Managing Knowledge of Human Performance Assessment: A Human Sigma Approach

Carol Paris & Joan Johnston
Naval Air Warfare Center Training Systems Division, Orlando, Florida
carol.paris@navy.mil; joan.johnston@navy.mil

Peter Chu
Naval Sea Systems Command
Washington, D.C.
peter.chu@navy.mil

James Brewer
Novonics Corp.
Norfolk, VA
jbrewer@novonics.com

ABSTRACT

The Department of Defense (DoD) has adopted Six Sigma as the basis for a disciplined methodology to improve its processes. To reduce variability of human processes is a challenge introduced by "Human Sigma" advocates. It focuses on adapting Six Sigma discipline to measuring and improving human competencies. DoD is one of the largest employers, with an ongoing need to measure and assess human performance effectiveness in the context of mission accomplishment. Accurately measuring performance, although labor-intensive and expensive, is most importantly, science-driven. Conserving valuable resources is at the heart of current knowledge management efforts using human performance metrics. With shrinking budgets and workforces, and the complexity of new technologies and warfighting environments, the time is ripe for instituting such an initiative. This paper illuminates efficiencies, cost-savings, and quality to be gained through developing a knowledge management system (KMS) for human performance assessment (KMS-HPA) for the military. We examine recent United States and International efforts to capture and manage knowledge about measures and their development, from such domains as air traffic control, human factors, and command and control. We examine lessons learned as they apply to enterprise level initiatives, since no such capability exists to support the evaluation of warfighter performance. As complexity of future military operations escalates, and as multinational militaries unite in common operations, we face new and unknown challenges for measuring mission success. Improvements in processes for measuring performance effectiveness should enhance future readiness.

ABOUT THE AUTHORS

Carol Paris is a Senior Research Psychologist for Naval Air Warfare Center Training Systems Division. She has eleven years of DoD experience, conducting research in tactical decision making, team performance, human systems integration, and simulation training technologies. She currently serves as Chair of the North Atlantic Treaty Organization Human Factors and Medicine Research and Technology Group 156, "Measuring and Analyzing Command and Control Performance Effectiveness." She obtained a Ph.D. in Human Factors Psychology from University of Central Florida and is certified by the International Society of Performance Improvement as a Certified Performance Technologist.

Joan Johnston is a Senior Research Psychologist for the Naval Air Warfare Center Training Systems Division. Her research focus is tactical decision making under stress, team performance and team training technologies, and distributed simulation-based training. Dr. Johnston obtained an M.A. and Ph.D. in Industrial and Organizational Psychology from the University of South Florida.

Peter Chu is an Advanced Science and Technology Engineer at the Naval Sea Systems Command Program Executive Office for Training Systems. His current focus is developing training technologies that can be employed to improve tactical operational proficiency. Mr. Chu obtained a B.S. in Electrical Engineering from University of Maryland.

James Brewer is a Senior Engineer for NOVONICS Corporation. Over the last 10 years, he has spearheaded development of training models and architectures for the Naval Sea Systems Command Total Ship Training System. Mr. Brewer obtained a BSEE from California Polytechnic University, San Louis Obispo, and an MSSM from the University of Southern California.

Managing Knowledge of Human Performance Assessment: A Human Sigma Approach

Carol Paris & Joan Johnston
Naval Air Warfare Center Training Systems Division, Orlando, Florida
carol.paris@navy.mil; joan.johnston@navy.mil

Peter Chu
Naval Sea Systems Command
Washington, D.C.
peter.chu@navy.mil

James Brewer
Novonics Corp.
Norfolk, VA
jbrewer@novonics.com

INTRODUCTION

Over the past decade, shrinking budgets and aging warfare systems have resulted in the Department of Defense (DoD) widely deploying Six Sigma as a disciplined method for improving business processes in order to reduce costs. Six Sigma strives to reduce variability in processes, eliminate waste, and increase quality by making processes repeatable (Miller, 2007; Robinson, 2008). Consequently, money saved on the corporate side means more funding for effectively maintaining and employing weapon systems. However, we propose that focusing on corporate business processes only achieves a fraction of cost savings. Significant cost savings can be realized through human-centered design of weapons systems and their deployment. This is especially important given the expanded missions under the Global War on Terror (GWOT). GWOT has required a rapid response with surging forces, deploying new technologies (e.g., unmanned vehicles), increasing joint and coalition operations, and reconfiguring legacy with new weapons systems, employing them in new ways. Consequently, enormous pressure has been placed on ensuring warfighter performance readiness. Evaluation of human performance effectiveness in the context of mission accomplishment can provide feedback to improve processes required for designing, developing, building, and employing warfighting systems.

