Development of an Instructor Aid to Diagnose Performance

Meredith B. Carroll, Roberto K. Champney, Laura M. Milham, David L. Jones, Dervon Chang*
Design Interactive, Inc.
Oviedo, FL
Meredith, Roberto, Laura, David@designinteractive.net

***Now with Basic Commerce & Industries, Inc.**

Glenn A. Martin

Institute for Simulation and Training
Orlando, FL
martin@ist.ucf.edu

ABSTRACT

Given the increasing complexity of team simulations, there is a need to aid instructors by providing tools that can facilitate After Action Review (AAR). Without these tools, instructor evaluation and feedback might not be optimized due to human limitations in attention and memory (e.g., capability to simultaneously monitor and remember intricacies of all team members' performance). By providing a tool which can facilitate detection of root-cause errors, the diagnostic value of the AAR or system feedback provided to trainees may be improved, potentially increasing training efficiency by targeting the specific skills that need improvement.

The Performance ASsessment and diagnostic Tool (PAST) addresses these needs by 1) presenting instructors with the pertinent type and amount of information to properly assess training performance and 2) providing trainees in a deployed setting, with a "Trainer in a Box", allowing them targeted feedback about specific errors on which they need to focus. PAST achieves these goals through 1) performance measurement via system collected and semi-automated observer collected metrics, 2) performance diagnosis of root cause via metric relationship logic, 3) interpretation of performance trends through metric linkages to relevant training goals/objectives, and 4) intuitive and usable presentation of performance diagnosis and interpretation to facilitate an understanding of the root-cause of performance decrements and areas in need of improvement. PAST was developed in close collaboration with United States Marine Corps (USMC) subject matter experts (SMEs) and validated with instructors at a range of different schoolhouses, including a series of evaluations testing its functionality and usability. This paper discusses PAST functionality, evaluation and testing data collection with SMEs, results and the redesign resulting recommendations..

ABOUT THE AUTHORS

Meredith Bell Carroll is a Senior Research Associate at Design Interactive, Inc. and is currently supporting training system design and evaluation for the office of Naval Research's Human Performance, Training and Education (HPT&E) Program. Her work focuses on Task Analysis, HCI design of VE training environments, design of Performance Measurement and Training Management tools and design and conduct of training effectiveness and transfer evaluations and experimentation. Her research efforts have been directed towards human/team performance and training in complex systems in aviation and military domains, with focuses on performance measurement and virtual training technology. She received her B.S. in Aerospace Engineering from the University of Virginia in 2001, her M.S. in Aviation Science from Florida Institute of Technology in 2003 and is currently a Doctoral Candidate in Human Factors and Experimental Psychology at the University of Central Florida.

Roberto K. Champney is a Senior Research Associate at Design Interactive, Inc. and has had significant experience in systems design and evaluation having performed several system evaluations in industry and academia. His has worked on several multidisciplinary teams where he has developed an expertise in capturing the Voice Of the Customer (VOC) for transformation into design requirements and specifications. Most recently he has worked in developing methodologies for performance assessment and diagnosis in training systems. Roberto holds an M.S. in Human Engineering and Ergonomics, and is an ABD Ph.D. candidate at the University of Central Florida. His dissertation research interests are in Design, Usability and Emotion, primarily in the use and development of integrative tools and methodologies for informing the design and evaluation of products with regards to user experience criteria.

Laura M. Milham, Ph.D. is the Director of Training Systems at Design Interactive, Inc. She received her doctorate from the Applied Experimental and Human Factors Psychology program at the University of Central Florida in 2005. She has worked as Principle Investigator in support of the development and assessment of the effectiveness of training systems design. For this, she has performed task analyses; assessed design and training needs for VE systems; performed usability evaluations; developed design recommendations for system HCI; developed research plans; conducted Training Effectiveness Evaluations (TEEs) and Transfer of Training (TOT) evaluations. She has also been responsible for the design of training content, from scenario development to human performance assessment toolkits for individuals and teams. In addition she has developed guidelines for training tactical teams.

David L. Jones is a Senior Research Associate at Design Interactive, Inc. His primary research focuses are the design of multimodal interactive systems and training systems. He has a Masters in Industrial Engineering from the University of Central Florida, with a focus on human-computer interaction (HCI) and usability. He has a Bachelors in Human Factors Psychology from Embry-Riddle Aeronautical University, with a focus on human factors and computer science. To support the Navy's VIRTE and HPT&E efforts he has performed TEE's and TOT evaluations on advanced training systems in various stages of the development process. Under this work has experience performing detailed task analyses and developing metrics to evaluate transfer performance with a particular focus on doing so under the special constraints associated with a military field environment.

