2013

Measuring Human Factors Success in Acquisitions

Jason Demagalski

Dr Branka Subotić

Dr Jerry Crutchfield

Dr Dwayne Norris

MEASURING HUMAN FACTORS SUCCESS IN ACQUISITIONS

Jason Demagalski / Dr Branka Subotić
FAA / NATS
Washington DC, USA / London, United Kingdom

Dr Jerry Crutchfield / Dr Dwayne Norris
FAA / American Institutes for Research
Oklahoma City, USA / Washington DC, USA

Human factors specialists work as part of multidisciplinary teams supporting acquisition programs. Depending upon the acquisition model and culture of an organization, the exact points of involvement of human factors specialists may vary, as too will the metrics of success for each project. Acceptable/optimal level of operational performance is the goal of any Air Traffic Control (ATC) system. Human performance is a key component of operational performance. Measuring human performance can be an area of academic debate often yielding vague answers or need for more research. This is of little use or interest to project and senior managers who make decisions. This paper presents a high level overview of the difficulties of measuring performance, provides a case study of how this has been addressed in the United Kingdom ATC organization and concludes with thoughts on how the human factors community can support acquisitions and achieve what we would deem, success.

Human factors specialists work as part of multidisciplinary teams with a variety of stakeholders within acquisition programs and projects. Depending upon the organization and its acquisition structure, the exact stage/s and magnitude of involvement may vary. These inputs may range from the initial concept development stage through to operational implementation or merely an opportunity to “sign-off” at a set point in the acquisition process when anything but acceptance would be unwelcomed. While everyone involved in a project is seeking success, it is not sufficient to assume that all involved have the same criteria for success.

There are many intended and unintended consequences to an ambiguous definition of project success. If we ensure that the project success criteria include valid and verifiable measures of human performance then the chance of consulting with and incorporating the design recommendations from human factors specialists up front will be enhanced. The extent to which this can be achieved will reflect the dominant measures of success for the ATC organization. The ultimate goal of human factors specialists is for auditable and traceable human factors inputs to track from early requirement definition to final validation of human performance.

The central problem for human factors specialists lies in what organizations call success in acquisition programs. Is success when the hardware is installed? When the software functions correctly? When the operator can use the new tool/equipment? Ultimately, success must be when the overall system performance, as a result of the acquired system or tool, meets given criteria, one of which being that the human operator’s task performance is within the acceptable range. If an organization is striving for an acceptable level of system and task performance following any change, then success of the project should be based on achieving it.

Within ATC, operational performance can be measured using a variety of metrics, such as traffic throughput, safety statistics and system availability. Different stakeholders will judge their success against the metric most closely aligned with their specialism. However, within large ATC organizations, there is considerable risk that the stove-piped nature of development of tools, technology, airspace changes, safety initiatives and such like, when implemented, lead to a resultant performance level less than that of the individual components. As such, if acquisitions and operational implementations are poorly managed strategically, then the Air Traffic Controller Specialist (ATCS) is left to individually integrate their knowledge and understanding of each newly acquired tool/procedure to allow them to perform effectively. Projects may be deemed to have been successful upon installation and activation of the new tool in the ATC facility regardless of how the controllers overall performance is affected.

In the last 20 years the pace of development of potential new ATC systems and the opportunities for their implementation have increased. As a consequence, this has increased the burden on organizational acquisition which in turn has increased the importance of having useful human factors requirements, the importance of testing and validating proposed systems and the importance of having consistent and reliable measures to indicate whether a system has fulfilled the requirements and is fit for purpose.

A key component of being able to measure human performance is to have defined performance standards. Performance standards describe the expected level of performance, in measurable terms, on critical work tasks. Developing reliable and valid performance standards requires the identification of the current work requirements. Work requirements are typically comprised of some combination of the tasks that are performed on the job; the knowledge, skills, abilities, and other attributes (KSAOs) required to perform those tasks; and the tools and equipment incumbents used to perform those tasks. Such requirements are identified through a job analysis and the resulting requirements can be used to support a variety of activities from personnel practices, acquisition requirements and establishing performance standards.

Performance standards define the level of performance required or expected at a given level of expertise. While the job analysis identifies the relevant work tasks, performance standards capture how the tasks are done or their expected results. Said another way, task lists allow organizations to differentiate between jobs; performance standards allow organizations to differentiate degrees of controller proficiency/performance on a specific job task or group of tasks. Performance standards should be realistic or challenging, specific, measurable, consistent with organization goals, and understandable. Performance standards can be used to develop assessments for a variety of purposes, or for setting the cut scores for those assessments. In addition, they provide a way to communicate the expectations of the organization. The level of performance at which the standards are built should be based upon the purpose for which the standards will be used.