Six Sigma programs are not designed to address the pressing requirement for this extended and rapid expansion in warfighting missions. Instead, we propose applying the "Human Sigma" approach, a concept resulting from extensive research by the Gallup Organization involving over 80,000 managers in more than 400 companies (Fleming & Asplund, 2007; Fleming, Coffman, & Harter, July-August 2005). The approach takes knowledge about individuals in the work environment and applies it to create a more productive and consistent workforce. The original study analyzed how to assess, manage, and improve employee-customer encounters. The same approach can be applied to other human-centric endeavors. The idea is the same as Six Sigma - to reduce variability in processes and improve quality - but it is focused on assessing and managing human processes, interactions, and reducing human variability (in lieu of manufacturing or service processes).

Human performance measures are used to track changes in human performance processes. Good measurement enables understanding, decision-making, forecasting, record keeping, diagnosis and self-improvement—all of which lead to process improvements, mission readiness, and return on investment. However, for such measurement to be science-driven, it can be both labor-intensive and expensive. Furthermore, chances for error are multiplied by the process of creation, selection, application, analysis, and interpretation of performance measures and their results. For example, an analyst may detect a performance issue using a particular performance measure, in a certain situation or scenario. However, during subsequent validation processes (i.e., during developmental or operational testing), there is no guarantee that the implemented solution (e.g., training program or modernization) will use the same measure in the same situation or scenario(s). Without such congruence, it is impossible to determine whether the original performance shortcoming has actually been improved. More fundamentally, no efficient technologies exist to ensure the measures employed are good or appropriate for a particular problem in the first place. This problem extends beyond simply ensuring that a performance measure is held constant throughout the process.

For these reasons we believe a Human Sigma approach cannot currently be employed more widely because it lacks a unified Knowledge Management System for Human Performance Assessment (KMS-HPA) that persists over the performance problem-to-solution lifecycle. A KMS-HPA can enable congruence among the fleet, acquisition, and

Manpower, Personnel, Training & Education (MPT&E) communities during efforts to meet mission performance requirements. A KMS-HPA would provide a source of human performance measures, metrics, and methodologies that are conceived within a unified, systemic model of performance measurement. The main goal of a KMS-HPA is to promote quality, consistency, and coordination of measurement within and across domains. Therefore, the purpose of this paper is to describe how a KMS-HPA can lead to significant efficiencies, cost-savings, and quality that can be achieved through assessing human performance with respect to mission effectiveness. We describe a typical problem, define requirements for a KMS-HPA model, review existing prototype systems and provide lessons learned for a way ahead that will enable employment of Human Sigma for warfighter readiness.

USE CASE EXAMPLE

Adopted by the Six Sigma and Human Sigma advocates, the Define -- Measure -- Analyze -- Improve -- Control (DMAIC) methodology is a systematic approach for validating improvements in new business and human performance processes (George, Rowlands, Price, & Maxey, 2005). As a tool for Human Sigma, the DMAIC approach is used to identify human performance challenges and develop, evaluate, and validate improved performance processes. Next we describe a hypothetical use case to demonstrate the need for employing DMAIC to address typical human performance problems with respect to mission readiness.

Hypothetical Scenario

This scenario demonstrates the extent to which measurement can have an impact on very important decisions, and includes how lack of coordination and communication among the stakeholder activities can be a typical barrier to adequate employment of measures. Note in the chronology of events the number of times that measurement and assessment is needed (**bold numbers in parentheses**), the lack of measures congruence, and the consequences when coordination from requirements specification through development and testing fails to materialize.

- It was noticed that shipboard Air Defense Coordinators (ADCs) are not performing the **(1)** "detect-to-engage" task criteria using the current configuration of communications equipment and information displays. Through further analysis it was concluded that operator **(2)** situation awareness (SA) was not adequate to perform the tasks during critical periods of the mission. Given the fact that all equipment operators had been formally qualified, it was suggested that the source of error must be equipment-related because the initial suspicion had been a usability problem with system interfaces.