Dervon Chang is a Human Factors Engineer at Basic Commerce & Industries, Inc. (Dahlgren, VA) supporting the Human Systems Integration (HSI) division. She currently works on subcontracted projects for the USMC and DHS developing HSI inputs for requirements documents, including efforts toward a Program of Record (PoR) for Defense Acquisition. Past work includes evaluating operator-system interfaces, conducting usability testing, developing HSI redesign recommendations, and building training effectiveness evaluation (TEE) approaches suited for military training programs. Ms. Chang received her B.S. in Psychology at the University of Illinois Urbana-Champaign in 2003, where she is also earning a M.S. in Psychology from the Visual Cognition Human Performance (VCHP) department (based in the Institute of Aviation). Her academic research experience has focused on multi-modal, multiple resource related work with Unmanned Aerial Vehicles, UAV and pilot-ATC communications, with extensive use of Frasca simulators and head-mounted eye trackers.

Glenn A. Martin is a Senior Research Scientist at the University of Central Florida's Institute for Simulation and Training where he leads the Interactive Realities Laboratory, pursuing research in multi-modal, physically realistic, networked virtual environments and applications of virtual reality technology. He has worked on numerous projects related to virtual reality including a testbed for evaluating VR for training uses, an investigation of network architectures, a library for VR applications and the study of human factors. He received his B.S. in computer science in 1992, and his M.S. in computer science in 1995 both from the University of Central Florida.

Development of an Instructor Aid to Diagnose Performance

Meredith B. Carroll, Roberto K. Champney, Laura M. Milham, David L. Jones, Dervon Chang*
Design Interactive, Inc.
Oviedo, FL
Meredith, Roberto, Laura, David@designinteractive.net

***Now with Basic Commerce & Industries, Inc.**

Glenn A. Martin

Institute for Simulation and Training
Orlando, FL
martin@ist.ucf.edu

INTRODUCTION

Without useful After Action Review (AAR) tools, the evaluations and feedback provided by instructors may not be optimized. Performance evaluations rely on the ratings of instructors possessing domain expertise, but often due to the dynamic and complex nature of the training task they do not address the underlying sources of poor performance (i.e., they do not drill down far enough to expose the root cause of training deficiencies)._ Rather, they focus on outcome measures or proximal errors. For example, in the CAS domain, instructors might determine that a student has difficulties scanning and detecting aircrafts in the sky for failing to do so, when the real root cause of this difficulty could have stemmed from preceding errors or vulnerabilities propagating through subsequent tasks (e.g., the timing of the mark was planned too close to the aircraft TOT, requiring that the student search for the mark when the aircraft is on the final approach). Additionally, given the need to allow deployed troops to maintain high levels of skilled performance without access to formalized training and expert instructors, it is necessary to provide training management tools in hand with simulators to turn practice opportunities into actual training opportunities which promote skill improvement through diagnostic feedback. By providing a tool which can facilitate instructors and trainees in detecting the root cause of error, the diagnostic value of the instructor's AAR or system feedback provided to deployed warfighters may be improved while at the same time potentially increasing training efficiency by targeting the specific skills that need improvement.

The Performance ASsessment and diagnostic Tool (PAST) was developed to address these needs by providing an effective and efficient AAR tool aimed at supporting both instructors in training and trainees in a deployed setting. The proposed approach to accomplish this was through two primary goals. The first aims to present instructors with the optimal amount of pertinent information to properly assess training performance by providing a high-level, global picture of performance and errors allowing instructors to drill down in more detail to find the root cause of training deficiencies. The second goal aims to provide trainees in a deployed setting with enough diagnostic information regarding performance to provide them with some guidance from an "Instructor in a Box", which provides them detailed feedback about specific errors they need to focus on.

In order to enable these capabilities in a system, four basic general functionalities are needed, 1) performance measurements, either via system, manual or semi-automated observer collection, 2) performance diagnosis or interpretation of the performance measurements (e.g., analysis of root causes via predefined logic), 3) interpretation and linking of performance trends and areas in need of improvement to relevant training goals and objectives, and 4) presentation of performance diagnosis and interpretation to users in a manner which facilitates an understanding of the root cause of performance decrements and areas in need of improvement.

PAST was designed as a training management component to work with the Multi-platform Operational Team Training Immersive Virtual Environment (MOT^{2}IVE), a USMC Fire Support Team (FiST) training system, and developed by University of Central Florida's (UCF) Institute for Simulation and Training (IST) as on overlay to the Dismounted Infantry Virtual After Action Review System (DIVAARS) under the Office of Naval Researches (ONR) Virtual Technologies and Environments (VIRTE) Program. It was intended to work as a stand-alone application, which in conjunction with a data logging and metric calculation application (e.g., AAIRS; DIVAARS) can yield contextually meaningful metric output valuable for diagnosing training performance. PAST was developed in close collaboration with Subject Matter Experts (SMEs) and tested and validated with instructors at a range of different schoolhouses. This document describes in detail the functionality of PAST, including the steps taken in development and evaluation of the tool.