There are a number of ways to set performance standards. A common method for setting performance standards, subjective standard setting, while the simplest, can also be the least effective. Using this method, standards are set without reference to criteria for effective performance. Instead, they are set arbitrarily, typically based on convention or rules of thumb (e.g., 70% is a passing score on a test). Norm-referenced methods are based on performance relative to others; for instance, training instructors might decide that the top 50% of a given training class has high enough performance levels, while the bottom 50% might need additional training. Criterion-referenced methods are a different and more thoughtful way of setting performance standards. These methods require that standards are based on specific criteria that are decided in advance (e.g., mastery of a work tasks at specified level of performance). Given the parameters within ATC, performance standards defined by ATCSs are ideally established using a criterion-referenced approach.

Within the criterion-referenced method, there are two general procedures that can be used to establish performance standards. Judgmental procedures require SMEs to make judgments about how work tasks are to be performed with respect to a target level of performance (e.g., the level of performance for a minimally acceptable candidate). Alternatively, empirical procedures establish performance standards at a level that differentiates task performance between two groups that vary in their level of qualifications (e.g., contrasting experts and novices). Both methods are acceptable for setting criterion-referenced performance standards. Krokos, Baker, Norris, and Smith (2007), developed performance standards for ATCS in tower, Terminal Radar Approach Control (TRACON), and Enroute environments. An example can be found in table 1.

Table 1. Enroute Control Center Performance Standards: Sample

Example Tasks	**Performance Standards.** The Successful Controller...…
1. Project mentally an aircraft's future position 2. Ensure separation using data from various sources of information including range/bearing function, vector lines, route function, or User Request Evaluation Tool (URET) function etc. 3. Determine whether aircraft/airspace separation standards may be violated 4. Review radar display to ensure aircraft compliance with clearance	**Behavioral Indicators of Certified Professional Controller (CPC)-level Performance:** • Appropriately uses tools to observe alerts, have vector lines out to watch projected traffic • Uses the Range Bearing Indicator (in URET) to measure distance • Initiates control actions prior to being alerted to a potential conflict **Cognitive Indicators of CPC-level Performance:** • Takes wider scan in order to understand how control actions affect the aircraft around them now and in the future • Develops multiple potential plans of action to resolve a potential conflict • Projects future position of aircraft **Results of CPC-level Performance in this Sub-Activity:** • No conflict alerts occur, and no “immediate” or “expedite” type control instructions (e.g. turn right immediately) are issued. • Surrounding controllers are not negatively affected by controller’s actions. • All aircraft are separated safely.

Performance standards describe the expected level of performance on each activity at a given level of expertise. It is important to remember that these performance standards are not measurement tools. However, the standards could be used to develop such tools, thus allowing them to realize their full potential. For example, knowledge tests, job simulations, performance rating scales, or checklists could be developed to measure performance based on the standards.

Efforts to identify a set of generally accepted performance measures for air traffic control human factors evaluations have been plagued by challenges for decades (Hopkin, 1980). Challenges include access to data (due to technical feasibility, resources needed to collect the data, etc.), reliability of subjective criteria and the relationship of subjective criteria to objective criteria, as well as the largely cognitive, and therefore unobservable, nature of the air traffic control task. Additionally, these data are complex and dynamic in nature which can lead to analytical challenges, for example, comparing results collected from one type of sector to another type of sector or within the same sector but with different air traffic situations. Continuing advances in human factors methodologies and in information technology have over time presumably mitigated some of the challenges. With the increasing tendency of proposals to enhance the air traffic control system to include wider and more far reaching changes, it seems as important as ever to identify a set of standard performance measures so that the results of evaluations of proposed changes to tools and procedures can be consistent and comparable. It may be time once again for our community to examine the performance measures available and try to come to a consensus on what the performance measures should be.

Even with improvements in general methods and greater access to data, the challenges unique to the air traffic control domain still remain. New technologies and procedures that are evaluated in different ways and by different organizations in the industry only exacerbate these unique challenges. A standard set of scenario events and conditions would help to put results from different studies in context with relation to each other. Events would likely include off-nominal occurrences (e.g. convective weather, radio outage, runway shut-down) and conditions would likely be based upon predictions of traffic levels and complexity at certain milestones in the future. An application of our understanding of air traffic and air space complexity to a method to compare human performance experiencing different air traffic or using different air space may prove feasible and beneficial as well.

