Development of a tool to predict organised team performance in process execution

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This paper was presented at the Annual Conference, Ergonomics Society, Keele University, UK, 13-15 April 2010. It is a preprint of an article whose final and definitive form has been published in the Proceedings of the Annual Conference of the Institute of Ergonomics & Human Factors, ‘Contemporary ergonomics and human factors 2010’ [© CRC Press, Taylor & Francis]: http://www.crcpress.com/product/isbn/9780415584463

Metadata Record: https://dspace.lboro.ac.uk/2134/9283

Version: Submitted for publication

Publisher: © CRC Press (Taylor & Francis Group)

Please cite the published version.
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.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
DEVELOPMENT OF A TOOL TO PREDICT ORGANISED TEAM PERFORMANCE IN PROCESS EXECUTION

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

¹Loughborough University, ²BAE Systems

The paper describes the development of a tool to predict the success of a team executing a 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 a good argument for Human Factors Integration in projects.

Introduction

Slowly, we have developed a 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 verifications and validations are planned. 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 will 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.

There is still debate about a name for the tool. Hence, it will be called ‘the tool’ in this paper.
Development of the tool

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

- A user-defined constraint: If the tool needs more that 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 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, 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, without recourse to minuscule script. 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, user-oriented 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. station staff in London Underground, yacht crew, helicopter crew, student project groups). In all, 8 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 user 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.

Structure of the tool

Input data on individuals cover five variables. These variables map to Thurstone’s ‘Five Factor Model (Thurstone 1934) and to Salas et al’s ‘Big Five’ model (Salas, Sims et al. 2005). The first is an identifier for each person, to aid the user to understand the intermediate steps. 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 the process. Finally, the fifth variable assesses the authority of the person within the team. For the first 4 variables, 4-point rating scales are provided. For the last, there is a 5-point scale. Albeit simple, these scales provide adequate sensitivity, as discussed later.
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). High-performance people will have a PSF less than 1.0; for it poor performers 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). 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. This might be feedforward, feedback, or 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 individual.

Secondly, the team’s binding to the process is addressed. It was discovered that three classes of binding cover a 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 for each of these teams are presented for the choice of the user.

**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 ‘go-fer’, 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
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, the tool predicted a $p(\text{success})$ for the whole team of 0.1; in other words, 2.6 aircraft should have undergone the process with no rework, no extra resources, etc. In reality, the success rate was 3 aircraft (the rest required some degree of rework).

**Current state of the tool**

Due to the adoption of a form of Rapid Prototyping as a development methodology, the tool is now in spreadsheet form, in its eighth version. Version 6 was in a testable state; verifications and validations have taken place with versions 6, 7, and 8, and these are described below.

Functionally, the tool is being made more presentable and understandable to the intended user. This is necessary for testing of the user manual, currently being written. In accordance with the first constraint above, a two-page version and a more detailed version are being produced.

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 be fully tested, the constraint will be removed. With cognizance of this constraint, the tool 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. It can accept distributed teams, and teams where some members do mainly physical tasks whereas others do mainly cognitive tasks.

**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. Normally, one would then describe the verification (does it do what we say it does? Does it behave as we might expect?) of the tool first, then its validation (does it predict in the real world?). Because of an unexpected finding during the validation phase, it seems better to discuss the validation first, then the verification.

**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 approximately 57 billion. There appears to be sufficient sensitivity for most purposes.

**Validation of the tool**

To date, 14 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 the
criterion was subjective, since the processes were not repeated.