- The acquisition program office conducted a **(3)** heuristic usability evaluation and found **(4)** errors of detection related to the interface symbology representing aircraft. It was then concluded that the lack of SA was related to the usability issue.

- The Office of Chief of Naval Operations determines this is a serious issue that must be rectified, leading to the requirement that a new materiel solution be developed (e.g., new symbology library). The requirement states that the new equipment shall provide the ADC with adequate SA in order to perform certain missions.

- The acquisition program begins the development of new systems in order to resolve the problem. However, no specific data regarding the nature of the problem, performance context, or measures or criteria are forwarded to the acquisition program X during development. The new systems undergo test and evaluation to determine whether it is meeting mission requirements of improved SA. Development analysts select **(5)** an SA measure based on what they think will be appropriate. However, they lack critical information about the original problem. After the evaluation, the results indicate that SA is "moderately high." The new system is, therefore, judged as adequate to meet the requirement, and is procured and deployed as an upgrade to the existing platform.

Has the new system really resolved the problem? The following list of issues reveals this question is still open.

- During testing of the old system, SA may have been measured using one particular measure based on the evaluator's preference. No information was exchanged that ensures that the same measure is employed during developmental or operational testing as was used initially.

- Few precautions were taken to ensure that the initial measurement conditions, such as physical environment, stress and cognitive load, and how the measure was collected, matched those under which the later tests were administered.

- It is not known whether either measure of SA was appropriate or valid in its application, nor was there assurance that sufficient rigor was used during testing.

- An acceptable criterion for SA was not established at the beginning of the problem. Therefore, using different measures of SA invalidated conclusions about how the problem was solved.

- Once the new software system (symbols) was deployed it was not evaluated to determine whether the same **(6)** errors in overall mission task performance were observed.

- Had objective measures and data been available and shared across activities (MPT&E, acquisition, and operational) it may have become apparent that it was not an equipment issue at all. Instead, the problem was common to a set of operators receiving training around the same timeframe, with a lapse in applying their **(7)** knowledge. Skill decay was ultimately the source of the problem.

- Even if SA had been adequately improved by the changes no **(8)** "acceptability analyses" were performed at the fleet level with regard to the new symbology. Therefore, it is possible that familiarity with the old symbology had resulted in such a strong **(9)** cultural preference for the old system that would prevent this change from ever being viable.

As the preceding scenario illustrates, failure to employ DMAIC occurred at numerous points in the process, mostly due to the lack of correspondence and congruence among the various activities (fleet, acquisition, and development) involved.

KNOWLEDGE MANAGEMENT SYSTEM FOR HUMAN PERFORMANCE ASSESSMENT

To address these problems we propose a software-based KMS-HPA would enable "controlling" human performance information developed through the DMAIC process. The KMS-HPA would provide a capability for objective and reliable measurement at all levels—mission, organization, unit, team, individual. The vision is a web-based capability for defining, measuring, analyzing and improving warfighter capabilities. The KMS-HPA would enable authoring, storing, selecting, implementing, and exporting performance measures, metrics, and methodologies. Figure 1 presents a conceptual model of a KMS-HPA software tool for implementing the DMAIC process. Specifically, the KMS-HPA would enhance the DMAIC process by:

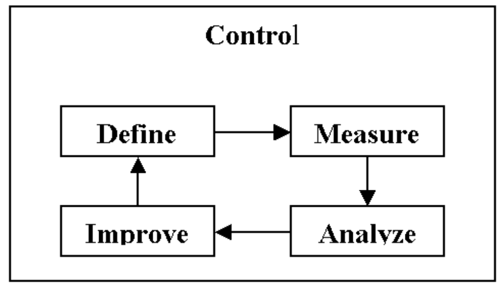

Figure 1: Conceptual Model for Managing Knowledge of Human Performance Assessment

Define

- Providing a database of lessons learned from various problem definition programs

- Creating a traceable and iterative approach to problem definition

- Evaluating the original problem definition based on the results

Measure

- Allowing measures, metrics, and methods to be standardized, re-used, and tailored for specific applications

