Executive Summary

Article Title: Using Design-Based Research to Develop a Virtual Human Interface for Police Nystagmus Training

Authors: Marjorie A. Zielke; Djakhangir Zakhidov; Joel Rizzo; Erik DeFries; Laay Trivedi; Cecelia Marquart; Matthew Dusek; and James Prochaska

Study Purpose or Objective(s):

- In this paper the authors discussed design-based research principles that helped them iteratively design, develop, test and implement the Individual Nystagmus Simulated Training Experience (INSITE) Horizontal Gaze Nystagmus/ Vertical Gaze Nystagmus (HGN/VGN) police training simulator which is being used as a training aide in the Texas' Advanced Roadside Impaired Driving Enforcement (ARIDE) program.
- The authors described the simulation, the educational implementation contexts, the learning activities, and the needs of the various users/learners. The users/learners included law enforcement student officers, ARIDE and Drug Recognition Expert (DRE) trainers, researchers and program administrators.
- The primary focus of the article highlighted the design considerations for modeling physiological symptoms of HGN/VGN in a virtual human; strategies for implementing virtual human technology into ARIDE HGN/VGN police training sessions; and detailing how iterations of the INSITE technology impacts user/learner interactions based upon qualitative and quantitative feedback.

Key Findings:

- During HGN field testing, officers must check for conditions which could impact the validity of the HGN/VGN test. Disqualifying testing conditions include a subject with unequal or dissimilar pupil size, an inability of the eyes to track a moving stimulus sequentially (together), and resting nystagmus. Researchers referenced images and videos of impaired subjects and created a high-fidelity 3D virtual character that presents visual indicators of alcohol impairment and other conditions in which HGN/VGN tests becomes inappropriate.
- The researchers maintained acute focus on fidelity and realism through four phases of the art pipeline: concept art and character design, 3D modeling, rigging, and eye-movement generation. Math-driven unity scripts map simulated eyes and human features for virtual physical attributes associated with impairment.
- Using stimulus position data tracking motion sensing, scripts and animations were created to simulate eye and facial features that virtually follow real-world stimulus movement. Leap Motion sensors measure movements and interfaces with sub-millimeter accuracy.
- The Leap Motion controller is used to track the stimulus movement during the performance of HGN/VGN testing. The translational data recorded by the Leap Motion sensor measures heights, angles, and speeds of the moving stimulus and the resulting data are passed into the INSITE system which drives simulated eye movements that replicate realistic eye tracking.

Study Strengths:

- The authors suggest that the use of a virtual human for HGN/VGN testing allows officers to avoid performance mistakes while allowing for continuous practice opportunities that reinforce learning.
- The authors suggest that INSITE technology provides for a rare training opportunity not usually available during traditional training methods. Varying degrees of eye redness, eye wetness, dissimilar pupil sizes, prosthetic eyes, or resting nystagmus, which may indicate a pathology or high dose level of alcohol/drugs for that individual are programable/testable attributes.
- The authors used a design-based research method to estimate the effectiveness of INSITE technology on officer HGN/VGN testing performance. Design-based research is rooted in an immersive educational context with a focus on design and testing of significant interventions. Using mixed methods, introducing and improving multiple iterations, and involving collaborative partnerships between researchers and practitioners drive the research methodology especially when collaborative contexts, activities, and cultural structures support blended learning relationships.
- The article suggests that design-based research integrates reasoning into the mechanics of the simulation and by using a virtual human to represent an HGN/VGN test subject in an enforcement environment, it allows the user/learner to interact intuitively with the simulation. This creates authenticity within the simulated scenarios while allowing the instructor to provide immediate feedback on the user's performance.
- The authors' methods involved a collaborative partnership between the science of learning, technology use in learning, and its impact on the practitioner's ability to perform to set testing standards. They also assessed the evolution of the INSITE system design over time by evaluating differing iterations of the technology; the simulations impact on the state of practice; testing environments grounded in a real-world educational contexts; and focusing on the design and testing as a significant intervention used to aid learning and retention of practical testing techniques.
- The authors suggested that one critical factor in the successful development and utilization of INSITE was the developers in-depth understanding of educational and workplace contexts where technology and process overlap. INSITE was purposefully designed to complement a blended learning methodology that enhances opportunities for practical evaluation techniques that are taught in the classroom to be linked to a virtual environment.
- The authors emphasized that INSITE technology was developed to improve officer confidence in performing HGN/VGN testing. High fidelity, easily accessible virtual reality trainers such as INSITE offer flexible, measurable and cost-effective ways to enhance student learner environments while forming meaningful solutions that raise student learner confidence in administering HGN/VGN tests. Ultimately, officers who retain knowledge, skills, and abilities are able to better access that knowledge which enables them to make better arrest and release decisions.
- The authors strongly suggest that INSITE technology provides a virtual construct that when blended with traditional classroom cognitive learning methods, can enhance retention of knowledge and improve HGN/VGN testing evaluation performance.

Study Weaknesses:

- The authors of the article provided a good definition of what design-based research principles are and how the principles are intended to compliment simulations as an educational countermeasure for improving ARIDE training, however, there was limited information provided on the research design that was used to measure performance improvements of officers using the INSITE technology pre- and post-treatment.

  Research design fits into the evaluative process by framing questions that analyze and report officer testing performance improvement. The theory begins with a supposition and uses a method to help guide researchers with detecting and collecting their observations. Ultimately, each observation should provide a measured value of the worth that supports the theory expressed. Using deductive reasoning to derive a set of propositions from the theoretical exercise does this.

  It is recommended that the authors expand and develop the linked benefits of how using the INSITE technology improves learning and retention of knowledge in the ARIDE training. In this light, the reader will be able to make meaningful connections to how improvement metrics were assigned, assessed, analyzed and interpreted to support future use of INSITE technology for supplementing traditional ARIDE training.