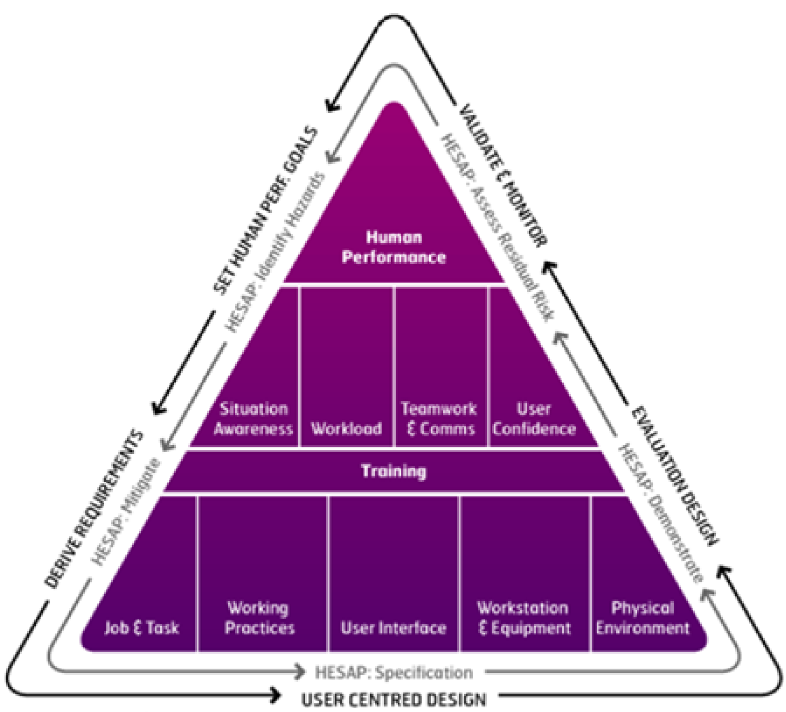
Human factors design through to performance in an international aviation organization (NATS, UK)

NATS is the United Kingdom's (UK) largest air traffic navigation service provider. NATS provides services to aircraft flying through UK controlled airspace and at several UK and international airports. In 2010/11, NATS operational staff handled 2.1m flights (87% of flights to/from UK, 13% overflights; NATS corporate presentation, 2012).

Interim Future Area Control Tools Support (iFACTS) for one of the NATS Area Control Centre (ACC) was born out of Research and Development (R&D) driven innovation that matured over years into a set of control tools. iFACTS represents a set of controller tools that use the current position of the aircraft, the flight plan and complex mathematics to predict the likely future position of all the aircraft in a sector. These positions are then compared for up to 18 minute into the future and any potential conflicts are highlighted to the controller. The most noticeable change from the previous system is the removal of the paper flight progress strips previously used by the controllers, and the subsequent requirement to enter all tactical clearances into the system electronically.

iFACTS completed transition into service in November 2011. On this particular project, the NATS Human Factors (HF) team was involved from design stages, through implementation, to monitoring of the system in live operation within its first six months of operation. All HF activities conducted on iFACTS were structured using the NATS HF assurance framework (Figure 1). This framework is split into emergent properties (depicting how the whole system performs) and design characteristics (describing characteristics of the individual parts of the system).

Figure 1. Human Factors Assurance Framework



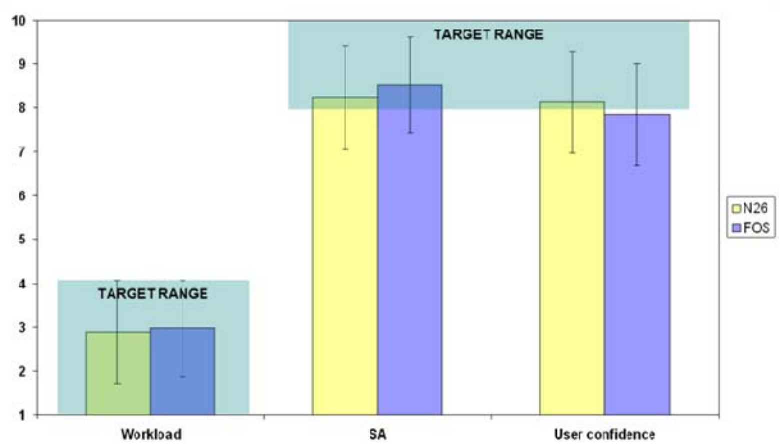
At the bottom of the triangle lie design characteristics describing individual parts of the system. The design characteristics can be directly designed. Therefore, we can specify requirements and objectives for these areas and then verify (through test and audit activities) that they have been met. It is by influencing these design characteristics during the design phases of the project, that we can measure user and system performance, assess design improvements and influence the ultimate concern of effective task performance.

