

# USING CRISIS ACTION PLANNING EXERCISES TO ASSESS PROGRAM LEARNING OUTCOMES IN SUPPORT OF OUTCOMES- BASED MILITARY EDUCATION

Commander Daniel Post (USN), PhD;  
and Amanda M. Rosen, PhD

**Abstract:** Crisis action planning exercises (CAPEXs) and simulations, as limited forms of wargaming, are increasingly being used in military academic settings to evaluate learning objectives at the individual and program level. While there are reasons to believe that these exercises may be useful tools for programmatic assessment, questions remain about how to determine the conditions under which they should be used for this purpose. This article explores the tradeoffs of using these tools to evaluate program effectiveness and offers a tool to assist assessment designers in deciding when and how to adopt a crisis simulation for program assessment. Using evidence from the CAPEX at the U.S. Naval War College as a case study, the authors argue that using simulations specifically for program assessment requires additional cautions and considerations.

**Keywords:** crisis action planning exercises, CAPEX, simulations, wargaming, program assessment, professional military education, PME, outcomes-based military education, OBME

Outcomes-based military education (OBME) is now a reality at

U.S.-based professional military education (PME) institutions. No longer can programs claim that their students learned based on showing what is on the syllabus; as Kristin Mulready-Stone, former chair of the Assessment Committee at the Naval War College, outlines, programs must demonstrate students have achieved the stated program learning outcomes (PLO).<sup>1</sup> The burden is now on institutions to conduct program-level assessment and provide evidence of student learning. This movement necessitates important changes, among them the need to find some way to engage in program-level assessment when the focus has traditionally been on course-level assessment. With many other demands limiting the time and energy of faculty, administrators, and students, academia must find high-quality, cost-effective, useful ways of conducting program assessment.

Capstone-style simulations represent a possible solution. Games, simulations, wargames, and crisis experiments are experiencing a renaissance as instructional and experiential activities in the PME classroom. Many professors of strategy and related topics use wargames in the classroom as a primary educational tool for their courses.<sup>2</sup> James D. Fielder of Colorado State University notes the immersive qualities of gaming and their ability to create an alternate reality. From Fielder's perspective, "great games are viscerally lived experiences that mimic the emotions and learning of real events."<sup>3</sup> This is, indeed, the most important claim of wargaming and crisis simulation enthusiasts. In their view, wargames and crisis simulations are particularly engaging, immersive, and emotional in a way that traditional classroom methods such as lectures and seminar discussions simply cannot be. Across political science and security studies, simulations have become a common-

<sup>1</sup> Kristin Mulready-Stone, "A New Form of Accountability in JPME: The Shift to Outcomes-Based Military Education," *Joint Forces Quarterly* 112, no. 1 (2024): 30–38.

<sup>2</sup> On the utility of gaming in the classroom, for example, see Victor Asal and Elizabeth L. Blake, "Creating Simulations for Political Science Education," *Journal of Political Science Education* 2, no. 1 (2006): 1–18, <https://doi.org/10.1080/15512160500484119>; Mark Harvey, James Fielder, and Ryan Gibb, eds., *Simulations and Games in the Political Science Classroom: Games without Frontiers* (New York: Routledge, 2022), <https://doi.org/10.4324/9781003144106>; and James Fielder, "Pedagogical Spotlight: Gaming in the Classroom," *Western: Newsletter of the Western Political Science Association* 12, no. 2 (2022).