- *"The angle of onset for nystagmus determines the severity of impairment, and directly correlates to a subject's Blood Alcohol Concentration (BAC)." (p. 1)*

  While the angle of nystagmus onset does suggest a greater BAC for the individual, the severity of impairment cannot be specified or quantified with its use. Each individual is different by physiological/biological design and as such outliers exist that make quantification of blood alcohol concentrations impossible using angle of onset. Formulas such as 50 – Angle of Onset = BAC may be used to estimate a possible BAC, but the formula cannot be used to quantify a specific BAC for the subject. The formula is meant to be a "ballpark" estimation and not intended to be anything more than just an approximation. The authors should consider revising the sentence since BAC levels and the severity of impairment cannot be directly correlated to a subjects BAC based upon angle of nystagmus onset.

- *"Scientific evidence establishes that the horizontal gaze nystagmus test is a reliable roadside measure of a person's impairment due to alcohol or certain other drugs" (p. 1)*

  Similar to the first bullet, the authors have suggested that HGN can be used as a measure of a person's impairment. The authors should consider revising the sentence since the presence of HGN cannot specify/quantify a person's level of impairment.

- *"In addition to HGN, Vertical Gaze Nystagmus (VGN) is present in eye movements up and down, and can also help determine a subject's level of impairment." (p. 1)*

  Similar to the first and second bullet, the authors have suggested that VGN can be used as a measure of a person's level of impairment. While the presence of VGN can indicate a high dose level of alcohol for that person or indicate the presence of some drugs other than alcohol, it does not allow for an evaluator to quantify or determine a subject's level of impairment. The authors should consider revising the sentence since VGN cannot specify/quantify a person's level of impairment.

Study Limitations:

- Decisions about the use of virtual and other simulation-based educational technologies should consider the link between the need for educational outcomes and how technology-driven tools can be used appropriately.
- Little to no information was provided on how the research team measured the INSITE technology's impact on the learner's ability to retain the HGN testing methods or procedures.
- Little to no information was provided on how the collaborative joining of the learner and the technology enhanced learning; helped to measure improvement in conducting valid HGN tests; or improved the learners ability to discern eye clues that were associated with impairment.
- There was some discussion on improving officer performance from using INSITE. Officers included in the rollout trainings indicated that use of INSITE greatly increased their confidence in administering the test in three areas: Overall performance of the HGN test -- 22 percent increase in confidence; Improvement in HGN testing techniques -- 24 percent increase in confidence; and the Ability to make an arrest decision -- 44 percent increase in confidence. However, while these numbers are impressive, they stem from self-evaluations (post-training feedback) and are not reflective of actual observed performance improvements based upon measured metrics of pre and post HGN/VGN testing procedures.

Future Considerations:

- The INSITE technology has tremendous potential to help with how law enforcement officers learn HGN/VGN testing techniques. Using virtual simulations to aide in knowledge gain and retention is strongly encouraged to help improve an officer's ability to detect impairment in a subject's eyes. But in order to make an impairment determination, the officer must first administer and interpret the HGN/VGN test results correctly.

  The use of INSITE technology in a virtual environment is a positive first step in that direction. While the integration of technology into the field of law enforcement is highly encouraged, researchers must be careful to ensure that evidence-based research methods are performed in robust ways that validate the technology's usefulness and positive influence on learning. With regard to using INSITE technology to improve learning, there appears to be some evidence that supports the theory. However, it appears that more research is needed to conclude with any degree of certainty that the use of the technology can empirically improve administrative and interpretive performance of officers in the field post-completion of ARIDE training.

Subcommittee Commentary

Prosecutorial Perspective:

- Not much for prosecution other than alerting to a new method of training that some officers may be undergoing.
- This could be important for prosecutors to know about who train officers on DWI enforcement and to get additional feedback on whether they like using the INSITE
- Finally, it might be helpful if a prosecutor is using INSITE to demonstrate HGN to juries.

Enforcement Perspective:

- The INSITE simulator has the potential to be of great value in training officers in how to properly carry out the Horizontal Gaze Nystagmus (HGN) and Vertical Gaze Nystagmus (VGN) tests. The ability to program in the clues and obtain feedback as to whether the student's movements and observations are correct, are valuable from a training perspective. INSITE could be a bridge to enhancing the student's understanding between the instructor demonstration and drinking subject testing.
- While the system is claimed to be "cost-effective," the paper does not indicate how many INSITE systems were used during a training class, the number of times each student was allowed to utilize the system, and the length of time it took to complete student testing. HGN is only one of three tests that each student must be evaluated on, and the paper indicates that one HGN/VGN test takes 15-20 minutes. The ARIDE or SFST Basic Practitioner classes have a set amount of time for each portion of the training, including for HGN/VGN with HGN in the Basic class being more than just demonstrating the movements. Classes for ARIDE can have over 20 students, so how much time will each student be able to spend with INSITE during a 16-hour class? Do they use multiple systems in order to process more students? What is the cost of each system and how much does the cost increase if multiple systems are needed? Would additional instructors be needed with an increase in the number of systems used? The ARIDE student should be prepared to demonstrate proficiency in the SFST when they arrive at the class, so is INSITE a cost-effective method of showing experienced SFST trained officers how to do a test they should already be proficient in?
- The ARIDE participants and cadets indicated significant increases in confidence in their performance of the test, improvement in the HGN test technique, and their ability to make an arrest decision. Are there differences in the confidence of ARIDE officers as opposed to that of cadets? As mentioned above, ARIDE participants are all experienced and trained officers, while cadets lack knowledge and any real-life experience in the test. You would expect the confidence of a cadet to increase dramatically after using the INSITE system. The ARIDE students should be experienced in field testing of impaired driving suspects. NHTSA recommends that the Basic Practitioner SFST course utilize a wet lab, allowing students to test live drinking subjects. Had the cadets already tested drinking subjects prior to using INSITE or did they do so afterwards, if at all? It would be helpful to do follow up testing with participants to see if they found a noticeable improvement in their performance of HGN/VGN and correct arrest/no arrest decisions after using INSITE. A study

comparing testing and correct decision-making between INSTITE and non-INSITE trained students is needed to tell whether the INSITE system brings lasting value to the training experience.

Toxicology Perspective:

- This seems like a promising tool to use to provide more practice for officers for nystagmus
- training, but it should not replace traditional training techniques with live subjects and
- measurable alcohol levels.

