Developing Models of Expert Performance for Support in an Adaptive Marksmanship Trainer

Benjamin Goldberg, Charles Amburn, Keith Brawner
U.S. Army Research Laboratory—Human Research and Engineering Directorate
Orlando, FL
benjamin.s.goldberg@us.army.mil, charles.amburn@us.army.mil, keith.w.brawner@us.army.mil

Marko Westphal
German Federal Office of Bundeswehr Equipment—Information Technology and In-Service Support
Koblenz, Germany
marko.westphal@gmail.com

ABSTRACT

The U.S. Army's Engagement Skills Trainer (EST) uses sensors on simulated weapons to collect valuable data about a soldier's performance during marksmanship exercises. That data is available to an instructor for coaching and remediation purposes. However, experience shows that accessing the data, reviewing the data, and providing feedback to a trainee can be a time consuming process. This environment presents challenges when considering the number of trainees who must complete this training and the limited number of instructors available. This also assumes that instructors are capable of accurately interpreting the data and applying effective remediation. Simulators like the EST are prime candidates for the incorporation of an Intelligent Tutoring System's (ITS) capabilities. The goals of an ITS are to collect data from a system, make inference on that data as it relates to defined metrics, and to provide formative feedback when data is found to deviate from a specified standard. For this purpose, a system requires models to compare data against. In this paper, we will present the results of the first phase of a study to apply ITS technology to the fundamentals of marksmanship. Models created in this phase will be integrated into an adaptive training system prototype built within the Generalized Intelligent Framework for Tutoring (GIFT) for future experimentation. Data was collected across eight experts from the U.S. Army Marksmanship Unit's service rifle team as they conducted marksmanship tasks. These models are built around sensor data collected during execution, with each sensor being selected based on their link to the fundamentals of marksmanship. We will review the techniques applied to the data for model construction, trends found in the data that are generalized across each expert, and how the models will be used to diagnose error and trigger remediation.

ABOUT THE AUTHORS

Dr. Benjamin Goldberg is a member of the Learning in Intelligent Tutoring Environments (LITE) Lab at the U.S. Army Research Laboratory's (ARL), Human Research and Engineering Directorate (HRED), Simulation and Training Technology Center (STTC) in Orlando, FL. He has been conducting research in the Modeling & Simulation community for the past five years with a focus on adaptive learning and how to leverage Artificial Intelligence tools and methods for adaptive computer-based instruction. Currently, he is the LITE Lab's lead scientist on instructional strategy research within adaptive training environ-ments. Dr. Goldberg is a Ph.D. graduate from the University of Central Florida in the program of Model-ing & Simulation. Prior to employment with ARL, he held a Graduate Research Assistant position for two years in the Applied Cognition and Training in Immersive Virtual Environments (ACTIVE) Lab at the Institute for Simulation and Training. Dr. Goldberg's work has been published across several well-known conferences, with recent contributions to the Human Factors and Ergonomics Society (HFES), Artificial Intelligence in Education and Intelligent Tutoring Systems (ITS) proceedings, and to the *Journal of Cognitive Technology*.

Charles Amburn is the Senior Instructional Systems Specialist for the Advanced Simulation Branch of the US Army Research Laboratory, Human Research and Engineering Directorate, Simulation Training and Technology Center (STTC) in Orlando, FL. After obtaining both a Film degree and a Master's degree in Instructional Systems Design from the University of Central Florida, he began his DoD civilian career in the Advanced Instructional Systems Branch at the Naval Air Warfare Center Training Systems Division (NAWCTSD). There, he worked on special projects for the Navy and Marine Corps for 10 years before becoming the Lead Instructional Designer for the Army's Engagement Skills Trainer (EST) program at the Program Executive Office of Simulation, Training and Instrumentation (PEOSTRI), Orlando, Fl.

Dr. Keith Brawner is a researcher for the Learning in Intelligent Tutoring Environments (LITE) Lab within the U.S. Army Research Laboratory’s Human Research & Engineering Directorate (ARL-HRED). He has 8 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. The foci of his current research is in machine learning, active and semi-supervised learning, realtime datastream processing, affective computing, adaptive training, semi/fully automated user tools for adaptive training content.

Marko Westphal is a German Exchange Engineer for the Engineer and Scientists Exchange Program (ESEP) and member of the Learning in Intelligent Tutoring Environments (LITE) Lab at the U.S. Army Research Laboratory’s (ARL), Human Research and Engineering Directorate (HRED), Simulation and Training Technology Center (STTC) in Orlando, FL. He has 5 years of experience within the German Federal Office of Bundeswehr Equipment, Information Technology and In-Service Support in the field of acquisition and development of computer-based training software. He holds a Master Degree in Computer Engineering with a focus on Software Development.

Developing Models of Expert Performance for Support in an Adaptive Marksmanship Trainer

Benjamin Goldberg, Charles Amburn, Keith Brawner
U.S. Army Research Laboratory—Human Research and Engineering Directorate
Orlando, FL
benjamin.s.goldberg@us.army.mil, charles.amburn@us.army.mil, keith.w.brawner@us.army.mil

Marko Westphal
German Federal Office of Bundeswehr Equipment—Information Technology and In-Service Support
Koblenz, Germany
marko.westphal@gmail.com

INTRODUCTION

Marksmanship is a complex psychomotor skill demanding high physical and mental coordination. To strike a small target, possibly hundreds of yards away, a soldier must have a combination of proper breath control and body positioning, fine motor control of muzzle wobble and trigger squeeze, and sight alignment between the rear sight, front sight and target (Pojman et al., 2009). In the U.S. Army, all soldiers begin with Basic Rifle Marksmanship (BRM) training that focuses on the fundamentals of marksmanship, grouping and zeroing techniques with an M16-/M4-series rifle, and transitional firing that culminates with a qualification event. Traditionally, this training includes education in the classroom, practice on the Engagement Skills Trainer (EST) marksmanship simulator, and live fire on outdoor ranges. The EST is a simulator developed for initial exposure to the interaction components of BRM and supports all facets of firearms instruction, including weapon qualifications associated with basic training.