PAST FUNCTIONALITY

PAST was designed as a tool to facilitate AAR, with aims at supporting During Action Review (DAR) as the tool's concept evolves. The key to the diagnostic functionality offered by PAST is the internal logic linking each performance metric to other metrics in cascading relationships which represent the impact the performance or outcome of the first metric has on subsequent others, including final mission outcome metrics. This allows the capability to capture and flag errors as they occur throughout a mission with which root-cause relationships may be flagged during post mission. During post mission, the metrics containing errors and the linkages between them enable the creation of graphical representations illustrating error propagation from the initial metric (i.e., the top event, error source) to the final error represented by an outcome measure.

The four general functionalities mentioned earlier, (1. performance measurements, 2. performance diagnosis 3. linking of performance trends and areas in need of improvement to relevant training goals and objectives, and 4. presentation of performance diagnosis and interpretation) are instantiated within three main components in PAST: A Criteria Module, a Team Performance Data Observer Capture Tool (TEAM-DOC), and a Diagnostic Assessment Tool (DAT) as illustrated in Figure 1.

The Criteria module is a domain based representation of the task in the form of performance assessment measures or events, causal links between these measures (Diagnosis), and interpretation links between metrics and training objectives, trainees, mission phase, and training goals.

The TEAM-DOC tool is a module which provides access to observer collected measures which are presented to the rater as checklists of metrics or events.

The DAT is the output module allowing presentation of the data collected to instructors in a usable format such as root-cause diagrams, mission timeline to allow assessment of problematic time periods and mission phases, and performance summaries presenting aggregate data associated with training objectives, trainees, or mission phases.

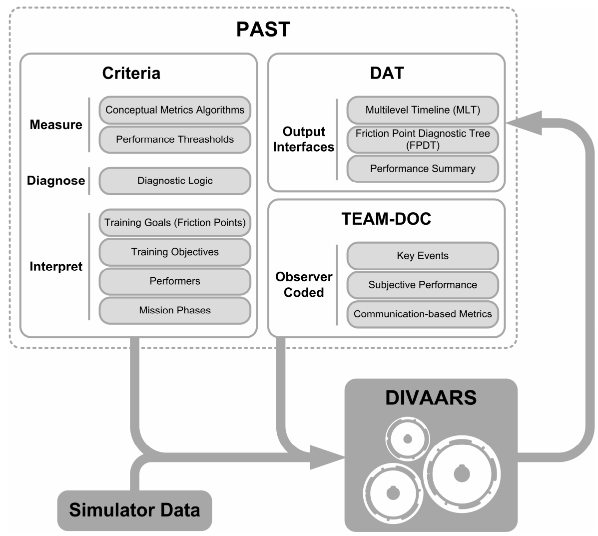


Figure 1. PAST Functionality

Measurement

An integral part of PAST is the metrics of which it is comprised. Some metrics depend solely on system collected data, while others rely on observer collected data (instructor/operator input; i.e., real-time event tagging or instructor evaluation input during or post mission) collected via a semi-automated event-based checklist.

PAST's TEAM-DOC allows the collection of observer assessed metrics via a semi-automated event based checklist that allows for objectivity and standardized performance data capturing,. TEAM-DOC is used in parallel with the data capturing application to allow instructors the ability to confirm task occurrences, code scenario events and communications, and leave amplifying remarks regarding performance either in real time or post-hoc. The TEAM-DOC input data is then fed forward to the independently running output module (DAT).

The TEAM-DOC's interface is designed to allow users to indicate (via a checkbox) whether or not an event/metric has occurred during an ongoing training scenario. Additionally, for selected metrics, the TEAM-DOC interface prompts users to enter important, task-relevant information required for an accurate assessment of error propagation (e.g., grid location communicated). In addition, system collected measures are captured via the Dismounted Infantry Virtual After Action Review System (DIVAARS).

Thresholds for each metric are set by which good/poor performance is gauged based on domain requirements. Metrics are calculated utilizing system collected data from the simulation regarding events occurring in the simulation and information entered into the semi-automated event-based checklist. Preset metric thresholds allow PAST to detect when a metric is not in line with expected performance and appropriately flag the metric for presentation to the instructors or trainees.

Diagnosis

Diagnosis is enabled by a logic-based framework in which cause-effect interconnections between metrics allow for potential error propagation trees (i.e., diagnostic trees), including root detected errors, to be identified and presented to the instructor/operator. This interconnection is in the form of causal relationships between domain-based tasks and event metrics that represent, in chronological order, the decomposition of a complex task (e.g., a mission objective). Such decomposition and relationship linking enable the identification of failures along the process flow (mission task), thus enabling PAST to suggest to an instructor where an actual failure occurred. In addition to each metric or event being linked to a process, these may have additional association to such things as type of skills (e.g., communication, procedural), mission phase, etc. This allows the possibility of additional diagnostic capabilities yet to be explored automatically through PAST.