<sup>3</sup> James D. Fielder, "Reflections on Teaching Wargame Design," *War on the Rocks*, 1 January 2020.

place alternative to traditional teaching modalities such as lecture and discussions, and research consistently shows that they have valuable impacts on learning.<sup>4</sup> As Princeton University's Center for International Security Studies (CISS) Strategic Education Initiative (SEI) points out: "Books, lectures, and discussions can teach a lot about strategic decision making, but even the best classroom cannot replicate the uncertainty, pressure, and friction that decision makers face in the real world."<sup>5</sup> The crisis simulations run by SEI are a prime example of how educators attempt to close the gap between theory and practice and give students practice in "making foreign policy decisions under conditions of strategic and bureaucratic uncertainty." They are particularly useful in teaching and practicing such procedures and skills, as they immerse students in the simulated environment and give participants a chance to "feel" the pressures and intricacies of real-world decision making and crisis action environments.<sup>6</sup> Moreover, they have been found to lead to longer-lasting learning than more traditional approaches such as lecture.<sup>7</sup>

Gaming and crisis simulations are also used as educational tools in numerous other fields such as business, law, and management. For example, Deloitte, a leading professional consulting and advisory firm, utilizes crisis simulations for crisis management training in a variety of business settings.<sup>8</sup> Law professor Shawn Marie Boyne writes about using crisis simulations to enhance decision-making

<sup>4</sup> See Michael K. Baranowski and Kimberly A. Weir, "Political Simulations: What We Know, What We Think We Know, and What We Still Need to Know," *Journal of Political Science Education* 11, no. 4 (2015): 391–403, <https://doi.org/10.1080/15512169.2015.1065748>; and Amanda M. Rosen and Lisa Kerr, "Wargaming for Learning: How Educational Gaming Supports Student Learning and Perspectives," *Journal of Political Science Education* 20, no. 2 (2024): 318–35, <https://doi.org/10.1080/15512169.2024.2304769>.

<sup>5</sup> For information on the Strategic Education Initiative and their crisis simulations, see "Simulations," CISS, Princeton University, accessed 26 September 2024.

<sup>6</sup> John R. Emery, "Moral Choices without Moral Language: 1950s Political-Military Wargaming at the RAND Corporation," *Texas National Security Review* 4, no. 4 (2021): 11–31, <http://dx.doi.org/10.26153/tsw/17528>.

<sup>7</sup> Adam Wunische, "Lecture versus Simulation: Testing the Long-Term Effects," *Journal of Political Science Education* 15, no. 1 (2018): 37–48, <https://doi.org/10.1080/15512169.2018.1492416>.

<sup>8</sup> "Perspectives: The Deloitte Perspective," Deloitte, accessed 26 September 2024; and "Making Crisis Simulations Matter," Deloitte, accessed 26 September 2024.

skills in legal settings.<sup>9</sup> Public relations professor Karen Olsen uses simulations to teach communication skills in crises.<sup>10</sup> In each case, the simulated reality of the crisis situation is said to uniquely engage the students and to enhance student learning. Across PME institutions, educational wargaming and table-top exercises are experiencing a renaissance, becoming a mainstream part of the student instructional experience.<sup>11</sup> In some cases, courses are being taught entirely on educational wargaming design, such as the ones at the Naval Academy and Georgetown University, which center on student teams researching, designing, developing, and play-testing an original educational wargame on a topic related to military strategy.<sup>12</sup> Crisis action planning exercises (CAPEX), therefore, have a rich history of use and provide a potential way forward for OBME.

What is newer and understudied is the use of simulations for assessment, rather than instructional or experiential purposes. In their typology of simulations, Nina Kollars and Amanda Rosen divide simulations by their overall purpose: formative or summative. While most simulations tend to be formative—that, is, aimed at developing student knowledge and skills—they can also be used as a summative exercise that evaluates student performance or

<sup>9</sup> Shawn Marie Boyne, "Crisis in the Classroom: Using Simulations to Enhance Decision-Making Skills," *Journal of Legal Education* 62, no. 2 (2012): 311.

<sup>10</sup> K. S. Olson, "Making It Real: Using a Collaborative Simulation to Teach Crisis Communications," *Journal on Excellence in College Teaching* 23, no. 2 (2012): 25–47.