The purpose of this project is to identify tools and methods to improve the diagnostic capability of the EST to support better training practices. Currently, much of the instruction is based on subjective assessment of performance outcomes. For instance, in a live training event on an outdoor firing range, an instructor reviews a target after a shooting event and applies his or her expertise on what may have caused errors in shot placement. Unfortunately, due to the high number of soldiers who must complete this training every year, the Army does not have the resources to provide a hands-on coach for every individual during every firing event. While an instructor is able to provide one-on-one feedback to a Soldier who is failing to perform to standard, he or she often does not have all of the data needed to immediately and effectively diagnose the cause of the errors. Furthermore, important behaviors involved in the execution of a shot (i.e., breathing, fine motor control during trigger squeeze, and sight alignment) are difficult to observe without using a system of sensors to collect data.

On the EST an instructor replicates this process, but has additional information at his or her disposal including aim trace, trigger pressure and cant of the weapon (before, during and after the shot), which is provided through data logged from sensors embedded directly on the weapon. The instructor can access and use this information to provide focused feedback to the group or individual in an attempt to improve performance. However, proper use of these sensor technologies on a simulator like the EST is no guarantee. Instructors have to take time to review the data, interpret the data, and then to provide individual recommendations to struggling shooters. Given the vast number of trainees who must complete this training and the limited number of instructors available, this environment presents challenges. The above instructor-in-the-loop scenario additionally assumes that the instructors are capable of accurately interpreting the data and applying remediation in a consistent fashion. The reality is that some instructors misunderstand parts of rifle marksmanship doctrine and/or inconsistently apply training techniques and procedures (James & Dyer, 2011). This can become additionally problematic when training is conducted using a complex simulator like the EST because of factors such as: (a) a lack of instructor familiarity with the system's features and capabilities, and (b) a lack of guidance with regards to best-practices for conducting training using the simulator.

One method of ensuring reliable, individualized, targeted diagnosis and instructional feedback is through the use of Intelligent Tutoring System (ITS) technologies. ITSs have already proven to be an effective and reliable tool for improving learning by personalizing instruction around the knowledge, skills, and abilities of a given learner, especially within well-defined domains of instruction like math and physics (e.g., VanLehn, 2011, Stottler, Harmon, &

Michalak, 2001, Sherry, Feary, Polson, & Palmer, 2000). What is not currently seen in the field is the application of these technologies to support psychomotor related domains, such as marksmanship. The psychomotor ITS training research presented in this work is the first of its kind. Conceptually, an ITS integrated into the EST could receive marksmanship behavioral and performance data produced by a trainee and instantly provide the instructor and/or student with feedback intended to correct performance. This could enhance the system's ability to provide effective training by eliminating the variability in accuracy and consistency of instructor error diagnoses. Further, by being able to address every trainee's problems immediately, an ITS could help the Army improve the efficiency and cost effectiveness of marksmanship training.

In related work, Chung, Nagashima, Espinosa, Berka, & Baker (2009) found that individualized instruction based on errors related to shot placement, body position, breathing, trigger squeeze and muzzle wobble was effective in improving shooting skills. In order to measure breathing and trigger pressure, a sensor-embedded weapon simulator was used like those on the EST. However, in that study, a human coach was used to review the data and then diagnose problems. Whereas in an ideal situation, an ITS would review the data (faster than a human could) and diagnose the problems (more consistently than a human could). This assumes, of course, that an ITS has the capability to analyze the shooter data and determine how it compares to the ideal performance. Thus the goal of this study is to collect performance data from expert shooters and use that data to create an expert model for use in an ITS.

Research Objective

The focus of this controlled laboratory study is to measure expert performance in the context of BRM. The overall research goal is to build models of expert performance on the fundamentals of marksmanship for inclusion in an adaptive marksmanship training system embedded with intelligent tutoring functionality. Models were built on sensor technologies that monitor variables linked to: (1) trigger pressure, (2) cant/angle of the weapon's orientation, and (3) breath control. Aim trace data was also collected, however we have not yet completed the analysis of this variable; as such, it will not be included in the results. The data was collected across a sample of expert marksmen, which was ultimately used to examine the efficacy of developing a generalized expert model based on trends and correlations present in data outputs. Resulting models will be transitioned into ARL's Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg & Holden, 2012) for the purpose of conducting follow-on experimentation, and assessing their use in an operational context.

Hypotheses

Sensor data collected during interaction with a simulated marksmanship training system will correlate across designated experts when performance is deemed excellent.

a. Prediction 1.1: A metric for application of trigger pressure, as gauged from the mounted sensor, will correlate across expert participants, with general trends observed individually
b. Prediction 1.2: Breathing patterns, as monitored by a breathing strap sensor, will correlate across expert participants
c. Prediction 1.4: The cant/angle of a weapon's orientation, as monitored by a mounted sensor, will correlate across expert participants.

