had the effect of increasing the basic error rate for each individual in the whole team, to $p(\text{error}) = 0.02$ for team members and 0.4 for gofers.

Combining all the information above, PEAT predicted a $p(\text{success})$ for the whole team of 0.082; in other words, 2.1 aircraft should have undergone the process with no rework, no extra resources, etc. In reality, the success rate was 3 aircraft.

Verification and validation of the tool

In other words, ‘Does it work?’. The evidence so far is that it does. A second question is, ‘Is it sensitive enough?’. Again, the indications are that it is. This sensitivity issue is discussed first below, followed by the verification and validation results.

Sensitivity of the tool. For a team of two, using CREAM to categorise the environment, 1360 values may be obtained between 0.0 and 1.0. For a team of 10, using HEART the range of values is well over 1 billion. There appears to be sufficient sensitivity for most purposes.

Verification of the tool. To date, some 400 verification tests have been carried out, changing at least one variable for each test. To date, two classes of teams have been investigated in some detail. Both are teams of four; the difference between them is in the communication patterns. In a ‘Linear’ team, the members are arranged linearly, with two-way communication between adjacent members (similar to a production line). In a ‘Cocktail’ team the arrangement is that of a star, with a central person communicating with all the others. The others have partial communications between them; the net effect is that each team member has a different number of communication links. In addition,

‘Granovetter’ tests have been conducted (Granovetter 1973), on a 10-person team, exploring the effects of different configurations and numbers of links.

A further restriction on all these tests is that all the tests have been performed using CREAM to characterise the environment.

From all of these tests, it seems that some general conclusions can be drawn.

1. Given that the tool has been constructed as a simple technique, with no feedback loops and no ‘if-then’ rules, it is gratifying that the behaviour is as expected; trends are consistent, and no test produced a prediction outside the range 0.0 – 1.0.
2. It is striking that in all sets of tests, a well-designed working environment is the biggest contributor to $p(\text{success})$. The other variables in the tool become more important as the working environment degenerates, but they cannot make up for it. In a good working environment, team variables are relatively unimportant; even when the team is of poor quality, $p(\text{success}) > 0.9$. For a military environment, this is equivalent to saying that if all the US DoD Systems Engineering requirements have been met completely and in full (or in the UK, the Defence Lines of Development have delivered in full), taking full account of human factors integration, then successful performance of the process is assured (note that this is not saying, ‘Victory is assured’).
3. In poor working environments, the most important variable is the quality of the team members. As long as there are several high quality members in central roles, able to communicate with the other members, then a level of performance ($p(\text{success}) \approx 0.25$) can be achieved. However, even with a high

quality team, it is not possible to lift p(success) above 0.35.

4. A good team is always better than 4 individuals.
5. Of the individual variables, Trust has the most effect, especially when allied to Authority.

Of course, the truth content of these statements depends on the validation studies below, and more of these are being carried out. But, insofar as there is truth in these statements, findings 2 and 3 together have a significant corollary; they provide a strong argument for the importance of Human Factors/ Ergonomics in systems engineering projects. One might imagine that, given the extent and complexity of military systems being sourced for the NATO forces for future decades (particularly the USA), and the likely manning issues for these, these conclusions will be of some interest to those in military procurement, and to those in civilian safety-critical systems management.

Validation of the tool. To date, 16 validation exercises have been carried out, summarised in Table 1. All are historical cases, with the results known; i.e. validation-by-criterion. However, in all cases but the last two the criterion was subjective, since the processes were not repeated. The last two were both repetitive processes, allowing comparison of probabilities of success. Extremely brief descriptions of the cases are included for reasons of space and confidentiality, together with the predicted success and the source's verbatim comment on the result.

Table 1: Aggregated data for validation studies

#	Team process	p(succe ss)	Comment on result
1	Insert new FCS in tank	0.7552	"That looks good."

#	Team process	p(succe ss)	Comment on result
2	Dev't of control system for UAV	0.5906	"OK, if a little generous"
3	Dev't of systems Health Mgmt System	0.8713	"I'm happy with that result -perhaps a little bit high"
4	Create engineering Technical Demonstrator	0.9195	"Result is OK; perhaps a bit high"
5	Dev't of comms system for navy ship	0.9464	"Rings reasonably true"
6	Bid prep'n for US DoD ITT	0.4541	"Result looks OK; wouldn't want to argue with it; perhaps a little bit high."
7	Execution of a Des & Build project for M.Eng degree	0.4875	"Result is a bit low"
8	Deliver HFI to manufacturing Technical Demonstrator	0.9234	"OK; but doesn't account for a weak team member. Got it together because of the efforts of the rest of the team"
9	Dev't of guide for UK Gov't dept.	0.9007	"That's OK"

#	Team process	p(succe ss)	Comment on result
10	Software development for UK NHS	0.5984	“Estimate is a bit low; would have expected about 0.75”
11	Development of a UAV ground station	0.4831	“A bit low - would have expected around 0.7”
12	Management team in University	0.8287	“Result is OK, but this isn’t a normal team”
13	Pension Bd of Trustees managing fund	0.9968	“That’s about right”
14	SAS patrols in hostile territory	0.63	“Difficult to assess whether correct; if contact with the enemy, plans change, hence ‘failure’. But no contact = failure, too. Overall, value looks good.”
15	Preparing aircraft for Gulf War	0.0817	“About right – only 3 of 26 went through without rework”
16	High-tech jobbing shop making military-standard RAM for development studies	0.97	“That’s interesting. Expected monthly performance for this process is between 0.90 and 0.98.”