Interpretation

Within PAST, each performance metric has associated training goals (e.g., Suppression of Enemy Air Defense) labeled as "Friction Points," Training Objectives (e.g., Team Coordination), performers (i.e., FAC, FO or FiST Lead), and mission phases (e.g., mission planning). Such a relational database allows the interpretation of which aspects of training are in most need of further attention. The current instantiation of PAST only provides prioritization of results based on Friction Points, yet having these additional relational data allows manual determination of further training priorities.

Presentation

PAST's Diagnostic Assessment Tool (DAT) contains output interfaces which present the data in both error propagation and timeline format as well as performance tables. This is intended to allow instructors to identify trainee weaknesses and performance trends with respect to Friction Points (i.e., training goals), Training Objectives, Performers and Mission Phases, to facilitate AAR feedback and aid in the selection of future training scenarios in order to drive the training focus. The DAT output component contains three (3) main functional sections: Friction Point Diagnostic Tree, Multi-Level Timeline (MLT), and Performance Summary.

These output interfaces were designed to present the correct *level* of information to instructors during an AAR, and the correct *way* to present it. Specifically, during a single scenario, many different events quickly accumulate and can prove to be difficult to dissect and later grouped into meaningful data. This useful but large amount of data can eventually prove to be useless and frustrating as it quickly becomes information overload to the user while also inhibiting the discovery of trends that the data represents (Chung, Chen, Chaboya, O'Toole, & Atabakhsh, 2004). The PAST interface has been designed to incorporate several of the philosophies currently found in the information visualization field. Information visualization has the capability of improving the human ability to process large amounts of data (Card, Pirolli, Mackinlay, 1999). First and foremost, as encouraged by Tufte (2001), good data speak for themselves. In this case, the PAST data must, at first glance, provide useful information, relevant to the instructor, with little to no difficulty in its retrieval process. This information must quickly give the instructor the information he or she needs at a high level. Secondly, the methodology, "Overview first, zoom, and filter, details on demand" (Schneiderman, 1996), employed by the PAST interface not only gives the instructor a high level overview, but allows him to take a closer examination of the data, then determine which data is most significant, and access the details associated with the most significant data. This process allows the instructor to further access the details of the errors he wishes to examine, thus allowing for relevant patterns to be formed by the instructor's expertise. The interfaces within PAST were designed to be user-centered by ensuring they provide meaningful information derived from a voluminous set of data in a usable form in order to diagnose errors to the instructor's desired level of detail.

PAST's DAT presents the instructors with three output types: a graphical representation of the diagnostic tree (and associated metric characteristics), a multi-level mission timeline, and a tabulated performance statistics.. These outputs allow instructors and trainees to understand the chain of events leading up to each error allows, identify earlier mistakes or other contributing factors that may have led to overall performance errors (outcomes), identify problematic mission phases, and allow operators to target specific portions of the mission.