METHODOLOGY

Apparatus

The experimental testbed is an integrated system of software and hardware components (see Figure 1). The software components are the SCATT marksmanship training software (SCATT Shooter Training Systems, 2013) and the GIFT (Sottilare et. al, 2012). The hardware components are an airsoft M4 carbine rifle that produces realistic recoil, a pressure sensitive trigger, a Weapon Orientation Module (WOM), and a Zephyr physiological monitor.

SCATT Shooter Training System

SCATT WS1 is used for fixed target marksmanship training at distances from 5 to 10 meters and imitates shooting up to 1000 meters away. The system operates through an electronic optical sensor that is fixed to the barrel of a weapon. The shooter aims at the SCATT electronic target. The system logs the point of aim in real-time as measured through the optical sensor. When the weapon's trigger is activated, the point of impact is recorded. All data can be logged accordingly for post-hoc analysis.

Hatalom Electronic Air Recoil Trainer (HEAT) Rifle
The HEAT is a simulated assault rifle weapon that has the look and feel of the real U.S. Army M4 (see Figure 2). Using a standard, commercially available 12 gram carbon dioxide (CO2) cartridge, the HEAT is able to produce a relative approximation of the noise and feel of weapon recoil without firing any projectiles.

Figure 1. The experimental system: Pictured from left to right are the target, the simulated weapon, a laptop running the marksmanship program, and the physiological sensor system.

Electronic sensors were attached to the HEAT in order to collect data about weapon orientation, trigger pressure, and aiming. One of these is an optic laser mounted near the muzzle of the weapon and pointed downrange at the target. This laser is used by the SCATT Shooter Training System (described above) to determine muzzle position (and thus, accuracy and muzzle wobble).

SCATT Trigger Sensor
The HEAT rifle was modified for the inclusion of the SCATT STS Trigger Sensor, which is used to record an individual's application of pressure on the trigger over the course of a shot. This includes the effort applied by the shooter leading up to the fire event as well as following the release of the shot. This sensor replaced the standard HEAT rifle trigger.

Inertial Labs OS3D Weapon Orientation Sensor
In addition to the SCATT laser emitter and trigger sensor being fitted directly on the HEAT rifle, an orientation sensor developed by Inertial Labs was also installed. This sensor, called the OS3D, combines sensor technologies of magetometers, gyroscopes, and acceleratometers to measure absolute orientation of an object (inertaillabs.com). The OS3D was mounted directly atop the barrel of the rifle enabling the collection of metrics that convey pitch, roll, and yaw (PRY). This information associates with the position of the weapon while an individual produces shots on the SCATT target and will be used to determine how experts and novices hold the weapon during task execution.

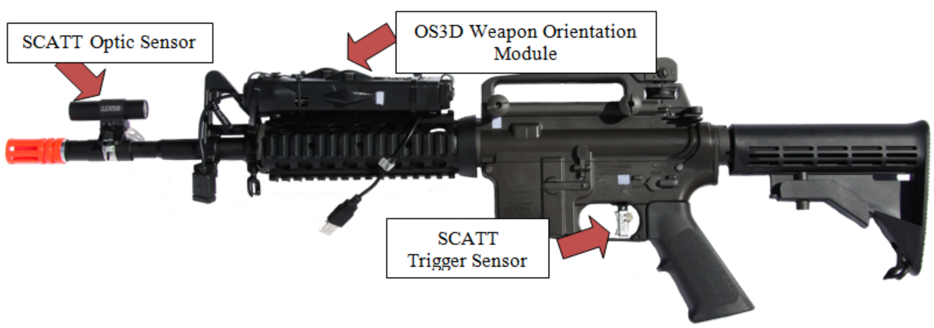

Figure 2. Simulated M4 Rifle with Mounted Sensors

Zephyr Technology BioHarness BT
The BioHarness BT is a compact electronics module that attaches to a lightweight fabric strap which incorporates Electrocardiogram and breathing detection sensors. The sensor wears like a heart rate monitor avid runners use when exercising and is completely wireless. For this experiment, the data of interest are the breathing metrics the device provides. This information will be used to monitor a participant's breathing pattern while executing groups of shots.

The Generalized Intelligent Framework for Tutoring (GIFT)
GIFT is an evolving architecture-based project intended to provide the tools and methods for authoring and delivering adaptive training in a variety of instructional domains (Sottilare et al., 2012). This generalized approach enables system developers to quickly construct intelligent tutoring capabilities through a set of standardized tools and messaging schemas. For this study, GIFT provided the architecture required to create a unified marksmanship system

from which data was collected and logged while subjects interacted with the SCATT training environment. Modules in GIFT were authored to enable control of the SCATT software along with the collection of sensor data being logged across multiple pieces of hardware. During the second phase of this study, the same modules will be modified for data assessment and feedback provisioning.

Participants

The expert marksmen for this study were recruited from the U.S. Army Marksmanship Unit's (AMU) Service Rifle Team out of Fort Benning, GA. Based on discussions with a leading researcher in the field of expert model development at the University of Central Florida, Dr. Avelino Gonzalez, the target sample for this study was eight. This number allows for exploratory analyses to determine if a generalized model of expert performance can be constructed that takes into account all subject data. In the event that no correlation was observed across experts, the use of a single expert (with best performance) would be the path forward. Of the eight experts recruited, seven were male and one was female, with an average age of 28 across all participants. Several members were repeat national champions in service rifle competitions, and majority were recognized as being one of the top one-hundred shooters in the country, having received the President's Hundred Tab. Subjects in the experiment were volunteers and received no compensation for participation. Superiors had no influence over subordinates in determining their participation nor were they present during data collection. All involved in the experiment will remain anonymous.