- Reducing cost and labor expenditures required for assessment activities

- Reducing measurement error, inconsistencies, and redundancies

- Identifying consistent, standardized, and psychometrically valid measures

Analyze

- Facilitating comparison of performance and providing a capability for base-lining performance and for systematically tracking performance over time

- Improving psychometric properties of performance data, resulting in better forecasting and trend analysis, and making better decisions

Improve

- Improving reliability of assessment processes so that process efficiencies developed in the Definition phase can be realized

Control

- Maintaining traceability and maintenance of the DMAIC process

- Enabling previously stove-piped DoD activities (e.g., training and operational) to align measures across system lifecycles and military careers based on better communication, collaboration, and leveraging of measures and assessment data from one another

- Refining measures and methodologies as they are used by various activities

TECHNOLOGY DEMONSTRATIONS

Table 1 presents capability features of recent KMS-HPA initiatives. These exemplar prototype and operational programs demonstrate that there is a grass roots movement for enabling a Human Sigma strategy. These sponsors have responded to the need for a common knowledge base of measures and methods as an enabling technology, one that is capable of satisfying the Human Sigma goals for reducing variability in performance processes and increasing quality by making processes repeatable. Not all efforts have been captured in software programs and web-based sources. Similar efforts have appeared in published texts, such as that authored by Gawron (2000) and Stanton, Salmon, Walker, Baber, & Jenkins (2005).

U.S. Air Traffic Control

Rantanen and Vlach (2007) describe a measures database that was developed from an extensive review of the literature related to Air Traffic Control (ATC) Research and Development. The ATC domain is characterized by many of the same measurement issues facing DoD. They found a considerable demand for measures for many purposes. For example, scientific research and subsequent engineering applications within the domain are dependent on methods of measurement and critically need reliable and valid measures. In addition, increasing traffic volumes and collateral demands to improve aviation safety and air transportation system efficiency present perpetual challenges. Advancing technologies attempt to address these issues. These technologies, however, fundamentally change ATC system functionality and air traffic controllers' working methods, strategies, workload, and performance. Evaluation of the consequences of these new technologies, both on system capacity and safety, and on the working conditions and performance of controllers, is of utmost importance. Success of these evaluation efforts is subject to the availability of valid and reliable measures.

Measurement in ATC is often hampered by unique difficulties not found in other domains. Collection of usable data from an environment as dynamic, variable, and multidimensional as ATC is not a trivial matter. Many of the variables of interest are not directly measurable, such as controller workload, situation awareness, and performance. The intricacies of measurement in the ATC domain have been exacerbated by the advent of new automated applications. Scientific investigation of the impact of new technologies has become increasingly difficult due to the escalating number of variables and their interactions in the ATC operational environments. Traditional measurement techniques within the ATC domain, that is, subjective, over-the-shoulder evaluation of controller performance by other experienced controllers, have become inadequate in the face of present challenges. Objective measures are highly desirable because they can be: used in high fidelity, realistic ATC simulations and operational settings; routinely collected and analyzed concurrently and unobtrusively during the task; and subjected to data mining techniques to detect trends in system performance before any problems are manifested as incidents or operational errors.

The taxonomy of ATC measures, 65 classes of direct measures and 35 categories of indirect measures, allows for cross-referencing between different types of measures, their purposes, and the required data. This feature facilitates the development of comprehensive models of ATC performance, as well as the development of new measures as new sources of data become available. Its contents include measure names, measurement scales, directly and indirectly measured variables, independent variables used in research articles cited, derivative variables, and primary and secondary references. The current beta version has not been subjected to systematic usability studies. Expansion is based upon voluntary contributions of researchers and users within the ATC community.