<table>
<thead>
<tr>
<th>#</th>
<th>Team process</th>
<th>p(success)</th>
<th>Comment on result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Insert new FCS in tank</td>
<td>0.7552</td>
<td>“That looks good.”</td>
</tr>
<tr>
<td>2</td>
<td>Dev’t of control system for UAV</td>
<td>0.5906</td>
<td>“OK, if a little generous”</td>
</tr>
<tr>
<td>3</td>
<td>Dev’t of systems Health Mgmt System</td>
<td>0.8713</td>
<td>“I’m happy with that result - perhaps a little bit high”</td>
</tr>
<tr>
<td>4</td>
<td>Create engineering Technical Demonstrator</td>
<td>0.9195</td>
<td>“Result is OK; perhaps a bit high”</td>
</tr>
<tr>
<td>5</td>
<td>Dev’t of comms system for navy ship</td>
<td>0.9464</td>
<td>“Rings reasonably true”</td>
</tr>
<tr>
<td>6</td>
<td>Bid prep’n for US DoD ITT</td>
<td>0.4541</td>
<td>“Result looks OK; wouldn’t want to argue with it; perhaps a little bit high.”</td>
</tr>
<tr>
<td>7</td>
<td>Execution of a Des &amp; Build project for M.Eng degree</td>
<td>0.4875</td>
<td>“Result is a bit low”</td>
</tr>
<tr>
<td>8</td>
<td>Deliver HFI to manufacturing Technical Demonstrator</td>
<td>0.9234</td>
<td>“OK; but doesn’t account for a weak team member. Got it together because of the efforts of the rest of the team”</td>
</tr>
<tr>
<td>9</td>
<td>Dev’t of guide for UK Gov’t dept.</td>
<td>0.9007</td>
<td>“That’s OK”</td>
</tr>
<tr>
<td>10</td>
<td>Software development for UK NHS</td>
<td>0.5984</td>
<td>“Estimate is a bit low; would have expected about 0.75”</td>
</tr>
<tr>
<td>11</td>
<td>Development of a UAV ground station</td>
<td>0.4831</td>
<td>“A bit low - would have expected around 0.7”</td>
</tr>
<tr>
<td>12</td>
<td>Management team in University</td>
<td>0.8287</td>
<td>“Result is OK, but this isn’t a normal team”</td>
</tr>
<tr>
<td>13</td>
<td>Pension Bd of Trustees managing fund</td>
<td>0.9968</td>
<td>“That’s about right”</td>
</tr>
<tr>
<td>14</td>
<td>Preparing aircraft for Gulf War</td>
<td>0.1032</td>
<td>“About right – only 3 of 26 went through without rework”</td>
</tr>
</tbody>
</table>

Table 10  Aggregated data for validations. Minimal information is included for
A Kolmogorov-Smirnov test (Siegel and Castellan 1995) 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 12 below.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Ideal cumulative</th>
<th>Actual cumulative</th>
<th>Proportional difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miss</td>
<td>0</td>
<td>3</td>
<td>0.214</td>
</tr>
<tr>
<td>Correct predictions</td>
<td>14</td>
<td>14</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 12  Kolmogorov-Smirnov test for departures from a ideal distribution. Getting 3 predictions wrong represents a proportion of 0.214. According to table E of Siegel (Siegel and Castellan 1995), this indicates that the Null hypothesis (no departures from Ideal) is not rejected (p > 0.05).

A further, Binomial test (Siegel and Castellan 1995) was executed on those estimates deemed above and below, to test for a bias in the predictions. Table 13 shows the results of this.

<table>
<thead>
<tr>
<th>Prediction below user’s opinion</th>
<th>Prediction above user’s opinion</th>
<th>Predictions deemed exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 13  Binomial test for bias in predictions. Only the first two columns were used for this test. According to Table D of Siegel (Siegel and Castellan 1995), 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 the tool can produce reasonable predictions.

**Verification of the tool**

To date, some 350 verification tests have been carried out on these teams, changing at least one variable for each test. Further verification tests are under way. To date, two particular 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. A further restriction on 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 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 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 \( p(\text{success}) \) above 0.35.

4. A good team is always better than 4 individuals.

Of course, the truth content of these statements depends on the validation studies above, and more of these must be 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 engineering projects. One could state that, ‘To the extent that HFI is ignored in systems design, greater rates of failure for those human processes are guaranteed’. Furthermore, this tool could provide a probability estimate for failure.

The exploitation of this is evident; currently, attempts to justify the need for resources to be devoted to Human Factors Integration are largely made by appeals to instances from the past; Hendrick’s article is a good example (Hendrick 1997), as is Oxenburgh, 1991. However, the argument above does not rely on instances; it is an \textit{ab initio} argument, in theory applicable to any team scenario. Its validity comes from the validation tests, discussed earlier. It is likely that those responsible for bids for work will be able to turn the statements in the bullet points immediately above into risk and cost estimates that will provide strong arguments for improved funding of HFI.

**Conclusions**

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

- A 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 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. 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. Their contributions are saluted.

References