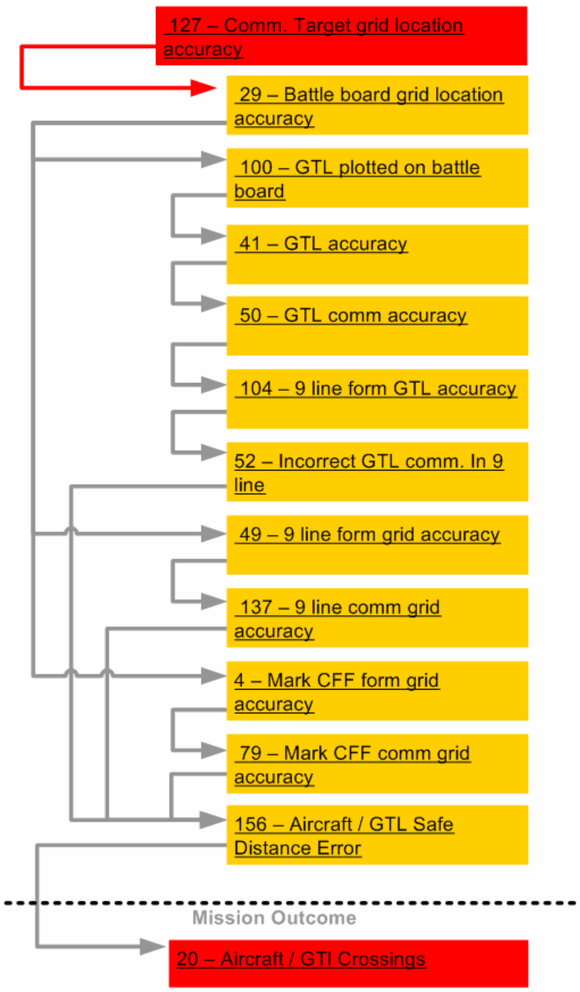


Figure 2. Section of Diagnostic Tree GUI

Friction Point Diagnostic Tree

The Friction Point (FP) Diagnostic Tree interface is designed to present the instructor with a 'big-picture' snapshot of the training scenario results, summarizing the entire training simulation with respect to the FPs. Summary data is organized and arranged by FPs to facilitate an instructor's ability to conceptually understand the causal relationships across failures during the mission. Specifically, PAST highlights how task failures propagate from an initial root failure to a potentially failed outcome metric, providing an additional level of insight to help instructors develop a more accurate diagnosis of team performance.

The error propagation format presents output data in clusters of related metrics which form a diagnostic tree. Figure 2 presents a small section of the graphical user interface (GUI) highlighting this propagation. These trees contain the outcome measures in which performance was not as expected, the failed metrics along the causal chain leading to the outcome metrics, and highlight the first failed metric in the tree (i.e., the root of the chain of errors). Rather than provide a single diagnostic tree, PAST is subdivided in seven (7) Friction Points (key mission performance components and often utilized as training goals in the domain). As such, corresponding metrics are grouped within each Friction Point.

Multi-Level Timeline

The multi-level timeline interface is designed to provide a chronological view of the scenario, which invites an opportunity to make performance assessments based on when events took place. This panel displays the timeline of simulation events (e.g., enemy/friendly actions) as they occurred in the scenario along with performance errors as indicated by performance metrics which did not meet minimum threshold values. Simulation events are presented on the bottom portion of the timeline while performance errors are presented above the timeline and color coded based error type (i.e., root error or not). This timeline format is designed to allow additional information such as physiological measures (e.g., Workload Arousal Meter; Hoover & Muth, 2004) to be layered on top of the performance measurement in the future to allow a more diagnostic assessment of problem areas. For example, given a cluster of root errors at close proximity in time, a workload/arousal reading at that point in time could indicate to instructors that trainee overload was an issue and may be a contributing factor to breakdowns. Figure 3 illustrates the multi-level timeline interface. These offer the instructor a more consolidated picture of performance decrements, including initial errors which may have caused later problems in performance as well as when the error occurred with respect to the timeline or trainee state.

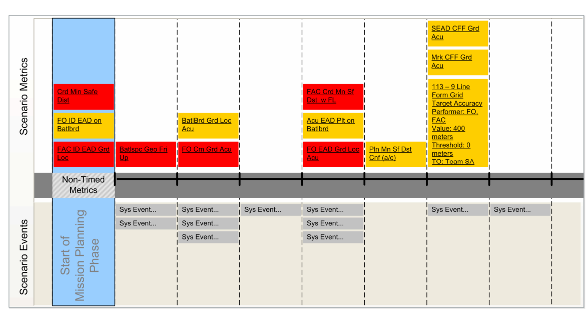
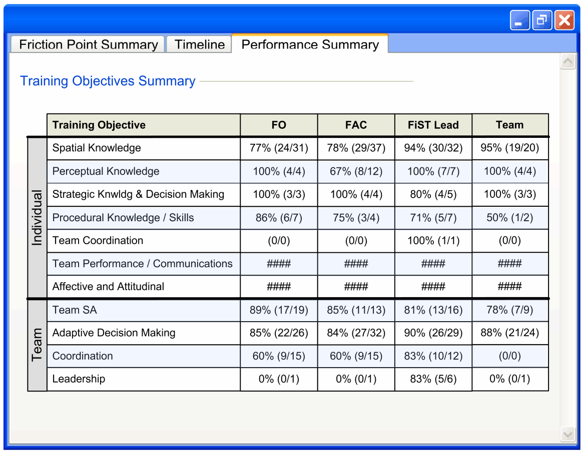


Figure 3. Section of Multi-level Timeline GUI

Performance Summary

Finally, the Performance Summary interface is designed to present individual-level performance data of trainees across multiple Training Objectives, Mission Phases, and Friction Points. This interface was designed to aid in interpretation of performance. Performance summary tables present individual-level performance data offering the opportunity to understand performance differences between team individuals as well as see and track performance trends. Figure 4 illustrates the performance summary interface.