Table 1: Knowledge Management System Human Performance Assessment Initiatives

	US Air Traffic Control	**Navy HSPAC Prototype**	**NATO C2 Assessment**	**United Kingdom HFI-DTC**
Developer	Federal Aviation Administration	US Navy	NATO and Partners for Peace (PfP)	The Human Factors Integration Defence Technology Centre (HFI DTC)
Content	Subjective and objective measures reported in ATC research	Acquisition-focused measures	300+ C2 variables; 200+ subjective and objective measures	Over 200 Human Factors methods and techniques
Psychometric Criteria for Content Inclusion	No	Content was not peer reviewed prior to submission for initial beta test	No, but measure attributes documented	Content drawn from extensive literature search
Taxonomy Capabilities	Search via any level of a taxonomy	Browse and search a taxonomy	Search via any level of taxonomy from a pull-down menu	Page reader format for Taxonomy
Software Application	MySQL open source, relational database	Protégé	MS Excel with Plone open source GUI; U.S. Navy Human Performance Assessment Tool (HPAT)	World Wide Web version conforms to Consortium (W3C) standards for: xhtml transitional (W3C XHTML 1.0) and W3C - CSS level 2
Availability & Life-Cycle Support	Available upon request; Facilitates continual updates	Access restricted to DoD employees and contractors. Was Tested on a .mil website but no longer available	Access restricted to NATO and PfP Countries, Web-Enabled is planned	For Human Factors practitioners; Copyrighted material; http://www.hfidtc.com; UK Ministry of Defense and Industry/Academia Consortium support

U.S. Navy Human Systems Performance Assessment Capability

In 2005, the Chief of Naval Operations sponsored a prototype Human Systems Performance Assessment Capability (HSPAC) program. This initiative coordinated efforts across Navy organizations to provide a Navy-wide capability to measure and assess human performance (Hughes & Davidson, 2007). Over an 18-month period, the Navy developed a prototype taxonomy of human systems performance measures to help define requirements for a shared repository that would be used to create, reuse, and store such measures (Winters and Pester-De-Wan, 2007). The taxonomy was incorporated into a temporary web-based prototype that allowed for a beta-test. Winters and Pester-De-Wan (2007) reported numerous lessons learned, including concerns that organizations and individual users would not be motivated to maintain and update a single Navy repository. As a response, the Naval Air Warfare Center Training Systems Division and the Naval Sea Systems Center Dahlgren Division recently collaborated on the development of a standalone prototype Human Performance Analysis Tool (HPAT) (M. Stretton, public presentation, March 20, 2008). The HPAT is a scenario-based performance measures authoring tool, using commercial off-the-shelf software, with menus that users select to develop assessment tools for individual and combat team performance. Scenarios and measures can be easily created and stored for re-use, employed to reliably collect and validate human performance data, and analyze results. The advantage of such a tool is that users can create and manage their own repository of scenarios and measures, and can control sharing them with other users. The HPAT design is currently under a beta-test by the North Atlantic Treaty Organization (NATO) Command and Control (C2) assessment program (see below).

NATO C2 Assessment

In October 2006, the NATO Research and Technology program, through its Human Factors and Medicine (HFM) panel, appointed a 3-year research task group (RTG-156)--"Measuring and Analyzing Command and Control Performance Effectiveness--to develop a KMS for C2 measures. RTG-156 will identify and recommend C2 measures and metrics that will be useful for future C2 systems (e.g., Network Enabled Operations), thereby establishing the assessment capabilities and deficiencies for this rapidly evolving area. The tools will be web-enabled so that NATO and Partner for Peace (PfP) nations can access and utilize it. The rationale behind this effort is that such a knowledge base could:

- Identify, examine, and catalogue C2 performance assessment tools (measures, metrics, methods, and tools) utilized by the NATO Alliance;
- Identify assessment deficiencies of human performance in C2 systems; and
- Make recommendations regarding use of the tools and share lessons learned gleaned from the analysis so that the nations can learn from each other's successes and failures in the area of C2 measurement and assessment.

The HPAT is being used as an initial repository for the C2 measures and it is being tested for its ability to support the user in identifying and pulling measures into an assessment plan. Eventually, it could be used in evaluations for training or operational exercises. Future growth and population of the knowledge base will depend on NATO implementation of HFM RTG-156 recommendations to be provided in their 2009 final report.

United Kingdom Human Factors Integration Defence Technology Centre

The United Kingdom (UK) Ministry of Defence (MoD) has established several Defence Technology Centres (DTCs) for the purpose of collaborating within the broader UK science and technology community to build better defense capabilities. The Human Factors Integration Defence Technology Centre (HFI-DTC) is supported by a formal consortium of government, industry, and academic stakeholders, and maintains a web-based database called the "Human Factors Design and Evaluation Methods Review." The database serves to describe and evaluate HFI methods in the design lifecycle for C2 systems. While not strictly a performance measurement and assessment KMS, this database includes many such techniques or methods. The database is also a published book (Stanton, Salmon, Walker, Baber, & Jenkins, 2005).