Research and Evaluation Perspective:

- There was limited information provided on the research design that was used to measure performance improvements of officers using the INSITE technology pre- and post-treatment. It is recommended that the authors expand and develop the linked benefits of how using the INSITE technology improves learning and retention of knowledge in the ARIDE training.
- The INSITE technology has tremendous potential to help with how law enforcement officers learn HGN/VGN testing techniques. However, it appears that more research is needed to conclude with any degree of certainty that the use of the technology can empirically improve administrative and interpretive performance of officers in the field post-completion of ARIDE training.

Using Design-Based Research to Develop a Virtual Human Interface for Police Nystagmus Training

Marjorie A. Zielke
Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab
The University of Texas at Dallas
Richardson, USA
margez@utdallas.edu

Djakhangir Zakhidov
Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab
The University of Texas at Dallas
Richardson, USA
dxz021000@utdallas.edu

Joel Rizzo
Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab
The University of Texas at Dallas
Richardson, USA
joel.rizzo@utdallas.edu

Erik DeFries
Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab
The University of Texas at Dallas
Richardson, USA
exs124530@utdallas.edu

Laay Trivedi
Center for Modeling and Simulation/Vitual Humans and Synthetic Societies Lab
The University of Texas at Dallas
Richardson, USA
lut170000@utdallas.edu

Cecelia Marquart
Texas Impaired Driving Initiatives
Sam Houston State University
Huntsville, USA
icc_cpm@shsu.edu

Matthew Dusek
Northeast Police Department
Texas
trooperdusek@yahoo.com

James Prochaska
Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab
The University of Texas at Dallas
Richardson, USA
jlp140630@utdallas.edu

***Abstract*— The Individual Nystagmus Training Simulation Experience, or INSITE™, is a virtual human simulation program to help train police officers in identifying one of the strongest clues of alcohol impairment in drivers – nystagmus, or involuntary rapid movement of the eyeball. In this paper we talk about the design-based research principles that helped us iteratively design, develop, test and implement this police training simulation as part of the Texas' Advanced Roadside Impaired Driving Enforcement (ARIDE) program. We describe the simulation, the educational implementation context, the learning activities, and the identified needs of the various users -- including trainees, trainers, researchers and program administrators. Most importantly, we discuss the evolution of the user experience over time in response to feedback. This paper focuses on: 1) design considerations for modeling physiologic symptoms of nystagmus in a virtual human; 2) the strategy for implementing INSITE™ into ARIDE police training sessions; 3) detail on the numerous iterations of the multi-leveled user interface and experience based on qualitative and quantitative feedback from trainees, Standard Field Sobriety Test (SFST) instructors, and subject matter experts; and (4) an overall summary of our experience on this design to date.**

Keywords—UX/UI design, human-computer interface, design-based research, simulation, tracking and sensing, modeling and simulation, computer graphics techniques.

I. Introduction

Every 50 minutes in the US, someone dies in a traffic accident involving an alcohol-impaired driver [1]. In 2016 this accounted for 28 percent of U.S. traffic deaths [2]. In our home state, according to the Texas Department of Transportation: "Texas has a DUI-alcohol crash every 20 minutes and 37 seconds" [3]. In 2013, "Texas led the nation with 1,337 deaths

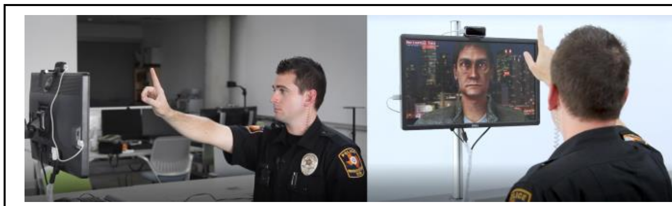

Fig. 1. Officer conducting the nystagmus test on a virtual impaired subject using the INSITE™ system.

caused by a drunk driver" [4]. In humans, nystagmus occurs naturally and is "an involuntary rapid movement of the eyeball, which may be horizontal, vertical, rotatory or mixed" [7]. Horizontal Gaze Nystagmus (HGN) refers to involuntary eye jerking motions when looking far to one side. The angle of onset for nystagmus determines the severity of impairment, and directly correlates to a subject's Blood Alcohol Concentration (BAC). HGN is one of three standardized field sobriety tests (SFST) [5]. "Scientific evidence establishes that the horizontal gaze nystagmus test is a reliable roadside measure of a person's impairment due to alcohol or certain other drugs" [5]. Many officers consider HGN to be one of the best clues to impaired driving, yet, HGN is subtle and difficult to see. In addition to HGN, Vertical Gaze Nystagmus (VGN) is present in eye movements up and down, and can also help determine a subject's level of impairment. Its presence and severity depends on the subject's personal alcohol tolerance level. As powerful as these techniques are, officers often don't rely on these important roadside sobriety tests because they lack confidence in recognizing these conditions [6].

Further training sessions often have to offer wet labs, which require "dosing up" human participants. Asking participants to become inebriated at various levels is expensive, presents potential liability, and is administratively intensive. The use of a virtual human allows us to avoid these pitfalls while offering continuous practice for officers. As shown in **Figure 1,** the Individual Nystagmus Simulated Training Experience, or INSITE^{TM}, allows focused practice of nystagmus tests -- regarded as the most critical portion of the SFST. INSITE^{TM} is designed to simulate the potential eye conditions of impaired drivers caused by varying levels of BAC. A virtual impaired character named Brian is able to accurately display a wide range of alcohol impairment such as 0, 0.05%, 0.08%, and 0.15% BAC through simulations of lack of smooth pursuit, varying degrees of onset for HGN, and distinct and sustained nystagmus at maximum deviation. As shown in **Figure 2** below, INSITE^{TM} enables rare training opportunities not usually available during conventional training methods, such as varying degrees of eye redness (panel A), eye wetness (panel B), dissimilar pupil sizes (panel C), prosthetic eyes (panel D), or resting nystagmus, which occurs when the eyes jerk as the subject is looking straight ahead -- and may indicate a pathology or high dosages of drugs such as PCP [10]. Checking for the angle of onset for nystagmus is illustrated for HGN in panel E1 and for VGN in panel F1. Distinct and sustained nystagmus at maximum deviation refers to being able to clearly see each eye jerking when the eye is gazing at a stimulus which has been moved as far to the side as possible and held there for about four seconds, as shown in panels E2 for HGN and F2 for VGN. Lack of smooth pursuit occurs when the eyes cannot follow a stimulus smoothly. In addition to providing abundant nystagmus practice, INSITE^{TM} assesses trainees' understanding of nystagmus at the end of each session.