<sup>11</sup> Erik Lin-Greenberg, Reid Pauly, and Jacquelyn Schneider, "Wargaming for Political Science Research," *SSRN Electronic Journal* (2021), <https://doi.org/10.2139/ssrn.3676665>. See p. 7 for a useful table displaying game characteristics. Some examples of recent work relying on wargame data include: Daniel R. Post, "On the Prospects of Escalating to Deescalate and Limiting Nuclear War: With a Focus on the U.S. Perspective" (PhD diss., Brown University, 2023), <https://doi.org/10.26300/x4jy-1j42>; Emery, "Moral Choices without Moral Language"; Reid Pauly, "Would U.S. Leaders Push the Button?: Wargames and the Sources of Nuclear Restraint," *International Security* 43, no. 2 (2018): 151–92; Jackie Schneider, "Cyber and Crisis Escalation: Insights from Wargaming" (unpublished paper, Naval War College, 2017); Erik Lin-Greenberg, "Wargame of Drones: Remotely Piloted Aircraft and Crisis Escalation," *SSRN Electronic Journal* (2020), <https://dx.doi.org/10.2139/ssrn.3288988>; and "The Project on Nuclear Gaming (PoNG)," University of California-Berkeley, accessed 19 November 2024.

<sup>12</sup> Sebastian Bae, "GU Wargaming Society," Basics of Wargaming Course, SEST 560-01, Center for Security Studies, Georgetown University, accessed 28 September 2024.

abilities.<sup>13</sup> While not uncommon in businesses or military training, it is rare to see a simulation used for course assessment in social science education—and even more so for program-level assessment.<sup>14</sup> With little written about the utility of these types of exercises as tools specifically for evaluation and assessment of programs, there is a clear gap that the push for OBME requires us to fill.<sup>15</sup>

To address this gap, this article outlines an initial framework for evaluating when a CAPEX-style simulation is an appropriate tool for program assessment in PME. To begin, the authors analyze the advantages and disadvantages of using simulations for assessment. They then offer a five-question decision tool to guide decision makers as they consider adding a capstone simulation to their institution. The authors then apply the tool to the experience of the U.S. Naval War College in its use of a CAPEX from 2022 to 2024, ultimately concluding that the advantages of this style of program assessment are currently outweighed by the challenges, and that future iterations of the CAPEX need to either improve their alignment to the assessment purpose or to instead refocus entirely on the instructional and experiential benefits of using a simulation as a formative capstone rather than a program assessment tool.

### Advantages and Challenges of Crisis-Simulation-Style Assessments

Institutions considering adopting a CAPEX-style simulation for program assessment should carefully weigh several advantages and challenges (table 1). While these activities can offer high authen-

<sup>13</sup> Nina Kollars and Amanda Rosen, "Simulations as Active Assessment?: Typologizing by Purpose and Source," *Journal of Political Science Education* 9, no. 2 (2013): 144–56, <https://doi.org/10.1080/15512169.2013.770983>.

<sup>14</sup> One example of a simulation being used at the course level is the Capstone Planning Exercise for the Naval War College's Joint Maritime Operations (JMO) Course. For information about the course, see "Joint Maritime Operations," USNWC.edu, accessed 19 November 2024.

<sup>15</sup> By program, the authors refer to master's degree programs or entire Joint professional military education (JPME) curricula, or similar programs, that span multiple courses all geared toward specific qualifications or designations. Programs may be considered academic structures that have the following characteristics: they offer a consistent set of experiences, such as a set of mandatory core courses; they require students to engage with the set of experiences over an extended period of time (usually multiple semesters); and they are intentionally structured to achieve some outcome (or a set of multiple outcomes). List reproduced from Keston H. Fulcher and Caroline O. Prendergast, *Improving Learning at Scale. A How-to Guide for Higher Education* (Sterling, VA: Stylus Publishing, 2021), 5.