Software Engineering Repositories

Software measurement repositories are being employed extensively within the field of software engineering. In this section we provide a brief review of capabilities, but it is far beyond the scope of this paper to describe them all. Software engineering relies heavily upon the reuse of software code, models, and metrics. Software measurements refer to the qualitative and quantitative assessment of the software development process, the resources used in development, and the software product itself. Employing software assessment tools leads to: a) improved management of software projects; b) better understanding of software development practices; c) improved software design and reuse; and d) empirically validated novel ideas and techniques (VanHilst, Garq, & Lo, 2005). The International Institute of Electrical and Electronics Engineers holds conferences on mining software repositories, and lessons learned from these techniques, as well as the data mining software, can be re-used in performance measures repositories for Human Sigma activities.

RECOMMENDATIONS

Following review of the efforts in Table 1 we identified four major capabilities that should be addressed in KMS-HPA development to ensure optimal workflow: 1) submission of new performance measures data, 2) post-submission (e.g., content verification and cataloguing), 3) preservation (short and long-term storage), and 4) structural management (user and user group management, archive structure and content management, policies and authorizations) (Zuccala, Oppenheim, & Dhiensa, 2008). Software and usability issues, such as DoD software restrictions and the need to interface with other systems and programs, span these workflow areas and must be considered prior to and during development. Next, we provide key technical and organizational recommendations that apply to the workflow areas listed above.

Technical

Make the KMS-HPA Interface User Friendly
Ease of data entry is an especially important area. If data entry is too time-consuming or difficult, users cannot be expected to willingly update the system. The KMS-HPA cannot grow and evolve over time if users will not enter new information.

Establish a Measures Taxonomy
A taxonomy is a hierarchical structure used to organize and classify constructs and measures. Whenever this approach is employed, difficulties are likely to arise with categorizing measures into the appropriate taxonomic category and level. Experience has already demonstrated that consensus on placement of new measures within the taxonomy is frequently lacking. Another related concern is that users of the taxonomy may have difficulty understanding nuances between constructs and pull the wrong measures out of it for their use. Even so, taxonomies are probably the most used or convenient method for organizing such repositories. Alternative approaches, such as meta-tagging the data, associations, and latent semantic analyses, may offer more robust search capabilities.

- With taxonomies, users generally require a visual depiction of the entire taxonomy when trying to decide where to place a new measure or metric in the taxonomy.

- Viewing the entire taxonomy in a graphical format makes it easier to identify potential locations for a metric.

- Annotations help to clarify some of the more subtle distinctions between related categories.

- Screening of new measures is needed to validate its placement within the taxonomy and to establish psychometric attributes for acceptance.

- The goal for mapping is generally to place the measure or metric as far “down” inside the taxonomy as possible, that is, to the lowest level possible

- One potential capability might be to create different taxonomy views for user communities and allow users to choose what view or combination of views they prefer to see.

Establish a Measures Lexicon
Accurate, complete, and valid definitions are necessary for all relevant terminology. It can be more difficult than imagined to obtain consensus on some terms. Use by multiple communities might require a lexicon of alternate terms, that is, common concepts identified by terms appropriate to each community, but with different terms specifically noted.

Establish Level Of Measurement Granularity
A significant issue is how to map measures to taxonomy variables. In other words, should entire measurement instruments be mapped to the taxonomy variables, or should specific items within instruments be mapped? The answer seems to depend on the taxonomy. If the variables are more general or inclusive, then it may be more appropriate to map entire instruments, even though their items address several constructs. If the variables are very specific in nature, then it may be necessary to extract from an instrument only those items that relate to the variable of interest. This is a problem encountered by the NATO group. The NATO C2 Reference Model being utilized as a taxonomy has over 300 third-level variables, most of which are very specific in nature. Extracting items from validated instruments and mapping them to these specific variables carries with it the problem that the items are no longer considered psychometrically valid. However, the HFM-RTG 156 group determined this was the reasonable and necessary approach, both to provide the needed utility to the knowledge base and to satisfy the deliverable requirements to NATO. An additional benefit from this approach is the database could be used to author new measures. The variables of interest can be assembled, along with their measurement items, into a newly constructed instrument that could then be tested and validated.