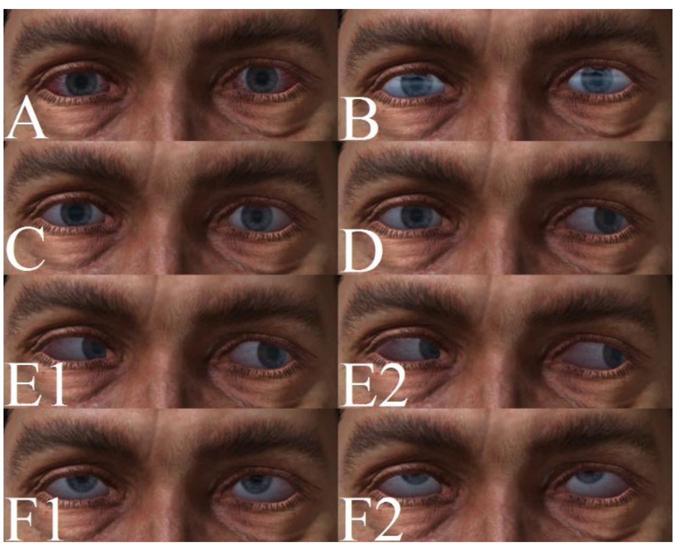

Fig. 2. Possible eye conditions which can be represented in Brian.

II. SIMULATING NYSTAGMUS IN A VIRTUAL HUMAN

The HGN test is based on the premise that the automatic tracking mechanisms of the eyes are affected by alcohol [8]. Alcohol slows down the eyes' ability to rapidly track objects and causes the eyes to jerk before they normally would in a sober person [9]. During a roadside stop, an officer checks for three

Funded by internal funds from the University of Texas at Dallas, Sam Houston State University and support from the Texas Department of Transportation (TxDOT).

Fig. 3. Brian is a virtual impaired character able to represent physiological symptoms of horizontal and vertical gaze nystagmus, lack of smooth pursuit, resting nystagmus, and other conditions.

signs of HGN: lack of smooth pursuit in the left and right eye, distinct and sustained nystagmus at maximum deviation in the left and right eye, and the onset of nystagmus prior to 45 degrees in the left and right eye [10]. To determine the angle of onset of nystagmus, the stimulus is moved slowly to the side. As each eye moves to the side, if jerking occurs before the eye reaches the 45 degree angle of gaze, impairment is present and the BAC is .08 or higher, depending on the angle of onset [10]. The officer also needs to check for conditions which may prevent the validity of an HGN test. Such conditions include pupils of dissimilar size, a subject's inability to track equally with both eyes, and resting nystagmus. We have researched reference images and videos of impaired subjects to create a high-fidelity 3D virtual character, Brian, who is able to represent visual symptoms of alcohol impairment and conditions in which an HGN test is not valid, as described above.

A. 3D Character Design

As depicted in **Figure 3**, a high-fidelity 3D virtual impaired character with photorealistic facial features and a lifelike appearance has been developed in Maya software. When creating Brian, we maintained acute focus on fidelity and realism through the four phases of our art pipeline: concept art and character design, 3D modeling, rigging, and eye-movement generation. Brian employs a sophisticated series of math-driven Unity scripts that map Brian's eyes and physical features to physical phenomena associated with impairment. By using stimulus position data tracked by a Leap Motion sensor, these scripts and animations allow Brian's eyes and facial features to follow a real-life officer's stimulus movements.

B. Leap Motion Sensor for Stimulus Tracking

The Leap Motion controller is a sensor for hand gesture controlled user interfaces with sub-millimeter accuracy [11]. In INSITE^{TM}, the Leap Motion controller is used to track a police officer's stimulus during the performance of the HGN test. The Leap Motion controller is placed on top of a monitor display on which Brian's face is projected in near life-size proportion. Trainees can place a vertically oriented stimulus 12 to 15 inches away from Brian's face projected on the monitor display, then move it in a horizontal line parallel to the suspect's shoulders as recommended by the National Highway Transportation Safety Administration (NHTSA) guidelines [12]. The translational data recorded by the Leap Motion sensor includes measurements of height, angle, and speed of the moving stimulus. These data are passed into the system and drive Brian's eye movements in order to allow for realistic eye tracking. Using linear interpolation and a detailed mathematical model, Brian's eyes in the virtual world are coded to smoothly follow the real-life position of the trainee's stimulus. Based on the calculated angle and speed the trainee's stimulus exhibits at any moment, Brian's eyes are further coded to reflect the conditions he has been set to have, such as resting nystagmus and HGN.

III. Principles of Design-Based Research

Design-based research (DBR) provides a substantial platform for cross-disciplinary work such as presented in our project. A design-based research is one that is: situated in a real educational context, focuses on the design and testing of a significant intervention, uses mixed methods, involves multiple iterations, and involves a collaborative partnership between researchers and practitioners [13]. "Intertwining design and research is especially important for establishing collaborative contexts, or activities and cultural structures that support collaboration leading to learning" [14]. DBR allows scientists to consider details, such as multi-layers of participants, e.g. instructors and trainees; nuances such as the learning context, e.g. law enforcement education and culture; and integration of these factors with technology research in an iterative design approach [15].

IV. DBR in Health-Related Educational Simulation Research

The literature offers specific examples of DBR when applied to health-related simulations. For example, Koivisto et al. used DBR to iteratively design, develop, test, and refine a simulation game for nursing education for improving clinical reasoning [16]. Koivisto et al. describes several iterative cycles of design, development, and testing and the resulting design principles which emerged based on the empirical knowledge gained from each iteration cycle. The design principles included: integrating clinical reasoning into the game mechanics, using high quality graphics to represent a 3D patient and hospital environment, allowing the user to interact intuitively with the patient and the hospital environment, creating authentic and realistic patient scenarios, and providing immediate feedback on performance [16]. Dornan et al. used DBR to explore how medical students learn in a self-directed way in the clinical environment [17]. The researchers explored if students could be entirely self-directed in a clinical environment and concluded students were rarely fully autonomous or subservient [17]. They valued affective and pedagogic support, and relied on teachers to manage their learning environment [17]. McGaghie and colleagues posit that effective use of simulation depends on a close match of education goals with simulation tools [18]. Virtual Reality simulators are now in use to educate surgeons and medical subspecialists in complex procedures that are too dangerous to practice on live patients. However, decisions about the use of these and other simulation-based educational technologies should consider the match between goals and tools [18].