Procedure

Upon arrival, participants received a brief overview of the study and were asked to fill out an informed consent. Next, subjects were fitted with the BioHarness breathing strap, and the sensors were then synced to GIFT for time-synced data logging. This process took approximately 10-15 minutes. Participants then filled out a Demographics Questionnaire covering experience in the Army Marksmanship Unit and the various accolades and awards they won throughout their career.

Following completion of the surveys, participants were presented a short PowerPoint slideshow that reviewed the purpose of the study and the details of the experimental equipment. This included information on the SCATT Shooter Training System, the HEAT rifle they would interface with, and all the associated sensors that were logging data during task execution.

Next, participants were given the opportunity for familiarization training with the system. They were instructed on task procedures, and then practiced with the weapon for 5 minutes. Upon completion of the practice phase, subjects were prepped for the main data collection window where they were instructed to produce shot groupings consisting of five rounds, with the goal to get all shots as close to center of target as possible. There was no set limit for the number of shot groupings to produce. All subjects were given 18 minutes, across two BRM stances (prone and kneeling), to produce as many groupings as they could, with self-regulated breaks taken in between. Upon completion of data collection with the SCATT Shooter Training Systems, participants completed a post-experiment survey. This was followed by a short debrief where the subject was given the opportunity to ask questions to the experimental proctor and to discuss next phases of the research.

Independent Measures

The Independent Variable (IV) for this study is the associated interaction mode with the SCATT trainer. The interaction modes consist of two unsupported firing stances: (1) prone and (2) kneeling. As the goal of the study is to build models of expert performance, the IVs will distinguish marksmanship behaviors to assess if differences are observable across the two firing conditions. To support model development, correlations and trends will be examined with relation to the collected Dependent Variables and how they measure up with performance for the associated interaction mode.

Dependent Measures

Dependent measures are associated with two distinct categories. There are measures linked to performance outcomes that dictate the quality of a firing event, and there are measures linked to operator behaviors that occur during the execution of a shot (see Figure 1). Behavior metrics are used to observe how an expert functions during a firing event and to determine if experts consistently do the same things across events. The measures were selected based on their relationship to the recognized four fundamentals of BRM (Army, 2008). Table 1, below, describes the measure being collected and its relationship in the process of expert model development.

Table 1. Description of Associated Dependent Measures

Category	Measure Type	Description	Data Source	Expert Model
Performance	Shot Grouping Distance	• Measure computed by averaging the distance of each shot to the designated center point of each 5-shot cluster • Used to gauge consistency across shots	SCATT Shot Placement Values	n/a (direct)
	Shot Grouping Quality	• Measure computed by averaging distance of each shot to the center of target • Used to gauge accuracy across shots	SCATT Shot Placement Values	Sight Picture (direct)
Behavior	Respiration Rate	• Provides real-time monitoring of respiration patterns during execution of a shot	BioHarness BT Breathing Strap	Breath Control
	Weapon's Position	• Provides real-time pitch, cant, and yaw measures used to monitor weapon position during execution of a shot	OS3D Weapon Orientation Module	Body Position
	Aim Trace	• Provides real-time capture of optical sensor reading in relation to the SCATT target. Used to monitor stability and control of the weapon leading up to the execution of a single shot.	SCATT Optical Sensor	Incomplete (analysis in progress)
	Trigger Pressure	• Provides real-time monitoring of the pressure applied to the rifle's trigger during the execution of a shot	SCATT Trigger Sensor	Trigger Pressure

Data Analysis

The analysis has been conducted in the following phases:

Phase I - Introductory exploration for a match between experts. The basic research question that this analysis seeks to answer is "Are there commonalities for marksmanship performance observations among experts?" Data will be normalized, graphed, and statistically profiled for similarity measures. In the event that there are no significant correlations among experts, then the expert which is most "representative" of the expert group will be chosen for Phase II. Generally, a "n=1" sample is not statistically significant, but this method of analysis has been used in other expert analysis studies (if the person is a 'representative expert', then instruction towards this standard is logically sound).

Phase II - Construct statistical profiles of expert(s) data, and establish boundary conditions which seek to identify an event which would not occur among experts. Appropriate thresholds for error identification will be constructed via standard deviation metrics (ie. "you are doing something dissimilar from an expert marksman"). These standard deviation metrics will be tested in a future pilot study. Leave-one-out cross-fold validation is used to validate the resulting models.

RESULTS

Qualitative Within-Subject Analysis

The within-subject analysis generally comes to the conclusion that the manner in which experts fire across shots is internally consistent. As an example of this internal consistency, Figures 3a, 3b, and 3c show trigger pull, breath control, and weapon position data across 30 individual shots taken from one expert. Qualitative analysis reveals general trends across the data sets that represent consistent behavior during the execution of a shot. The middle point in each of the below figures represents the period after the trigger has been pulled. Similar trends are observed across all other experts.

