THE PROMISE AND PITFALLS OF EXTENDED REALITY IN NAVAL TRAINING

Extended reality (XR) is a broad term that encompasses virtual reality, augmented reality, and mixed/merged reality. Virtual reality is a completely immersive experience for the user—all generated by computer. Augmented reality adds information to the real environment, such as a drawing that appears in the user's real-world field of vision. Mixed/merged reality allows for virtual and real worlds to interact—virtual characters can respond to gravity, and real-world objects grasped by the user can become part of the virtual scene. While low-cost XR initially focused on gaming, today's XR is also used in art, healthcare, manufacturing, and education. In a naval context, XR is often used in training.

The Naval Education and Training Command (NETC) is the largest shore command in the Navy, responsible for overseeing the Navy's force development. This includes recruiting sailors and conducting their initial and technical training. NETC continuously evaluates the effectiveness of naval training. These evaluations have traditionally used the results of students' quizzes, tests, and fleet feedback surveys, often supplemented with instructors' appraisals of how well students work with hands-on procedures using real equipment.

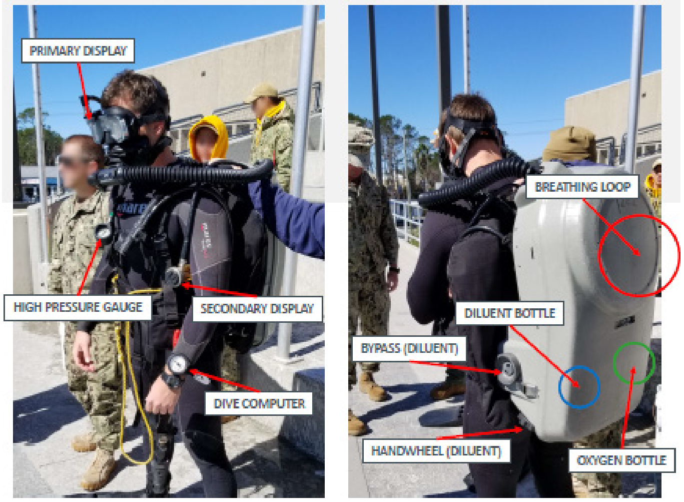
Watching students using real equipment is often more desirable for assessment than written tests, but there are many limitations to assessments with physical equipment. Sometimes equipment is not available or breaks if used too often. Stockpiling real equipment and supplies takes up classroom space, and assessing students individually is time-consuming. A class of 30 students requires instructors to watch students 30 times.

XR training devices, like written tests, allow the assessment of multiple students simultaneously, after or while they are receiving instruction. XR training provides similar economies of scale to written testing, but places more emphasis on application of knowledge, particularly for operating and maintaining equipment.

Despite the high potential of XR for assessing performance, three questions remain unanswered:

1. Can current XR systems provide reliable, useful performance metrics that students and instructors accept?
2. If current XR systems do not yet provide acceptable metrics, what obstacles are hindering the development of this capability?
3. How can the Navy foster the development of acceptable XR performance metrics?

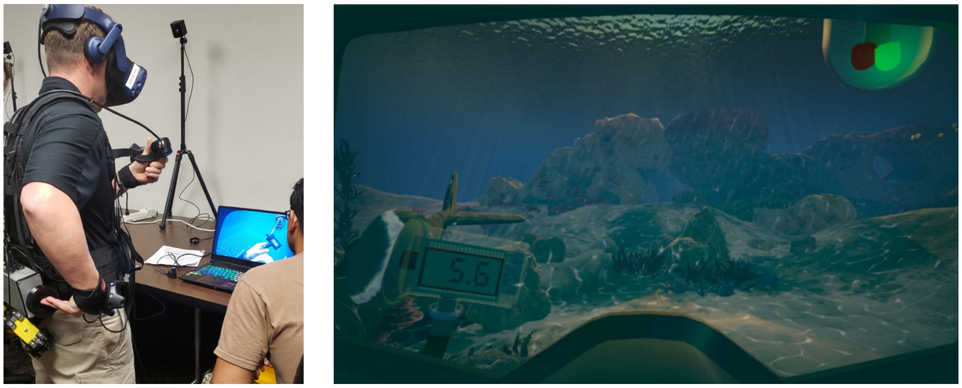
NETC asked CNA to address these questions and make a "roadmap" of recommendations for the future. To assess these issues, CNA examined a high-quality prototype XR trainer that instructs underwater divers on emergency procedures (EPs) and collects data on their EP performance. We observed divers using the XR trainer, interviewed instructors, performed a literature review, conducted discussions with XR subject matter experts across the Department of Defense, and analyzed the data collected by the XR device. Classes, varying in size from 2 to 20 students, were given a two-hour session to use the prototype.