The last two cases present quantitative validations of some accuracy (for this knowledge domain), and further quantitative independent validations are under way. The subjective comments in the tests mostly indicate acceptance of the predictions, by the ‘expert’ on the team, and can be subjected to statistical evaluation using standard methods in the human sciences.

A Kolmogorov-Smirnoff test (Siegel 1959) was carried out to discover any significant departure from accuracy (note that a Chi-squared test is not appropriate due to Cochran’s criterion (Cochran 1954) .This is shown in table 2 below. Getting 3 predictions wrong represents a proportion of 0.188. According to table E of Siegel (Siegel 1959) this indicates that the Null hypothesis (no departures from Ideal) is not rejected ($p > 0.05$).

Table 2: Data for Kolmogorov-Smirnoff test, from Table 1

Classes	Ideal cumulative	Actual cumulative	Prop. difference
Wrong prediction	0	3	0.188
Correct prediction	16	16	0

A further, Binomial test (Siegel 1959) was executed on those estimates deemed above and below, to test for a bias in the predictions. Table 3 shows the results of this.

Table 3: Data for Binomial test, from Table 1.

Prediction below user’s opinion	Prediction above user’s opinion	Predictions deemed correct
3	4	9

According to Table D of Siegel (Siegel 1959), this indicates that the Null hypothesis (no bias in predictions) is not rejected ($p > 0.05$).

We may conclude from these tests and tables that, to this point in time, PEAT can produce reasonable predictions.

Conclusions

In recognition of the fact that further testing of the tool is required (particularly validation tests), conservative conclusions are in order:

- A prototype tool with some predictive power has been produced. However, it is not yet clear how powerful the tool is, nor how extensive is its range of application
- It is believed to be the only tool able to provide estimates of team performance available for systems designers and human factors experts to use in the early stages of design.
- Planned developments to the user interface should make the tool usable by non-experts in human factors. This is believed to be an asset; if the engineer's own use of the tool shows that human factors issues must be addressed, this is likely to be a convincing argument.
- That the tool is in EXCEL spreadsheet form is good, but insufficient; it is necessary that a web-based version is available, with a management process associated with it, both to maintain its integrity, to capture lessons from its use, and to develop the tool within the organisation hosting it. Some plans exist for this.

Acknowledgements

This tool has been developed within the ambience of an Engineering Doctorate. Thanks are therefore due to the Engineering and Physical Sciences Research Board and the sponsor, BAE Systems, without whose funding this would not have happened. Many, many other have provided insight and knowledge to this exercise; those most directly involved in this are Dr Siemieniuch, Prof Henshaw, and

Prof. Haslam, all of Loughborough University, and Mr. Evans of BAE Systems. Mention should also be made of the early work which contributed to this by Dr. Jonathan Benn, Billy Fiske, Laura Matthews, Nick Reid and Isabel Smith. Their contributions are saluted.

References

- Cochran, W. G. (1954). "Some methods for strengthening the common Chi-squared tests." Biometrics **10**: 417-451.
- Granovetter, M. S. (1973). "The strength of weak ties." American Journal of Sociology **78**: 1360-1380.
- Hollnagel, E. (1998). Cognitive Reliability and Error Analysis Method (CREAM). Den Haag, Elsevier Science.
- Siegel, S. (1959). Non-parametric statistics. New York, McGraw-Hill.
- Williams, J. C. (1986). HEART – a proposed method for assessing and reducing human error. 9th Advances in Reliability Technology Symposium, University of Bradford, UK.

Biography

M.A. Sinclair, B.Sc., M.Sc. Currently an Eng.D student at Loughborough University, UK. In a previous existence he was a Senior Lecturer in Ergonomics in the University. He has many years of experience in European Union projects, and in the automobile and aerospace domains.

C.E. Siemieniuch, Ph.D. Currently a Senior Lecturer in the Dept of Electronic and Electrical Engineering. Her specialty is the Human Factors aspects of systems-of-systems engineering, based on many years of experience in UK and European Union projects, in both military and civilian domains.

M.J.d.C. Henshaw, Ph.D. After many years in the aerospace industry, initially as an aeroelastician and latterly as a senior project manager, he joined Loughborough University as Professor in Systems Engineering. Together with Dr Siemieniuch he heads the Engineering Systems of Systems centre.

L. Evans, B.Sc., M.Sc. Currently Head of Human Factors, Advanced Technology Centre, BAE Systems UK. His role as a senior manager includes, among many other responsibilities, the nurturing of applied, strategic research projects connected with Human Factors Integration in the military domain.