Make Measures Available For Reuse
A key risk is that limited information, in terms of measures, metrics, and methodologies, may be found. Additionally, few measures and metrics have actually been validated, so the psychometric quality (i.e.,

attributes such as reliability, validity, specificity of measures, etc.) of identified instruments may be unknown. However, by making the instruments available for reuse, the likelihood of future validation increases. Some measures, metrics, and methodologies will be more suitable for reuse than others, in that they are adaptable to multiple communities, multiple user expertise levels, and multiple contexts. For all measures, a standardized list of measure attributes should be included to help the user understand the differences between measures and help them make selections among available options.

Identify Performance Standards

If performance results are included in the KMS-HPA, those results can be used to establish future standards of performance. While measures and metrics are the tools needed to conduct evaluations, performance standards are the necessary companions, as they provide the comparison points to ascertain whether the performance results are good or bad. It should be clarified that performance standards are specific to a context or application, and while there is no reason that they could not be included in a KMS (with the exception of security issues), we are not specifying that these should be included up front in a metrics KMS. One further note about results (reports) is that the level of detail is likely to vary across reports, so decisions regarding standardization of information to be included may be needed. One option is to simply allow the amount to be determined by what information is available.

An additional information source within the KMS-HPA could be the requirements that drive assessments. In other words, requirements could be tied to the measures, metrics, and assessment reports. Requirements for assessments are often high-level or abstract, and are often expected to be user-defined. This capability would provide accountability that the assessment covers the intended topic.

Define Search Engine Requirements

Search capabilities are a key and critical requirement for the KMS-HPA. As multiple communities use different terminology to refer to the same behaviors, it will be important that the system be robust in returning the same information each time. This was the rationale for an ontology approach used in the US Navy HSPAC tool. It is a means by which users, using different terminology, can store, access, and manipulate repository content. The Navy's ontology used a taxonomy structure.

- Experience indicates that having multiple search mechanisms (e.g., browsing the taxonomy hierarchy and free text search) are beneficial.

- Linking measures—for example, measures that have been used together in a previous evaluation—can be useful for providing the current user with a more complete set of measures for his or her planned application.

- Book marking items of interest can help a user easily return to the content of interest.

The Navy identified a list of desired enhancements for the HSPAC repository and many of them focus on the search functionality. Some examples are:

- Correlations based on similarities of measures: Develop a visualization tool that easily conveys to users how measures are similar to one another. For example, provide support with phrases similar to those found online: "if you are interested in this topic, you may also find the following of interest."

- Correlations based on usage statistics: Tracking the measures different users employ or access over time could indicate patterns in measure selection that would then be useful to other users. Knowing how measures have been combined or coordinated over time could help users develop more comprehensive plans.

- Intelligent automated selection: Intelligent agents could assist users in selecting the most appropriate measures for the specified context. Different types of search engines will need to support different levels of users (expert vs. novice). Intelligent agents would need to recognize similarities across different operational tasks or other attributes.

- Organization of measures in a category: For measures within a category, it may be necessary to provide different attributes of the measures (e.g., name, organization, etc.). This would reduce the likelihood of duplication and facilitate the user finding information. It would be an essential tool for categories with a significant number of measures.

- Complement searches with functions that assist the user in developing an assessment plan, through the use of templates or software wizards. Tools and measures the user wants to employ would be selected and tracked, and the user would continue to update the information.

Organizational

Organizational Support And Resources Are Key To Developing And Maintaining A KMS-HPA
Many questions about developing and maintaining a KMS-HPA emerged in the review of capabilities described in Table 1. The first step is identifying the desired scope of such a capability:

- Should it be built as a multi-service tool with requirements from all Services built in up front?
- Should it be built as a single Service tool, to be used as an example for other Services?
- Should a Service be assigned to take the lead?
- Should the tool be built from the bottom up or from the top down—with senior levels involved from the beginning, or as a grass roots movement gaining momentum?
- If the latter, can we merge what currently exists?
- Is a research prototype needed to minimize risks and resolve challenges?