The underwater breathing apparatus.

The XR trainer we examined, the underwater breathing apparatus simulator (UBASIM), is a mixed/merged reality device. It provides more immersive, realistic indicators of emergency situations than can be simulated in a pool. Using replicas of real equipment with embedded sensors, UBASIM allows divers to develop "muscle memory" of appropriate EP responses. Divers can physically turn dials and valves, getting realistic feedback from their actions. They can learn to recognize emergency situations through consulting simulated sensor readings, and then see the result of their actions in real time.

For example, the diver can look at the sensor-laden secondary display through the goggles to help diagnose the situation. If the diver responds correctly, sensor readings will indicate that their actions are improving the situation. Conversely, inappropriate actions can cause worsening sensor readings. Wrong actions can even lead to simulated symptoms of oxygen toxicity, such as tunnel vision, ringing in the ears, or unconsciousness—represented by images and sounds on the screen and headphones. (See photos.) In other words, the XR prototype provides a more dynamic training and assessment environment. Previously, students memorized a sequence of EP steps, were quizzed on them, and then were asked to demonstrate the procedures in a pool when shown a particular emergency situation on a card.



The image on the left shows a student raising his left wrist to get a reading that shows in the virtual scene in the goggles on the right. On the left, you can see a dial that he could use to control air flow.

UBASIM has three modes:

Crawl mode allows divers to choose which EP they want to learn. It provides step-by-step instruction, and no performance data are collected.

Walk mode is more difficult because although the student can choose the EP, the student's responses must first show that they recognize and diagnose which emergency situation has occurred, given visual inputs from the goggle device and audio cues from headphones. Hints in Walk mode are less specific than in Crawl mode, and Walk mode collects real-time metrics on diver performance.

Run mode is the most difficult of all, because an intelligent tutoring system selects which EP the student will be presented and there are no hints or instruction. Students must diagnose the emergency by themselves. Performance data are collected, and students receive an after-action report that provides remediation.

UBASIM trains and collects performance data on 21 separate diving emergency procedures. UBASIM's developers recommend that students go through all the EPs in Crawl mode, then all of EPs in Walk mode before attempting Run mode. Performance data collected include:

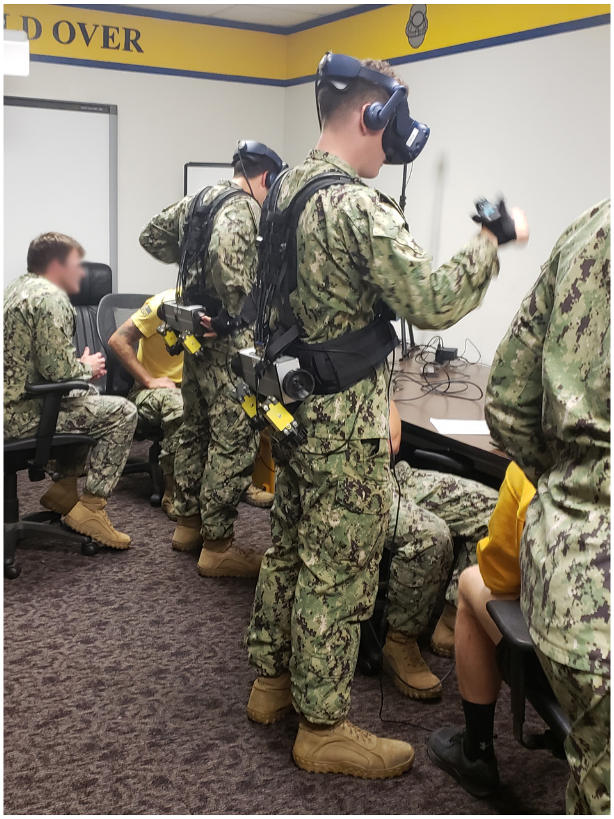
(1) whether steps were performed

(2) when they were performed

(3) whether the diver made life-threatening safety errors

Unfortunately, the XR prototype captured very low student performance levels, with divers passing fewer than 20 percent of Walk and Run mode sessions. This prototype probably was not always recording actual performance reliably and accurately for several reasons: There were several sensor difficulties, instructor-taught order of steps sometimes differed from those assumed by the prototype XR trainer, and the system had difficulties with steps performed simultaneously. The developer is working to address these issues.