V. Applying DBR Principles to INSITE™

DBR incorporates specific characteristics that fit our project's research goals. These include: involving a collaborative partnership between learning science and technology researchers and practitioners; evolution of design principles; comparison to action research; practical impact on practice; being situated in a real educational context; and focusing on the design and testing of a significant intervention [19]. One of the key tenets of DBR is involving end-users early on in the creative process of designing a simulation and doing

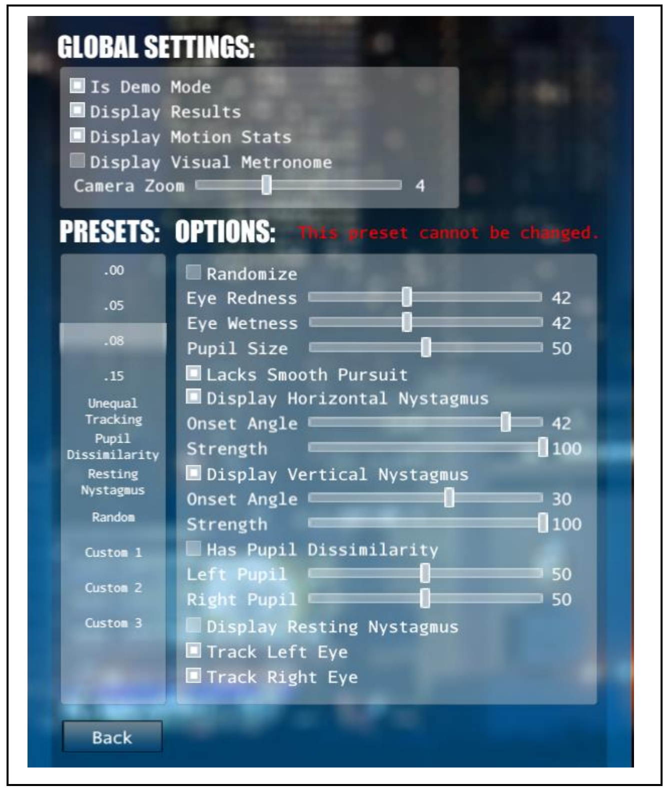

Fig. 4. INSITE™ offers a variety of settings for constructing unique scenarios.

early usability experiments. In our case, three key types of end-users have been identified: trainees, trainers, and administrators. As such, critique, validation and other input from all of these users was collected consistently throughout every design and development cycle of INSITE™. Another critical factor in the successful development and utilization of an assistive technology, such as INSITE™, is in-depth understanding of the educational and workplace context in which this technology will be used.

VI. Educational Context

INSITE™ has been designed to supplement the traditional law enforcement training offered by the Texas' Advanced Roadside Impaired Driving Enforcement (ARIDE) program. ARIDE courses are taught by Drug Recognition Expert (DRE) instructors, last about two days, and cover the full battery of SFST, including nystagmus. ARIDE training of nystagmus consists of lectures, practice sessions with a non-intoxicated partner, and proficiency testing by a DRE instructor. In addition to ARIDE courses, INSITE™ was also implemented at SFST wet labs, which involve practicing HGN on volunteers who have consumed alcohol. INSITE™ has been designed to complement existing blended learning curriculum and enhance training opportunities by providing more focused and hands-on training of HGN than is currently available in ARIDEs and wet labs. As mentioned above, INSITE™ also allows training on subjects with rare conditions not typically encountered at ARIDEs and wet labs, such as subjects with dissimilar pupil sizes, prosthetic eyes or resting nystagmus.

VII. The Evolution of INSITE™

The initial prototype of INSITE™ was developed in 2015 by researchers at the Center for Modeling and Simulation/Virtual Humans and Synthetic Societies Lab at the University of Texas at Dallas in collaboration with subject matter experts from Sam Houston State University and Eye T Plus. The initial prototype consisted of a simple user interface, a much shorter overall user experience, and a settings menu with limited parameters. INSITE™ received funding from the Texas Department of Transportation (TxDOT) in 2017 to roll out in seven Texas ARIDE programs and two SFST wet labs, educating 150

Fig. 5. The motion stats display the angle and distance of the user's finger relative to Brian. The visual metronome enables trainees to see live feedback about the speed and timing of their finger/stylus movement.

officers. Currently, with additional funding from TxDOT in 2018, INSITE™ is being further integrated into the Texas ARIDE program including instructor training. INSITE™ is positioned to help educate up to 500 officers and up to 50 instructors in 2019, with potential to educate officers from the Department of Public Safety, Parks and Wildlife, Sheriff Departments, City and University enforcement entities and other types of law enforcement officers who attend the training.

VIII. INSITE™ User Experience

In this section we focus on the current INSITE™ user experience, and in the next section we will describe the iterations of the user interface which led to the current user experience. The current INSITE™ user experience lasts 15-20 minutes on average, and consists of five steps: 1) Pre-Survey; 2) Pre-test; 3) the Nystagmus test; 4) Assessment; and 5) Post-survey. The current user experience is a result of numerous iterations to the user interface, as result of feedback from trainees, training instructors, and subject matter experts.

1) Pre-Survey: The user completes survey questions regarding their demographic data (age, education, gender, etc); their levels of experience and confidence in properly performing the Nystagmus test; and their experience as a law enforcement officer.

2) Pre-Test: The user checks Brian's condition, looking for the three primary pre-test conditions: equal pupil size, equal tracking, and absence of resting nystagmus. If these three conditions are normal in Brian, the user can proceed to the HGN and VGN tests. However, if the user notices that Brian has unequal pupil size, resting nystagmus, or unequal eye tracking, then the user has the opportunity to ask him questions, and/or stop the test.