For each of the experts, there is a qualitative analysis where similar behavior to the above is observed across all shots. Trending information was noted for each expert during this analysis in an attempt to answer the question "Does the expert exhibit similar behaviors for each shot taken"? The observed trends can be simply described and appear, on the surface, to be consistent with the fundamentals of BRM (Army, 2008). This validates the approach in that the correct items are being measured by the system. In brief, these items are:

- For trigger squeezing data, pressure applied gradually increases up until a shot is taken
  - Observed in data through gradual, mostly linear, slope in the trigger pull data
- Breathes are held during the time period prior to shots and trigger squeezes, and show little variability
  - Observed in data through breathing data coalescing to a single value across all shots
- Pitch, Roll, and Yaw (PRY) values stabilize in a time period prior to the shot, showing little variability
  - Observed in data through variability of the PRY values before and after the shot

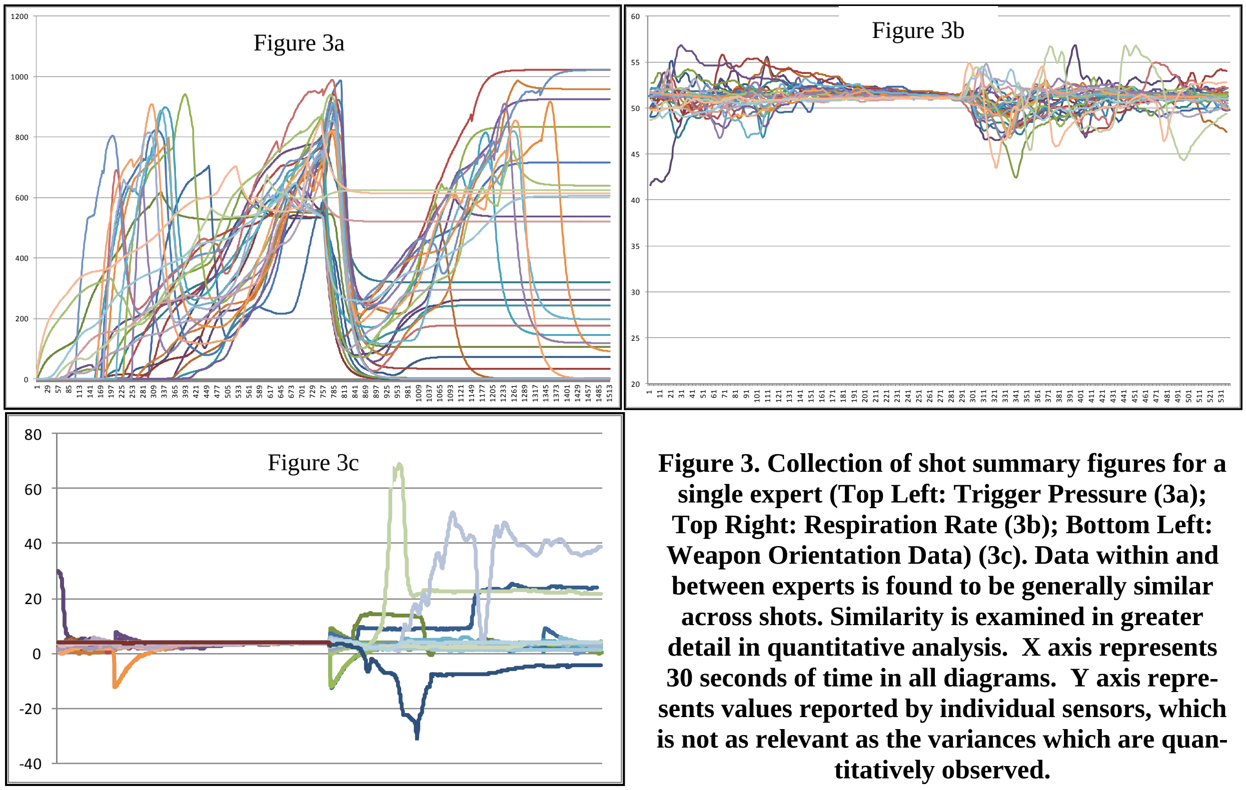

Figure 3. Collection of shot summary figures for a single expert (Top Left: Trigger Pressure (3a); Top Right: Respiration Rate (3b); Bottom Left: Weapon Orientation Data) (3c). Data within and between experts is found to be generally similar across shots. Similarity is examined in greater detail in quantitative analysis. X axis represents 30 seconds of time in all diagrams. Y axis represents values reported by individual sensors, which is not as relevant as the variances which are quantitatively observed.

Of true interest, however, is the creation of a general model of marksmanship fundamentals. The answer to the question of whether marksman fired in a manner which is similar to the other experts was among the things that were investigated, beginning with the following segment.

Model Development Considerations

This data represents the recording of superior performers in their environment of expertise. This results in, expert shot performance that is vastly superior to typical performance by novice shooters. Models created for this effort are intended to be *descriptive* in the behaviors associated with expert performance, with an intention for them to provide *prescriptive* capabilities within an ITS. These models will be used in order to diagnose misaligned (non-expert) novice behavior, rather than to catch novice errors, although these are similar tasks. There is no guarantee, however, that these models will catch the most common novice mistakes. In order to create novice-prescriptive models, expert-annotated novice performance data is required. The collection of such data is left to a second phase of this study, wherein novices will be

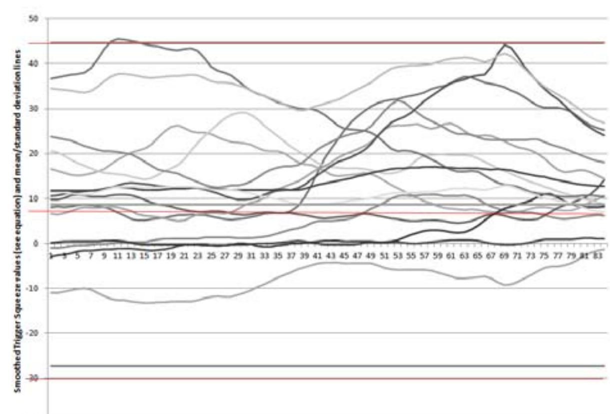

Figure 4. A sample model for one expert, applied to himself, across 11 shots. Red lines indicate mean and variance boundaries. X-axis is time, Y-axis is derived values (see below).

assessed via the descriptive expert models described in the following subsection. This is discussed more thoroughly in the future work section.