These difficulties were magnified by circumstances not directly related to the technology. UBASIM was not part of the required curriculum; this was a pilot test of a prototype, designed to help the developers improve the trainer. Possibly as a result of limited time, students often rushed or skipped to Run mode—contrary to the developer's recommendation. Classroom space and power constraints reduced the number of students who could use devices at the same time. Lastly, instructors told us that COVID-19 curtailed the amount of time that technical representatives could visit the site, making it more difficult to address technical problems quickly.



Student divers using UBASIM in the classroom.

Despite these challenges, the XR trainer did provide considerable training practice and could potentially assist instructors to efficiently identify both the EPs and the EP steps that give students the most trouble. With data from time stamps and sensors, XR might determine why a trainee had difficulty with particular steps, either by itself or in tandem with an instructor. An instructor will be better able to identify and interpret certain factors, including human responses to visual cues and common pitfalls in student understanding.

As a reliable and acceptable assessment device, however, we concluded that current XR technology is not yet sufficiently advanced to gain widespread instructor acceptance. Though instructors gave positive reviews of XR's potential as a training tool, they expressed reservations about using the device for assessment. The four main obstacles to acceptance involve **(1) sensors**, **(2) step order**, **(3) logistical constraints**, and **(4) a need for frequent onsite technical help while the device is perfected**.

How can the Navy overcome these obstacles and foster development of XR assessment for skills? Based on our analyses, we developed five key takeaways for the development and fielding of XR systems for training and assessment. These could help inform NETC's guidance to learning centers, policies, training materials, and program of instructional quality assurance. Some may require further research:

1. Testing and evaluation for XR training must be robust. This ensures that the system reacts correctly to unexpected trainee actions (e.g., performing steps out of the usual sequence or simultaneously).
2. "Correct" procedures in the XR system should reflect the full range of possible "correct" real-life options. Otherwise, systems may inadvertently train people against effective behaviors, such as developing shortcuts that improve efficiency.
3. XR is best used for training and assessment where steps and actions are well-defined.
4. Onsite technical support is key to avoiding lost training opportunities.
5. Automatically collected training assessments from XR systems have huge potential. Data collection and analysis should be part of the program plan from the beginning, and should be adequately resourced.

PRINCPLES TO CONSIDER IN MOVING FORWARD WITH XR

1. ***Why use XR? If it's impossible, dangerous, or expensive to train any other way***
2. ***How to use XR? With sufficient resources, time, and opportunities to interact***
3. ***Where to use XR? Where required space and power are available***
4. ***When to use XR? Immediately before training on real equipment or deploying***
5. ***Which type of XR to use? The XR most appropriate for the purpose (see table)***

Each of the three types of XR—**virtual reality** (VR), **augmented reality** (AR), and **mixed/merged reality** (MR)—differs in the user experiences it offers, and each has distinct strengths and weaknesses. The suitability of a type of XR will depend in part on its purpose. The table below summarizes the capabilities of the three major types of XR to meet various training needs.

Training need to meet	VR	AR	MR
High degree of immersion in a virtual world	Yes	No	Yes
Blocking the physical world for sufficient immersion and standardization	Yes	No	No
Augmenting the physical world with images or information	No	Yes	Rarely
Real world physical experiences	No	Yes	Yes
Perceptual interactivity between the physical and virtual worlds	No	No	Yes
Manipulation and response from both the physical and virtual worlds	No	No	Yes

XR holds great potential for the Navy, but the Navy must take care to ensure it is used for the right purposes, in the right situations and under the right conditions.

ABOUT CNA

CNA is a nonprofit research and analysis organization dedicated to the safety and security of the nation. It operates the Institute for Public Research — which serves civilian government agencies — and the Center for Naval Analyses, the Department of the Navy's federally funded research and development center (FFRDC). CNA develops actionable solutions to complex problems of national importance. With nearly 700 scientists, analysts and professional staff, CNA takes a real-world approach to gathering data, working side-by-side with operators and decision-makers around the world. CNA's research portfolio includes global security and strategic competition, personnel and training analyses, homeland security, emergency management, criminal justice, public health, data management, systems analysis, naval operations and fleet and operational readiness.

For additional information, please contact: Anita Hattiangadi, Director, Business Development & Client Relations, Resources and Force Readiness | HATTIANA@cna.org

www.cna.org

DMM-2023-U-034596-Final