	Training Objective	FO	FAC	FiST Lead	Team
Individual	Spatial Knowledge	77% (24/31)	78% (29/37)	94% (30/32)	95% (19/20)
	Perceptual Knowledge	100% (4/4)	67% (8/12)	100% (7/7)	100% (4/4)
	Strategic Knwldg & Decision Making	100% (3/3)	100% (4/4)	80% (4/5)	100% (3/3)
	Procedural Knowledge / Skills	86% (6/7)	75% (3/4)	71% (5/7)	50% (1/2)
	Team Coordination	(0/0)	(0/0)	100% (1/1)	(0/0)
	Team Performance / Communications	####	####	####	####
	Affective and Attitudinal	####	####	####	####
Team	Team SA	89% (17/19)	85% (11/13)	81% (13/16)	78% (7/9)
	Adaptive Decision Making	85% (22/26)	84% (27/32)	90% (26/29)	88% (21/24)
	Coordination	60% (9/15)	60% (9/15)	83% (10/12)	(0/0)
	Leadership	0% (0/1)	0% (0/1)	83% (5/6)	0% (0/1)

Figure 4. Performance Summary GUI

PAST DEVELOPMENT

PAST development took place over the course of two fiscal years and included several stages of design, development and validation. The following paragraphs detail the stages of PAST design and development.

Conceptual Design

The first stage of development centered on conceptual design of the PAST tool. Driven by the need for more diagnostic performance assessment, the PAST conceptual model drew on root cause analysis methods from the accident investigation literature as well as performance assessment literature (e.g., Benner, 1975; Borgonovo, Smith & Apostolakis, 2000). This concept was further combined with the idea of a multilevel timeline (MLT) (e.g., Potter & Woods, 1991) which is a variation of the chronological timeline often use in significant event analysis (Concoran, Swanson, Evans & Cooley, 1997). The goal of the MLT was also to provide instructors with a more diagnostic picture of performance by presenting, at each point in time, throughout the mission both performance metrics and additional measures. This layering would allow the addition of measures such as physiological measures onto this timeline in order to provide insight into why performance was unfolding a certain way (e.g., Was the trainee experiencing high workload? Was the trainee experiencing high levels of stress?). Based on this concept, the PAST design began to emerge.

Metrics Definition

The second step in development of PAST was defining the metrics. Based on document reviews of Training and Readiness (T&R) manuals and field manuals detailing Tactics, Techniques and Procedures (TTPs) and simulation and live fire training observation, metrics were developed. Time, accuracy and communication metrics for each subtask within the FiST mission were defined this way. Through a series of collaborative meetings with USMC Tactical Training and Exercise Control Group (TTECG) instructors known as "Coyotes", the metrics were refined, including additions, deletions, and redefinitions. Also, with the help of the instructors, metric criticality ratings for prioritization purposes and relevant *friction points* (i.e., training gaps) to which the metrics were relevant were identified. What resulted after several iterations was a comprehensive list of approximately 150 performance metrics covering performance of a 3-man FiST team (Forward Air Controller [FAC], Forward Observer [FO], FiST Lead) for a full Close Air Support (CAS) mission including Call For Fire (CFF) and Suppression of Enemy Air Defense (SEAD) spanning mission phases from planning to execution.

Metric Relationship Logic

In order for PAST to be able to identify the root cause of performance errors, cause-effect relationships between the defined performance metrics needed to be established along with metric performance thresholds to distinguishing good from poor performance on each metric. Given the intricate nature of how errors can rapidly propagate throughout many aspects of mission performance, independent error propagation trees were developed for each of the seven friction points, including outcome metrics associated with each. For each of these, cause-effect linkages between relevant metrics were developed, allowing for identification of how each potential error could propagate through to effect mission outcome. Flow diagrams illustrating the potential error propagations were developed to facilitate validation with instructors. Prior to development of final system requirements, metrics, metric thresholds and error propagation (i.e., metric cause-effect relationships) logic were validated over a series of meetings involving a walkthrough of metric logic flow diagrams with TTECG instructors. Requirements for the internal workings of PAST were then developed including detailed conceptual algorithms for calculation of metrics, including from where input parameters would be pulled (i.e., simulation data vs. event-based checklist), metric thresholds and error propagation logic associated with each friction point.

PAST TESTING AND EVALUATION

Following its initial development phase PAST was run through a series of evaluations testing its functionality and usability as a tool intended to effectively diagnose FiST team performance. These were of two types: a summative evaluation to validate the accuracy of the metrics and logic, and a formative evaluation focusing primarily on usability and user acceptance collecting feedback from SME's to drive future redesign.

PAST Metric and Metric Logic Testing

To evaluate PAST's functionality, all metrics and metric logic were tested by running the tool during scenario performance, generating output data, and verifying the actual output data with previously contrived output data (i.e., experimenter generated output data based on metric logic during the conceptualization phase of PAST development). The goals of metric and logic testing were to 1) confirm the feasibility of PAST creating output data, and 2) validate whether the output data followed the metric logic as expected.