In terms of implementation techniques, there are an infinite number of possible models, with little way to evaluate relative superiority aside from comparison to other experts. As an example, consider a measure of average trigger slope in a time window of 4 seconds prior to shot. Why not use 6 seconds? 8 seconds? 0.5 seconds? An infinite number of windows condensed into an average shot profile? A time normalized shot sequence starting from first trigger touch to final shot? In the interest of making models simple, understandable, and transferable, the models presented below are aligned with marksmanship fundamentals and observed trends. They are standard deviation (SD) based, profile-based, and applied to the other experts with measured data. Such an approach has face and numerical validity in the description of expert behavior, but may be able to be bypassed by novices. This type of model adjustment is discussed in future work. An example of such an assessment is shown in Figure 4 with SDs serving as error thresholds.

Quantitative Between-Subject Analysis

For the between subjects analysis, the hypothesis is that experts perform similarly to each other. While this is qualitatively verified visually via graphs of various expert variables, it must be implemented mathematically in order to serve as an implementation inside of a psychomotor ITS. All measurements were investigated in series in an effort to create such a model.

Trigger Squeeze

The variable of trigger squeeze, as seen from the above graphs can be described verbally as a “slow squeeze preceding the shot”. This can be described mathematically as “the slope of the smoothed line in the time period prior to the shot is positive”. This smoothing was performed to remove the noise present as a byproduct of data collection (Cooper, 1998), and the above figure is shown post-smoothing. Given that the model is intended to evaluate whether a novice falls outside of two standard deviations of the expert model, this value is examined, and described in Equation (1).

$$y[n] = \frac{1}{\tau}x(n) + \frac{\tau-1}{\tau}y(n-1), \qquad \text{window value (tau) of 0.5 seconds was chosen empirically.} \qquad (1)$$

The discreet time derivative measure was also applied in order to slope over time. This is described in Equation (2).

$$\left(\frac{\mathrm{d}}{\mathrm{dx}} = y(nT) = \frac{1}{8} * T[-x(nT-2T) - 2x(nT-T) + 2x(nT+T) + x(nT+2T)]\right) \qquad (2)$$

The model produced by this method is that expert shooters in the prone position have an average smoothed derivative measure of 8.3, with a 3.9 standard deviation, in the shooting range of 10 seconds prior to shot, while kneeling shooters have values of 6.79 / 2.185 in the range of 13 seconds prior to shot. Generally, experts pull the trigger slower for a longer period of time when kneeling. These values were validated through the use of leave–one-out-cross-validation, where a model made from 7 experts was tested against the remaining expert. No expert fell outside of two standard deviations of the values produced from the other marksman in this condition, leading to an experimental confirmation of the hypothesis that experts fire in a manner which is similar to the other experts, and which is logically consistent with the field manual and personal feedback received by experts. In use as a prescriptive model, this indicates that none of the experts would receive feedback for discrepancy on trigger pull fundamentals.

Breathing

The variable of breathing, as seen from the above graphs can be described verbally as a “holding a breath prior to a shot”. This can be described mathematically as having “little absolute value change of breathing variable in the time prior to the shot”. The measure which was selected to create the expert model is the standard deviation, which measures the square root of the variance. The general expert trend lends itself to this type of analysis.

The model produced by this method is that expert shooters in the prone position have an average standard deviation measure of 0.11, with a 0.10 standard deviation, in the shooting range of 1.6 seconds prior to shot, while kneeling shooters have values of 0.15 / 0.13 in the range of 2.21 seconds prior to shot. Generally, experts hold their breath longer in the kneeling position and exhibit more variability. Breathing values, as reported by the BioHarness system, are generally between 40 and 65, but are dependent on chest size, which standard deviation values serves to remove. These values were validated in the same manner as trigger squeeze. In using these two measurement models, only 2 experts made mistakes related to this item twice inside of a five shot grouping (when feedback would have been assigned as part of an ITS). The vast majority of the time (90 5-shot groupings) experts breathe slowly prior to shooting, which may result in a 2% total error rate of the model, and was deemed acceptable for model creation. This generally implies experimental confirmation of the hypothesis that experts fire in a manner that is similar to the other experts and which is logically consistent with the manual, training, and personal feedback received by experts.

Weapon Orientation

Table 2. Descriptive Model Statistics for Weapon Orientation

	P SD mean	P SD SD	R SD mean	R SD SD	Y SD mean	Y SD SD
Prone	0.0700	0.0543	0.1140	0.0919	0.0717	0.1077
Kneeling	0.0939	0.0655	0.1449	0.0869	0.1161	0.218

Weapon orientation was measured with Interial Labs' OS3D sensor. The variable of orientation, as seen from the above graphs can be described verbally as a "stabilizing the weapon prior to a shot". This can be described mathematically as having "little absolute value change of orientation variable in the time prior to the shot". The measure which was selected to create the expert model is the standard deviation in the same manner as was performed for the breathing measurement. The general expert trend lends itself to this type of analysis. The model produced by this method is described in table 2, with a defined range of 5.2 seconds prior to shot for both prone and kneeling. Generally, although aligned in time, shooters expressed more variability in the kneeling position. Orientation values, as reported by the OS3D, are absolute values between 0 and 360 degrees across the three dimensions of PRY. Using these measurement models, only 1 expert made mistakes related to this item twice inside of a five shot grouping in the prone condition. The same behavior was observed for another, different, expert in the kneeling condition. This expert generated models significantly different than the other experts for all shot groupings, which generally represents doing something different and is cause for future analysis. This expert was the third worst in terms of performance, which gives confidence that these models will be usable for a general population. Inclusion of the abnormal expert indicates a 15% total error rate, while exclusion of the dissimilar expert represents a 0% total error rate in the prone condition, while the kneeling condition indicates a 15%/1% error rate.