A lesson learned is that having multiple grass roots efforts may disengage some stakeholders, and this should be prevented if the desire is to have a KMS-HPA that applies across many communities and organizations. Therefore, an ideal approach is to build a KMS-HPA built with organizational support, resources, and direction from the top, rather than try to integrate individual systems. It would make sense to establish an Integrated Product Team (IPT) to lay out functional requirements and software specifications for DoD, and establish it as a program of record with dedicated funds. The scope desired (e.g., single Service, DoD, or coalition) would ultimately determine which activity should provide oversight. The IPT should carefully address the issues of acceptance, use, and sustainment, as these are the key factors for success.

Community Support is Required.
With support from the top, then user community support, or "buy-in" from all communities is required. Awareness needs to be developed. Visibility for the tool is enhanced if multiple communities promote it concurrently, thus enabling faster implementation. Although a collective synergistic effort can produce a better product, one of the largest obstacles is to motivate individual organizations to post information if they fail to see a direct benefit for their organization. They may be motivated to post information if they would use it internally, but they would be less motivated to post it if it is only for the benefit of other organizations. Additionally, a reluctance to release information outside of an organization—especially information connected to evaluations of performance may be an obstacle to implementation. To overcome this, formal agreements, policy mandates, and incentives might be required to stimulate use. As long as the functionality, organization, and content support the needs and goals of the projected users, resistance may be overcome in time.

Accessibility Should Be Balanced With A Policy For Information Security
The KMS tool needs to be available and accessible to targeted users. Security issues need to be addressed and resolved early on in development. For DoD, multi-level access, spanning different security levels, is needed if the example of the Navy HSPAC prototype is followed, and performance results and limited distribution documents are included. If the KMS-HPA is limited to the identification of measures and methodologies, and no performance data are incorporated, then security issues can be managed in each individual organization, as is the case with the HPAT.

CONCLUSION

The DoD objective is achieving the highest levels of mission readiness possible. It successfully transitioned to a Six Sigma model where improved process efficiency and cost savings are paramount. However, to achieve both goals a Human Sigma approach is required to address performance processes and deliver measurable positive impact on human capacity to perform. The need for performance measurement and assessment underlies much of the acquisition and design of platforms and systems, selection and training, and operational performance. Therefore, a KMS - HPA has the potential to become a critical Human Sigma tool to address 21st century military needs by improving processes for measuring performance effectiveness to enhance future readiness. Although such a tool could be built as an enabling capability for individual commands, agencies or military services, a DoD-level system would realize optimal efficiencies and cost savings.

REFERENCES

Fleming, J.H, & Asplund, J. (2007). *Human Sigma: Managing the Employee-Customer Encounter.* Gallup Press.

Fleming, J.H., Coffman, C., & Harter, J.K. (July-August, 2005). Managing your Human Sigma. *Harvard Business Review—The High-Performance Organization.*

Gawron, V.J. (2000). *Human Performance Measures Handbook.* London: LEA.

George, M.L., Rowlands, D., Price, M., & Maxey, J. (2005). *The lean six sigma pocket tool book.* NY: McGraw-Hill.

Hughes, G., & Davidson, K. (2007). Human performance measurement: cultural changes. *Naval Engineers Journal, 119* (1) , 65–69.

Miller, J. (December 3, 2007). DoD is serious about Six Sigma. *Federal Computer Week.* Retrieved May 28, 2008, from http://www.fcw.com/print/13_42/management/150925-1.html.

Rantanen, E.M., & Vlach, P.M. (2007). Development of Air Traffic Control measures database. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting,* 1124-1128.

Robinson, B. (March 3, 2008). DoD rallies around Lean Six Sigma. *Federal Computer Week.* Retrieved May 28, 2008, from http://www.fcw.com/print/22_5/features/151766-1.html.

Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., & Jenkins, D.P. (2005). *Human factors methods: A practical guide for engineering and design.* Hampshire, England: Ashgate.

VanHilst, M., Garq, P.K., & Lo, C. (2005). Repository mining and Six Sigma for process improvement. *Paper in the Proceedings of the 2005 International Workshop on Mining Software Repositories.*

Winters, J., & Pester-DeWan, J. (2007). Developing a repository for human performance measures: Possibilities, preconditions, and pitfalls. *Proceedings of the 2007 Human Systems Integration Symposium.*

Zuccala, A., Oppenheim, C., & Dhiensa, R. (March, 2008). Managing and Evaluating Digital Repositories. *Information Research, 13* (1). Retrieved June 2, 2008, from http://informationr.net/ir/13-1/paper333.html