**Table 1.** Advantages and disadvantages of conducting program assessment through CAPEX-style simulations

Advantages	Risks/challenges
<ul style="list-style-type: none"><li>• Potential for high authenticity</li><li>• Immersive environments thoroughly engage students</li><li>• Simulations can serve as a high-impact, capstone experience</li><li>• Exercises allow assessors to observe process and skills</li><li>• Reduce student anxiety</li></ul>	<ul style="list-style-type: none"><li>• Requires intensive resources to be effective</li><li>• Difficult to observe and assess the desired individual-level behaviors</li><li>• Challenges in designing a simulation that clearly aligns with the intent of assessment</li></ul>

Source: courtesy of the authors, adapted by MCUP.

ticity, engagement, capstone-like environments, and opportunities for assessing hard-to-see processes, all while reducing student anxiety; they also pose costs and risks, notably in time required, difficulty in recording observations, and design alignment. Each of these requires discussion.

**Advantages and Benefits  
of CAPEX-style Assessments**  
*Potential for High Authenticity*

When employing outcomes-based assessment, it is critical that skills are explored in an authentic manner.<sup>16</sup> As a Rand report commissioned by the Joint Chiefs of Staff highlights, “measuring student performance using authentic assessments—that is, assessments that simulate real-world applications of desired outcomes—is critical to the successful implementation of OBME.”<sup>17</sup> Crisis simulations enable students to assume the roles and positions (or those like them) for which they are being educated and trained to assume, and these “authentic assessments” have been shown to

<sup>16</sup> See *Outcomes-Based Military Education Procedures for Officer Professional Military Education*, Chairman of the Joint Chiefs of Staff Manual (CJCSM) 1810.01 (Washington, DC: Joint Chiefs of Staff, 2022), A-3.

<sup>17</sup> Paul W. Mayberry et al., *Making the Grade: Integration of Joint Professional Military Education and Talent Management in Developing Joint Officers* (Santa Monica, CA: Rand, 2021), <https://doi.org/10.7249/RR-A473-1>.

improve student performance and skills, particularly in leadership development.<sup>18</sup> This can be done in simulations in a way that is not possible during an exam or written assignment, as the interaction of other human beings and the dynamic environment more closely resemble real-world situations. Though crisis simulations are not the only way to achieve authenticity, many in the field regard them as the best method of emulating the real-world.

### *Immersive Environments Can Thoroughly Engage Students*

Engagement is a critical concern during an assessment. If students are checked out, they may not be demonstrating the full range of their knowledge, skills, attitudes, and behaviors, thus rendering the data collected less accurate. Assessments that last multiple hours need to keep student engagement high to increase the likelihood that the results are reliable; otherwise, students assessed at the end of the day may show differences in performance that are not due to their program.

Simulations are engaging in a way that more typical assessment forms are not. It is almost universally argued by wargamers, and generally accepted by most who have experience in gaming and simulations, that these activities are more engaging, interesting, and memorable than traditional classroom activities such as listening to lectures or participating in seminar discussions.<sup>19</sup> There are several reasons for this. First, simulations and games give students a chance to role-play and empathize in a way that other activities do not offer. In effect, they get a chance to wear the shoes of the relevant decision-maker, advisor, or politician. This engages

<sup>18</sup> See, for example, Zahra Sokhanvar, Keyvan Salehi, and Fatemeh Sokhanvar, "Advantages of Authentic Assessment for Improving the Learning Experience and Employability Skills of Higher Education Students: A Systematic Literature Review," *Studies in Educational Evaluation* 70 (2021): 101030, <https://doi.org/10.1016/j.stueduc.2021.101030>; and Anna Wiewiora and Anetta Kowalkiewicz, "The Role of Authentic Assessment in Developing Authentic Leadership Identity and Competencies," *Assessment & Evaluation in Higher Education* 44, no. 3 (2018): 415–30, <https://doi.org/10.1080/02602938.2018.1516730>.