Cross-Validation Summary

Table 3. Leave-one-out Cross Validation Outcomes

KNEELING									PRONE							
Leave Out Expert	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Trigger Pressure																
Avg	6.5206	6.4420	6.5115	6.8753	7.2269	7.2238	6.8208	6.7053	8.2369	7.8943	8.0533	9.5212	8.0377	8.4645	8.3794	7.8640
Stddev	2.2394	2.2501	2.3957	2.0022	1.8520	2.1366	2.4487	2.1575	3.8445	4.0035	3.8347	3.7677	3.9862	3.9832	4.0045	3.9797
Breathing																
avg stddev	0.1549	0.1537	0.1276	0.1456	0.1570	0.1548	0.1357	0.1317	0.1184	0.1263	0.1090	0.1124	0.1209	0.1243	0.1077	0.0997
std stddev	0.1391	0.1472	0.1310	0.1301	0.1491	0.1448	0.1237	0.0953	0.1011	0.1066	0.0976	0.0988	0.1067	0.1050	0.0731	0.0760
PRY measures																
P avg stddev	0.0909	0.1008	0.0967	0.0818	0.0877	0.0980	0.1008	0.0943	0.0677	0.0763	0.0688	0.0556	0.0738	0.0712	0.0741	0.0726
P std stddev	0.0687	0.0716	0.0679	0.0488	0.0593	0.0703	0.0719	0.0651	0.0557	0.0613	0.0549	0.0332	0.0568	0.0571	0.0577	0.0578
R avg stddev	0.1459	0.1559	0.1443	0.1320	0.1444	0.1480	0.1480	0.1409	0.1178	0.1248	0.1135	0.0917	0.1222	0.1201	0.1090	0.1129
R std stddev	0.0928	0.0956	0.0840	0.0753	0.0895	0.0894	0.0925	0.0763	0.0974	0.1034	0.0962	0.0680	0.0980	0.0975	0.0812	0.0935
Y avg stddev	0.1221	0.1271	0.0897	0.1071	0.1173	0.1215	0.1254	0.1188	0.0762	0.0780	0.0686	0.0484	0.0769	0.0762	0.0736	0.0754
Y std stddev	0.2438	0.2451	0.0697	0.2284	0.2396	0.2371	0.2469	0.2351	0.1205	0.1221	0.0960	0.0487	0.1208	0.1202	0.1168	0.1164
Prescriptive Feedback Assigned?	No	No	No	Yes: Weapon Position (all)	No	No	No	Yes: Weapon Position (1 Group)	No	No	No	Yes: Weapon Position (all)	No	No	Yes: Breathing (1 Group)	Yes: Breathing (1 Group)

Leave-one-out cross-fold validation was used to validate the resulting models to ensure their efficacy for describing performance across the fundamentals of marksmanship in a generalized fashion. This is achieved by developing models across n-1 participants, leaving a single subject's data to confirm the resulting model matches their behavioral patterns. This was performed across all possible combinations of participants to ensure it held constant in all facets of model development. At the most basic level, this answers the question of "if this model created from the other experts was used to assess this expert, would it find this experts data to be differing?" Generally, the finding from Table 3 is that the created metrics for measurement of expert items can be created in a manner which describes the performance of other experts. If these were implemented in an ITS with the purpose of diagnosing non-expert performance, the vast majority of the experts would be able to interact with the system without receiving feedback. The exception to this observation is Expert 4, who appears to perform weapon orientation different from the other experts, and would be assigned feedback at each stage (see "prescriptive feedback assigned" in table). This expert had firing scores analyzed cursorily to answer the questions that this raises, such as "is this expert performing activities that make them better than the other shooters, which the other shooters may consider adopting?" The expert placed 5^{th} in a rank-ordered list of the 8 experts in the prone position, but 2^{nd} among the 8 experts in the kneeling position (when 33% of the values conformed to the other expert models). This warrants some amount of future investigation, as the expert conforming to other expert behavior results in increased performance, but gives confidence in the overall use of the models.

DISCUSSION

In an effort to create an adaptive marksmanship ITS, the above results support the application of a generalized model of expert performance which may be used to diagnose novice error across fundamental principles of BRM. The analysis is promising in that it highlights similar behavioral trends across experts when performing standard grouping procedures using the apparatus developed for data collection.

The next step is to integrate the established assessment parameters within GIFT. The models can then be used to support real-time diagnostic functions when a non-expert is interacting with the training platform. This will enable GIFT to collect performance and behavior data from a user, compare that data against expert performance, and to diagnose behavioral errors as informed by the expert models and the principles they are linked to. As expert models are built to designate specific trends found in the data with concepts linked to task execution, parameters will be defined for the purpose of recognizing individuals that deviate away from desired behaviors. To accomplish this, three primary tasks must be completed. First, the concepts of BRM will need to be represented within a GIFT domain model. This involves breaking down the task into a hierarchy of concepts and subconcepts. For each concept and subconcept, thresholds will be defined that designate assessment values of 'at', 'above', or 'below' expectation, as deemed by GIFT's standardized assessment logic (Sottilare, Brawner, Goldberg, & Holden, 2013). These thresholds will be used by GIFT's learner and pedagogical models to determine novice error and the type of remediation to provide in an attempt to correct and improve subsequent performance.