FiST mission scenarios were developed and run on the MOT^{2}IVE training system. Team performance on three of the scenarios was simulated by experimenters acting as FiST members. Additionally, a DIVAARS operator tagged and recorded mission data during the simulated performance (i.e., using TEAM-DOC). At the end of the scenario, the recorded data was compiled, and PAST was able to output the Friction Point Summary, Multi-Level Timeline, and Performance Summary (Goal 1). The Friction Point Summary was further scrutinized by comparing all of the highlighted metrics (i.e., root errors, process errors, and outcome errors) with previously constructed Friction Point Summary data based on the metric logic for that scenario. Root error metrics were cross checked between the actual PAST output and the expected output to determine that PAST was outputting the Friction Point Summary data as it was intended (Goal 2). A match between the two data sets suggested the metrics and logic were accurate.

PAST Usability Evaluations and Redesign

Still in its initial development phase, PAST usability data was collected to help facilitate further redesigns of the tool. PAST was presented to TTECG instructors, Infantry Officers Course (IOC) instructors and USMC Training and Education Command (TECOM) SMEs. IOC and TTECG instructors viewed video recordings of FIST team performance in MOT^{2}IVE scenarios and were then presented PAST output resulting from the scenario. The instructors were asked to give feedback on PAST's utility to novice instructors to diagnose team performance and to provide team debriefs. Both groups gave positive feedback about PAST's potential to capture the necessary mission data to diagnose team performance, but expressed concern about the overwhelming presentation of too much information on its output interface. Based on these comments, redesign was performed and mock ups were developed including two PAST interface redesigns (see Figure 5): 1), Simplified Friction Point Summary (FPS) and 2) Debriefing Points.

The Simplified FPS addresses both IOC and TTECG instructor concerns that the original FPS interface was cluttered with too much information. Taking into account their feedback, a simplified FPS was created only highlighting root errors (dark green boxes) under debriefing points (yellow boxes), classified by Friction Points. The simplified friction point summary is illustrated in Figure 5.

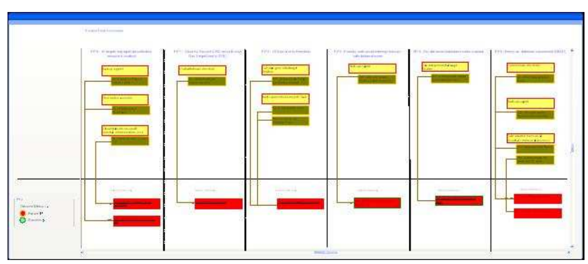


Figure 5. Simplified Friction Point Summary Interface

The Debriefing Points interface specifically addresses the TTECG instructors' suggestion to explicitly organize metric information under their instructor debriefing points (e.g., root error "SEAD Requests Not Made" falls under debrief point "Communicate effectively"). The Debriefing Points interface is illustrated in Figure 6.

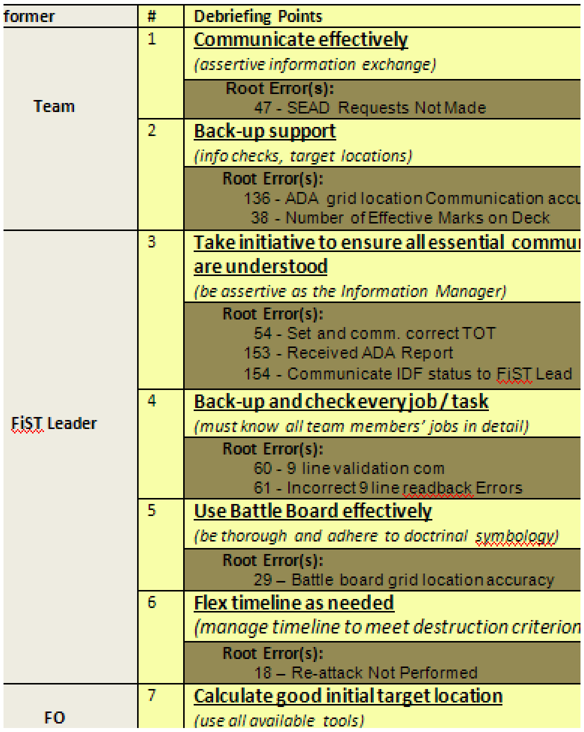


Figure 6. Section of Debriefing Points GUI

Initial responses to the mock-ups redesign interfaces were positive and suggested subsequent redesigns move in the direction of presenting PAST information in a more simplified and organized manner. TECOM was presented with both the initial and redesigned PAST interfaces and asked to comment on the general direction of interface redesign, preferred the redesigned views of the interfaces, but suggested that both the event tagging (i.e., TEAM-DOC) and the outputs could be further simplified by reducing the tasking involved with inputting tagging information (TEAM-DOC) and with extracting the pertinent team performance data (FPS).