3) The Nystagmus Test: The user performs stimulus passes in front of Brian's eyes, looking for signs of HGN or VGN. If nystagmus is present i.e. if Brian's eyes make involuntary jerking motions when they're looking far toward either side, the user will determine the angle from the center that nystagmus begins, and use that to estimate Brian's BAC.

4) Assessment: Following the nystagmus test, the user will be asked two assessment questions based on their observations:

a) Is distinct and systained nystagmus present?

b) If so, is the subject's BAC .08 or geater?

After answering, the user has the option to view the correct answers.

5) Post-Survey: The user is asked a series of usability questions about the INSITE system as well as questions about their levels of confidence recognizing nystagmus and performing the HGN/VGN tests.

Prior to the practice or test session, the instructor can adjust a number of variables in the settings, such as eye redness, wetness, and pupil size, left and right pupil dissimilarity, and set the angle of onset for HGN and VGN. By adjusting these settings, instructors can construct scenarios with varying levels of BAC, as well as situations that are rarely seen, like pupil dissimilarity.

IX. Iterations of the User Interface Development

Currently, the INSITE experience begins at the main menu. The user interface allows the user to start a session, change settings, or exit the program. As shown in **Figure 4**, in the settings menu the user can change global settings which includes system-wide functionality like camera zoom, or select presets which control Brian's eye behaviors. Some presets are hard-coded and cannot be changed, while others allow for custom preset definitions. Users have the ability to turn on or off: resting nystagmus, equal tracking, pupil dissimilarity, lack of smooth pursuit, HGN, VGN, and whether the system should randomize the settings. The user can also adjust the strength and onset angle for HGN and VGN, as well as change Brian's eye redness and wetness.

When the user begins a session, if the system is not in demo mode, they will participate in a short survey before progressing to the pre-test. The user has control over Brain's eyes during the pre-test and the HGN test, and Brian's eyes will behave according to the currently selected preset and settings. At any time the user has control over Brian's eyes, if "display motion stats" is enabled, the system will display certain statistics about the user's actions, such as current stimulus angle and distance. The visual metronome, if enabled, will also be displayed to allow the user to optionally practice their timing and speed. The visual metronome allows trainees to practice their technique by following along with a visual marker that represents the correct speed at which the HGN test should be conducted. **Figure 5** represents the motion stats and the visual metronome.

Upon completing the pre-test, the system will ask the user questions about whether certain pre-test conditions were noticed. Then, if "display results" is enabled in the global settings, the system will show the user whether their answers were correct. From there, the user can choose to conduct the HGN test, stop the test and provide a reason why, or ask Brian pre-programmed questions which change based on the context. If the user chooses to conduct the HGN test, they will once again have control over Brain's eyes and then just like the pre-test before, the user will answer questions based on their observations, and if "display results" is enabled the system will report the correct answers. Regardless of whether the user conducted the HGN test or stopped early, the system will offer the user an opportunity to start the test from the beginning to practice again, and if "randomize" is enabled, they will receive a totally different set of eye settings to practice against. Once the user is done testing, if demo mode is off, the user will complete a short post-survey, and then the session is complete.

A. *UI Design and Development Iteration #1 (2014-2015)*

The initial prototype of INSITE had only a small portion of the user interface described above and a significantly shorter overall user experience. The HGN test and the post-test observation questions were the only features implemented. The Settings menu allowed for changing of eye redness, wetness, turning on/off correct answer display and motion stats, and turning on/off HGN and pupil dissimilarity. The prototype was validated by DREs in Dallas/Fort Worth, and tested at two training academies: Oklahoma’s Council on Law Enforcement Education and Training (CLEET) and the Dallas Police Academy. The feedback was encouraging and resulted in a determination to pursue additional funding for further development of the technology.

B. *UI Design and Development Iteration #2 (2016-2017)*

This iteration focused on preparation of the INSITE system for 2018 ARIDE experiments. This iteration of development added new functionality for Brian to represent lack of smooth pursuit, resting nystagmus, and unequal tracking. This version also provided improved accuracy and realism of tracking system/eye behaviors which was an issue observed by one of our collaborators. To increase the utility of the system, also as suggested by one of the users, we added nystagmus onset angle and strength settings.

C. *UI Design and Development Iteration #3 (Early 2018)*

This iteration started with the early 2018 ARIDE experiments and garnered a large amount of quantitative and qualitative feedback from the users. One of the comments was that the pre-test and main test should be separated. We broke the pre-test apart from main test, and gave it three distinct sections. Users now could pick which pre-test conditions to check for in any order. This was done to give the system a more guided user experience, and to reduce noise incurred between pre-test sections and improve the accuracy of our data analysis techniques by breaking each pre-test section into a distinct, isolated, and thus easier to capture block of data points while maintaining the ability of the user to freely select pre-test sections. We also added new UI elements that provided textual feedback to users periodically. This was done to address concerns that the system had too little context and information throughout the pre-test and test portions. Another key new development was the visual metronome, which was added to address concerns that the system needed feedback about technique performance. In addition, the preset system was added to the settings menu. This feature was requested by trainers to allow for quicker and simpler simulation customization. By adding, removing, and setting as many presets as desired dynamically, trainers could switch between different configurations of Brian so as to simulate different levels of impairment without needing to individually and manually manipulate every possible setting between each usage of the system.

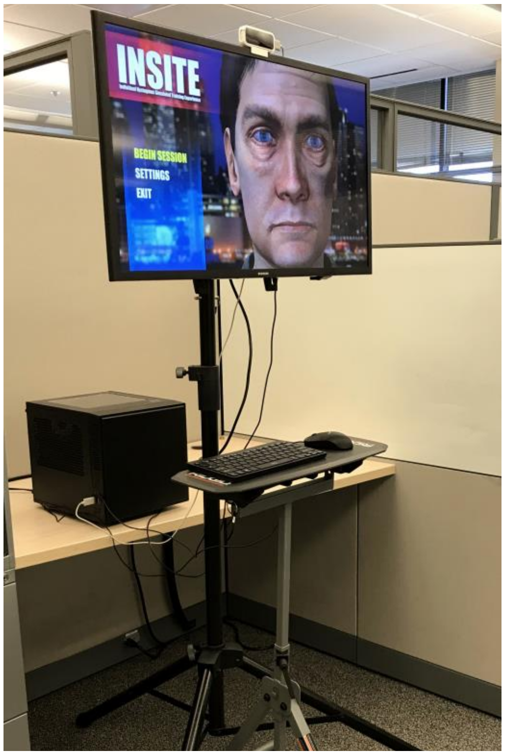

Fig. 6. The INSITE™ system provides a solution for taking the simulation out in the field and integrating into ARIDE sessions.