While this paper reports on the development of expert models for support of an adaptive marksmanship training capability, the ultimate outcome is to establish a validated testbed for the purpose of running studies across an array of topics, including feedback techniques and optimal training practices (e.g., blocked vs. random target practice). The last task before having a fully functional initial testbed environment is to develop instructional tactics and interventions for diagnosed errors and author the triggers that they act on. For the initial implementation, we will use video-based instructional interventions as previously applied by Chung et al (2009). The videos will include a SME describing the fundamental components of the concept assessed at 'below expectation' and providing tips on improving performance. For this phase of the study, this will be the only type of instructional tactic used for coaching and remediation. It is also important to note that the feedback authored will be fairly general in detail. This is due to expert models only being able to identify individuals not performing in an expert-similar manner, rather than identifying specific root causes of error.

FUTURE WORK

With an implemented version of GIFT for training BRM, phase two will result in an experiment using novice first-time shooters as the sample population. This experiment will serve two purposes. First, it will provide pilot data to test and validate the expert model's ability to interpret novice performance in real-time and diagnose errors based on outcomes. This will enable a thorough follow-on analysis to compare novices against experts and to locate common differences between both populations, as seen in (Berka et al., 2008). In addition, this experiment will provide initial evidence of whether individualized instruction autonomously delivered by an ITS is an effective form of training for learning BRM skills.

With a goal of improving the models used to diagnose error in novices, this phase will also include a panel of SMEs in BRM to observe the data generated from novices. As the initial expert models generated can be used to dictate what an individual is not doing similar to experts, these models do not have the ability to accurately determine what is truly causing the error, which in turn results in the system's inability to provide focused remedial feedback or to truly validate the approach taken during model creation. The data from witnessing experts on the use and validation of the models of marksmanship and associated feedback will be used to develop a buggy–performance library of common novice errors, as well as assess the overall validity of the above handcrafted models. The generation of a *labeled* dataset for marksmanship which identifies exactly which type of error a trainee is making will be beneficial to the research. It is worth noting that how errors are annotated will be a subjective task completed by SMEs. As James and Dyer (2011) note, this can lead to discrepancies in diagnosis as instructors and SMEs in this domain are inconsistent from each other in the tactics and procedures they teach along with their complete understanding of the concepts involved in the task. At this point, two competing models will be in existence, without the knowledge of relative superiorities. Recommendations from this follow-on work are intended to point the way to other types of psychomotor ITS systems (i.e., Do you have to collect annotated mistake information?).

CONCLUSION

This work is among the first to address the need for automated, personalized, and intelligent psychomotor training. Methods for measurement, model creation, expert validation, and novice diagnosis in this type of domain are not addressed well in other literature. This paper presents a way to measure items of interest, create models without a barrage of annotated or novice data, validate those models for reasonability, and apply them in a real context. It is hoped that future research in psychomotor training follows and improves upon the effort set forth in this work.

REFERENCES

Army, U. S. (2008). Rifle Marksmanship M16-/M4-Series Weapons (FM 3-22.9). *Fort Benning, GA.*

Berka, C., Chung, G., Nagashima, S. O., Musacchia, A., Davis, G., Johnson, R., & Popovic, D. (2008). *Using interactive neuro-educational technology to increase the pace and efficiency of rifle marksmanship training.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Chung, G. K., Nagashima, S. O., Espinosa, P. D., Berka, C., & Baker, E. L. (2009). An Exploratory Investigation of the Effect of Individualized Computer-Based Instruction on Rifle Marksmanship Performance and Skill. Report 754. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST).*

Cooper, G. (1998). Research into cognitive load theory and instructional design at UNSW. *Retrieved October, 2013.*

Inertial Labs, Inc. (2010).Weapon Oritentation Module (WOM) Data Sheet. Retrieved from www.inertiallabs.com.

James, D. R., & Dyer, J. L. (2011). Rifle Marksmanship Diagnostic and Training Guide: DTIC Document.

Pojman, N., Behneman, A., Kintz, N., Johnson, R., Chung, G., Nagashima, S., . . . Berka, C. (2009). Characterizing the Psychophysiological Profile of Expert and Novice Marksmen *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (pp. 524-532): Springer.

SCATT Shooter Training Systems. (2013). http://www.scatt.com/

Sherry, L., Feary, M., Polson, P., & Palmer, E. (2000). *Autopilot tutor: Building and maintaining autopilot skills.* Paper presented at the Proceedings Int. Conf. on Human Computer Interaction–AERO, Toulouse, France.

Sottilare, R., Brawner, K. W., Goldberg, B., & Holden, H. (2013). The Generalized Intelligent Framework for Tutoring (GIFT). In C. Best, G. Galanis, J. Kerry & R. Sottilare (Eds.), *Fundamental Issues in Defense Training and Simulation* (pp. 223-234). Burlington, VT: Ashgate Publishing Company.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). http://www.gifttutoring.org/documents.

Stottler, D., Harmon, N., & Michalak, P. (2001). *Transitioning An ITS Developed For Schoolhouse Use To The Fleet-TAO ITS, A Case Study.* Paper presented at the The Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist, 46*(4), 197-221.