The emergent properties of the system (i.e. situation awareness, workload, teamwork and communication, and user confidence) cannot be directly designed, but are the product of a number of aspects of the system design and the design process. Human performance (both safe and effective) sits at the top of the framework and is NATS' ultimate concern. The NATS HF lead sets performance objectives and criteria for emergent properties at the beginning of the project to inform the design work. These performance objectives and criteria are then validated towards the end of the project.

Due to its complexity and impact, the iFACTS project spanned years, and resulted in a range of staff working on it, including different human factors specialists. The consistency of approach was a must and a pillar of quality assurance over the decade of project life-cycle. To enable the required consistency of approach, human factors work on the project was structured around identification, management and fulfillment of relevant requirements, all structured around impacted areas of the HF assurance framework.

Relevant requirements originated throughout the project life cycle from human factors work, user-focused workshops, training needs analyses, process and procedural changes identified, and a range of risk management workshops. As the project matured, the more specific the requirements became. In the final phases of the project, a complete set of requirements was available for system validation, managed via IBM DOORS application. The type and nature of requirements was driving the design of system validation scenarios. Some of the requirements required specific event to be simulated for validation evidence to be captured, whilst others were quite generic and able to be captured in each and every scenario. Each requirement was mapped with relevant measures coming from subjective, observational/behavioral, and objective classes of measures.

Figure 2. Comparison on system validation and live operational data for the three key HF measures



Once the relevant measures were selected, it was necessary to define a target range for each measure of interest. This allowed the human factors specialists to set success criteria for validation of each requirement. This was usually based upon the baseline captured using the existing system, fine-tuned based on expert judgment and known properties of traffic. The target for each measure of interest was always defined prior to the beginning of each system validation activity.

iFACTS human factors assurance status was based on evidence derived from several sources: R&D work, past research, user-centered design, safety work, operational documents, system validation, and monitoring of live operations. In the final stages of the project, system validation evidence became vital. Figure 2 shows an excellent transfer of human performance findings (Workload, Situational Awareness (SA) and User Confidence) from system validation of iFACTS to its live operations.

CONCLUSIONS

Unless the activities that human factors professionals engage in within ATC acquisitions from initial requirements, design through to operational implementation, result in acceptable levels of human/task performance, we are not being as effective as a community as we should. Successfully bringing together human factors specialists, project managers and acquisition specialists requires compromise and useful goals that support and not hinder large organizational acquisition programs.

The NATS approach is a pragmatic one, combining academic knowledge with real world compromise. Acceptable levels of task performance were defined by human factors specialists. For complex systems or tools, lower level of task performance can be set as acceptable for the "O" date with appropriate operational mitigations. The higher levels of task performance should be expected within a predefined time in operations, when the operational mitigations, previously put in place, can be eliminated. This allows an incremental and 'soft 'implementation of complex systems rather than a 'big bang'. A key component to their approach to the application of human factors is traceability of all requirements within a project, all the way to documentation and final testing results (including verification and validation) as well as throughout monitoring of the system in live operation.

Historically there have been challenges to measuring performance in ATC human factors and the human factors community has not settled on a standard way to test ATC systems nor a standard set of performance measures used to assess whether a system performs successfully.

Within the FAA, we need a consistent method for testing new systems and human performance measures to support whether the requirements have been fulfilled. We must concede that we will never find a method that everyone universally agrees with, disagreements will be based around robust academic arguments. However, to acquisitions managers, those arguments are just that, academic. Technology is moving forward, systems are being procured and the longer it takes us to agree on such methods, the more systems are being operationally installed, in ways that are not what human factors experts would consider optimal. We need to select a consistent way to test new systems and a valid way to evaluate them for success. A model of human factors in acquisitions that drives through the whole process, defining and supporting each stage of acquisitions with appropriate tools/defined activities, is essential.

References

Hopkin, V. D. (1980). The measurement of the air traffic controller. Human Factors, 22(5), 547-560.

Krokos, K. J., & Baker, D. P., Norris, D. G., & Smith, M. A. (2007). *Development of Performance Standards for Air Traffic Control Specialists* (FAA Technical report submitted under Grant # 99-G-048). Washington, DC: American Institutes for Research.

Acknowledgements

The authors would also like to point out that the views expressed in this paper are primarily the views of the authors and may not represent the views and opinions of the FAA, NATS or AIR.