D. *UI Design and Development Iteration #4 (Mid-to-End of 2018)*

This iteration started with the testing of the first custom-built dedicated INSITE™ mobile systems. Six INSITE™ easily transportable systems were assembled to enable moving and setting up systems at remote locations. As shown in **Figure 6**, each transportable system is comprised of a computer, a 32" display monitor, a Leap Motion sensor, a custom Leap Motion mounting socket, a collapsible stand for the display monitor, a folding portable keyboard stand, a mouse, a keyboard, and cables. The INSITE™ system manual, a 20 page-long document, was also produced. The INSITE™ system manual provides detailed setup and operation instructions. User feedback from this iteration cycle was abundant and resulted in the following changes to the interface:

- We removed three distinct pre-test sections and unified them all into one block, with observational questions afterward. This was done to reduce confusion and system complexity.
- We added the "randomize" setting. This was done as response to feedback that system should be able to challenge and test user.
- We integrated our pre-survey and post-survey directly into the system itself as an onboard electronic component to replace the pen-and-paper solution used previously. This enabled the system to behave more autonomously by eliminating the need for separate survey distribution and collection, and helped to keep participant data well-organized and grouped.
- We added the ability for the system to conduct multiple tests in a single session. This was done to address the issue that any user wanting to practice more had to repeat pre and post survey needlessly.
- We added the demo mode. This was done to enable the system to be demoed without its pre/post survey components and without data tracking needlessly.

E. *UI Design and Development Iteration #5 (2018-2019)*

This iteration came after the completion of the 2018 ARIDE sessions and focused on updating INSITE™ for the 2019 ARIDE sessions. Based on user feedback, we reduced the length and altered the wording of the pre/post surveys. This was done to address feedback that the surveys were too long. We changed the indicator that originally displayed finger speed to instead display whether the Leap Motion sensor is correctly tracking the user's finger, as the visual metronome had effectively replaced the speed readout in functionality and purpose. We added the ability to turn off the visual metronome and made it less prominent. This was done to address concerns that the metronome was distracting.

F. *UI Design and Development Iteration #6 (Early 2019)*

This iteration started with a new round of 2019 ARIDE sessions. In addition to training the police officers, we focused on providing training sessions to instructors who will be in charge of using the INSITE™ system at ARIDE sessions. The feedback from instructors was very encouraging and valuable for ensuring we keep enhancing the system to be adopted by DRE instructors. As a result of instructor feedback, we added hard-coded, permanent presets; added a constant number of custom presets; and removed the dynamic presets system. The existing presets system was replaced with one that came pre-packaged with several fixed and common presets as well as a constant number of configurable presets; the ability to add and remove presets limitlessly was revoked. This was requested by instructors to ensure common presets would always be available and remain consistent across independent INSITE™ systems. We implemented VGN formally as a response to a growing prevalence of employing VGN in training regimens; we observed that VGN has become a more common technique and the importance of its inclusion in the system became evident.

X. Results From Experiments

The main focus of the INSITE™ training intervention, when integrated into ARIDE and wet lab sessions, was to improve officer confidence in performing the HGN test, since this can increase the impaired driver arrest rate. Officers in all ARIDEs and cadets in a local police academy included in the rollout indicated that use of INSITE™ greatly increased their confidence after using INSITE™ in three key areas:

1) *Overall performance of the HGN test* -- 22 percent increase in confidence.

2) *Improvement in HGN test technique* -- 24 percent increase in confidence.

3) *Ability to make an arrrest decision* -- 44 percent increase in confidence.

The data further indicate that officers with less experience performing the HGN test also had significant increases in confidence in each of the above areas, suggesting that INSITE™ provides important practice opportunities. These confidence increases were determined using a pre and post INSITE™ standardized t-test and are statistically significant at the 1 percent confidence level, n=130. We attribute these results to the successful implementation of the described DBR approach, which led to careful analysis of feedback from all key users and the numerous iterative enhancements of the experience for trainees, instructors, and administrators.

XI. Discussion

Our close collaboration with subject matter experts and careful analysis of feedback from system users enabled us to continuously refine the INSITE™ simulation. When analyzing feedback from users it is important to explain the three key types of users and their respective feedback. The police officers and cadets, or the trainees of the system, provided abundant feedback on system usability; the trainers, such as DREs, provided substantial input on how to make INSITE™ easier to use for the educators and how to better integrate INSITE™ into the ARIDE curriculum; and the administrators of the program, such as experiment executors and researchers, provided much useful feedback on how to make INSITE™ systems more portable and easier to operate and parameters for using the research in live training sessions.

A. Feedback from Trainees

The most useful comments on system usability from trainees focused on removing unnecessary complexity from the system, shortening the length of the experience without sacrificing educational outcomes, and improving the users expectations of what to expect at each step of the intervention through descriptive/transitional prompts. Trainees also shared many suggestions for future enhancements of the system, such as an ability to speak to the drunk driver. Many trainees desired a way to speak naturally to Brian during an HGN test. Trainees also suggested adding different types of virtual humans to represent alcohol impairment. Several trainees expressed interest in having an INSITE™ station in their police department, to be able to practice as needed on an ongoing basis.

B. Feedback from Instructors

The feedback from DRE instructors was particularly useful for better integrating INSITE™ into ARIDE sessions. Several DRE instructors expressed that they would feel comfortable using INSITE™ fully integrated into ARIDE course curriculum. Some DRE instructors expressed an interest in using INSITE™ for assessing officers' knowledge, understanding, and performance of the HGN test. Several DRE instructors mentioned another potential educational use of INSITE™ in courtrooms to explain nystagmus to jurors. Placement of INSITE into the overall curriculum was also part of the usability discussion from the viewpoint of instructors.