FUTURE OF TOOL

As VIRTE has drawn to a close, the recommendations resulting from PAST evaluations have not been fully realized into a working prototype. These recommendations, however, have and will shape future performance assessment tool development under ONR's Human Performance, Training and Education (HPT&E) and Capable Manpower programs, including Multi-Axis Performance Interpretation Tool (MAPIT) design under the Observational Skills Enhancement and Retainment Virtual Environment (ObSERVE) effort and the Instructor Support Station (ISS) design under the Next generation Expeditionary Warfare Intelligent Training (NEW-IT) effort. The paragraphs below discuss specifically how the PAST concept and functionalities are being leveraged in these two efforts.

A future goal is for PAST to track trends in these over trials and groups of trainees, allowing for pattern detection and identification of global trends.

MAPIT

The Multi-Axis Performance Interpretation Tool (MAPIT) has taken the root cause analysis methodology and applied it to perceptual skills performance assessment. Aiming for simplified metric propagation logic to both facilitate reduced development time and simplified output displays, MAPIT uses the root cause analysis methodology along with behavioral and physiological metrics to detect where in the perceptual process (i.e., attention, sensation, perception) a breakdown occurs to provide instructors with granular performance diagnosis data to facilitate tailored training feedback. MAPIT is currently under development and further findings will be reported as they become available.

NEW-IT

The Next generation Expeditionary Warfare Intelligent Training (NEW-IT) Instructor Support Station (ISS) is being designed to provide the training community with a generalizable (i.e., reusable, reconfigurable) instructor operator/training management system to be used across a range of Distributed Operations (DO) domains. PAST provides a useful foundation for the development of this tool. Identified limitations of PAST such as rigidity of intended use (i.e., designed for specific mission with specific team members and performance conditions) and the perceived complexity of input and output components by instructors as well as instructor input regarding usefulness of the different functionalities to current training practices are invaluable inputs to NEW-IT ISS requirements.

CONCLUSIONS

The use of intelligent diagnosis tools to assist instructors in understanding true failures appears to be feasible and of value, although not without challenges. Throughout the design of PAST two main challenges were confronted. First, while the design of interfaces to accomplish this goal may be accomplished through iterative designs and testing, their general application to different domains may prove more challenges. Nonetheless, certain aspects of such designs do appear to be readily transferable, such as those based on visualization principles (e.g., propagation diagrams). Secondly, the tradeoff between diagnosticity and transferability (i.e., the more specific and diagnostic a system is, the more domain specific it becomes such that its transferability to other domains is hindered and vice versa.) These are challenges that are being faced as the concept is explored further and implemented in different projects.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Office of Naval Research (ONR) and its Virtual Technologies and Environment (VIRTE) program, and the instructors at the Marine's Tactical Training Exercise Control Group (TTECG) who made this work possible. In addition, we would like to thank Dustin Chertoff, Jason Smith and Casey Thurston for helping in the development of PAST.

REFERENCES

Benner, J. (1975). Accident Investigations: Multilinear Events Sequencing Methods. Journal of Safety Research, Vol. 7, No. 2

Borgonovo, E., Smith, C.L., Apostolakis, G.E. (2000). Insights from using influence diagrams to analyze precursor events. In Proceedings of the International Conference on Probabilistic Safety Assessment and Management. November 27-December 1. Osaka, Japan.

Card, S.K., Pirolli, P., Mackinlay, J.D., (1999). Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann Publishers, San Francisco, CA.

Concoran, W.R., Swanson, R.N., Evans, R.W, Cooley, D.E. (1997) The comparative TimeLine highlighting human performance factors. In proceedings of the IEEE Sixth Annual Human Factors Meeting. Orlando Florida. Pp. 13-18.

Chung, W., Chen, H., Chaboya, L., O'Toole, C., Atabakhsh, H., (2004). Evaluating even visualization: a usability study of COPLINK spatio-temporal visualizer. International Journal of Human-ComputerStudies, 62(1), pp. 127-157.

Hoover, A. & Muth, E. (2004). A real-time index of vagal activity. Journal of Human Computer Interaction, 17 (2), 197-209.

Potter, S.S. and Woods, D.D. (1991) Event Driven Timeline Displays: Beyond Message Lists in Human-Intelligent System Interaction. In proceedings of the : Systems, Man, and Cybernetics, 1991. 'Decision Aiding for Complex Systems Conference. October 13-16. Pp. 1283-1288 vol. 12.

Tufte, E.R., (2001). The Visual Display of Quantitative Information, Second Edition. Graphics Press, Cheshire, CT.

Schneiderman, B. (1996). Inventing discovery tools: Combining information visualization with data mining. Information Visualization, 1(1), pp. 5-12.