C. Feedback from Administrators

The administrators of the program provided suggestions on improving the practical ease of use of the INSITE™ system. The systems were re-designed twice to make them lighter in weight for easier transport and more modular in design for the ability to easily replace missing items. Administrators were also tasked with streamlining the set-up and calibration of the INSITE™ system so that a non-technical individual could launch and run the intervention. For this purpose, a detailed setup manual was created, providing steps for setting up the system, launching it, collecting and extracting data, and explaining the settings menu. The administrator from the ARIDE program was also able to facilitate curriculum adjustments to accommodate the instructors' feedback.

D. Fedback from Different Types of Law Enforcement Users

Additional feedback was gained from conversations with law enforcement officers. A game warden commented on the usefulness of INSITE™ for training nystagmus testing to conservation officers responsible of making arrest decisions on water. Given the limited space on a boat, the walk-and-turn and one-leg stand portions of SFST are often not possible, and HGN can be the primary method for determining if a boat driver is impaired. Nystagmus testing is also useful when assessing alcohol impairment in subjects who are laying down on a gurney, perhaps as a result of injuries sustained during accidents. These detailed usage scenarios provide opportunities for additional enhancements of the INSITE™ user experience and user interface.

XII. Remaining Challenges

To date several challenges remain. One of the key challenges involves the automation of secure data collection. Currently, trainee performance data is collected manually from each INSITE™ system after an ARIDE session. The automation of this data collection is difficult because there is no certainty about the type of and security of internet connection available at varying remote ARIDE locations. Another ongoing challenge is the ongoing need to make the INSITE™ systems easy to transfer, assemble, and use. We continue to work toward making the nystagmus testing environment on INSITE™ as close to the actual field experience as possible.

XIII. Summary

High fidelity, easily accessible virtual reality trainers such as INSITE™ offer a flexible, measurable and cost effective means to provide a portion or the full scope of SFST training and hold promise as a major solution for raising police officers' confidence in administering nystagmus tests, thereby enabling better arrest decisions. Using DBR principles, we systemically adjusted various aspects of the user interface so that each adjustment served as a type of experimentation that allowed us to test user experience in naturalistic settings. We believe that through keeping end user requirements at the forefront of every design decision we have developed a learning tool that can get implemented as part of state-wide and national programs to train law enforcement officers to better identify alcohol impairment in drivers, make more accurate arrest decisions, and save more lives.

XIV. ACKNOWLEDGMENTS

We would like to acknowledge our research collaborators at Sam Houston State University and EyeT Plus. We would like to express gratitude to the Texas Department of Transportation, as well as Texas constituents including police entities, DRE instructors, and law enforcement officers who participated in testing and validating the INSITE™ system.

REFERENCES

[1] Impaired Driving: Get the Facts | Motor Vehicle Safety | CDC Injury Center. (2019). Retrieved from https://www.cdc.gov/motorvehiclesafety/impaired_driving/impaired-drv_factsheet.html

[2] National Highway Traffic Safety Administration. Traffic Safety Facts 2016 data: alcohol-impaired driving. U.S. Department of Transportation, Washington, DC; 2017 Available at: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812450 Accessed 16 April 2018.

[3] TxDOT.gov,. (2015). Labor Day 2015 - Drink. Drive. Go to Jail.. Retrieved 16 September 2015, from http://www.txdot.gov/driver/sober-safe/labor-day.html.

[4] Madd.org,. (2015). MADD - Texas. Retrieved 16 September 2015, from http://www.madd.org/drunk-driving/state-stats/Texas.html

[5] Hoffman, D. (2002). HORIZONTAL GAZE NYSTAGMUS: THE SCIENCE & THE LAW. Retrieved from http://dree.mshp.dps.mo.gov/LotusQuickr/dre/Main.nsf/a7986fd2a9cd47090525670800167225/675228ff8f942a3686256bee007106fc

[6] Busloff, S. E. (1993). Can Your Eyes Be Used Against You--The Use of the Horizontal Gaze Nystagmus Test in the Courtroom. *J. Crim. L. & Criminology*, *84*, 203.

[7] Dorland, W.A.: *The American Illustrated Medical Dictionary* , ed 22, Philadelphia: W. B. Saunders Company, 1951.

[8] Edward B. Tenney, II. *The Horizontal Gaze Nystagmus Test and the Admissibility of Scientific Evidence*, 27 N.H. B.J. 179. 180 (1986)

[9] John Seelmeyer, *Nytagmus, A Valid DUI Test*, Law & Order, July 1985, at 29.

[10] Burns, M. (2007). *The robustness of the horizontal gaze nystagmus test* (No. DOT-HS-810-831). United States. National Highway Traffic Safety Administration.

[11] Weichert, F., Bachmann, D., Rudak, B., & Fisseler, D. (2013). Analysis of the accuracy and robustness of the leap motion controller. Sensors, 13(5), 6380-6393.

[12] Meaney, J. R. (1996). Horizontal Gaze Nystagmus: A Closer Look. *Jurimetrics*, 383-407.

[13] Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research?. *Educational researcher*, *41*(1), 16-25.

[14] Hoadley, C. P. (2002, January). Creating context: Design-based research in creating and understanding CSCL. In *Proceedings of the conference on computer support for collaborative learning: Foundations for a CSCL community*(pp. 453-462). International Society of the Learning Sciences.

[15] G. Stahl, "Computer Support For Collaborative Learning: Foundations For a CSCL Community", in *CSCL 2002*, Boulder, CO, 2002.

[16] Koivisto, J. M., Haavisto, E., Niemi, H., Haho, P., Nylund, S., & Multisilta, J. (2018). Design principles for simulation games for learning clinical reasoning: A design-based research approach. *Nurse education today*, *60*, 114-120.

[17] Dornan, T., Hadfield, J., Brown, M., Boshuizen, H., & Scherpbier, A. (2005). How can medical students learn in a self-directed way in the clinical environment? Design-based research. *Medical education*, *39*(4), 356-364.

[18] McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical education*, *44*(1), 50-63.

[19] Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research?. *Educational researcher*, *41*(1), 16-25.