<sup>19</sup> See the discussion in Amanda M. Rosen, "The Value of Games and Simulations in the Social Sciences," in *Learning from Each Other: Refining the Practice of Teaching in Higher Education*, ed. Michele Lee Kozimor-King and Jeffrey Chin (Berkeley: University of California Press, 2018), 215–27, <https://doi.org/10.1525/9780520969032-018>.

the emotions of the participants. As John Emery highlights from an in-depth study of wargaming at Rand, “The capacity for empathy in wargaming comes from being made to feel the weight of decision-making and exercising ethical practical judgment in a simulated environment with a high degree of realism rather than abstraction.”<sup>20</sup>

Second, simulations, especially when time constraints are involved, may conjure real stress in the players and participants that, as James Fielder points out “when overcome, reinforces learning through practice and fosters trust amongst the players.”<sup>21</sup> In the authors’ participation in crisis simulations, they have seen first-hand the immersion players often experience and how when games and simulations wrap up, players are often exhausted, exhilarated, and sometimes feel as though they are coming up from out of a cave or some other alternate reality. For something as important as program assessment, deep student engagement and emotional involvement is certainly a plus and can contribute to stronger demonstrations of program learning outcomes.

### *Simulations Can Serve as a High-impact, Capstone Experience for Assessment*

An end-of-program simulation can serve as a collaborative, capstone experience for students that is valuable for their learning. The American Association of Colleges and Universities notes that such events are “high-impact” practices that can lead to lasting learning for students; that finding is supported by extensive research.<sup>22</sup> These findings focus more on the instructional and experiential benefits of simulations, but there is every reason to expect that those benefits remain even if the purpose of the experience is largely one aimed at assessment. So long as designers balance the demands of creating an immersive simulation with those of an assessment activity, there can be great value in crafting an end-of-program capstone simulation that doubles as a program assess-

<sup>20</sup> Emery, “Moral Choices Without Moral Language.”

<sup>21</sup> Fielder, “Reflections on Teaching Wargame Design.”

<sup>22</sup> Cindy Kilgo, Jessica Ezell Sheets, and Ernest Pascarella, “The Link between High-impact Practices and Student Learning: Some Longitudinal Evidence,” *Higher Education* 69, no. 4 (2015): 509–25, <https://doi.org/10.1007/s10734-014-9788-z>



ment. This is particularly valuable in programs that have disparate courses where it is not overtly obvious to students how the courses connect to each other. Creating a capstone experience can help students synthesize what they learned from the year, even as assessors gather to observe the learning that has already taken place.

### *Exercises Allow Assessors to Observe Process and Skills Rather than Just Knowledge*

Unlike exams or written assignments, crisis simulations allow assessors to observe the communications and processes taking place as part of the decision-making or other task performance.<sup>23</sup> This makes it possible to study and observe skills related to those processes such as decision-making structures, organizational behaviors, type and frequency of communication, leadership, or any other process-oriented outcome they wish to observe. As an example, in programs designed to enhance strategic decision making in a national security context, assessors will be able to see how individuals and groups organize themselves; see and hear how they compare and contrast various options; observe the adoption of leadership roles and styles; and identify questions and issue areas that arise in discussions—all of which would be much more difficult to observe in other types of assignments. Crisis simulations are most beneficial when the program learning outcome (PLO) under consideration is of the nature of these process-oriented types of skills. As Ellie Bartels, a Rand policy researcher, suggests that “war-games work best when used to explore a problem involving human decision-making in conflict and generate new potential solutions. That makes wargames particularly powerful early in decision-making processes when the nature of a problem is still unclear, and where wargames can suggest new frames or approaches to guide subsequent analysis.”<sup>24</sup>

Likewise, assessors will be able to observe and track the important considerations, questions and factors that arise in the group discussions and processes. This is beneficial because in other tasks, such as exams and written assignments, students are less

<sup>23</sup> Elizabeth Bartels, “Getting the Most Out of Your Wargame: Practical Advice for Decision-Makers,” *War on the Rocks*, 19 November 2019.

<sup>24</sup> Bartels, “Getting the Most Out of Your Wargame.”

likely to address questions or factors not specifically assigned to them as part of the assessment. Such tests tend to focus more on knowledge gained rather than skills or reasoning. In crisis simulations, however, it is possible to use the simulation to identify what exactly is important to the participants, what questions they would want to have answers to, and what issues arise that the players deem critical to the simulation. This can do at least two important things. First, it can shed light on the players thought processes and reasoning and demonstrate their level of mastery of program learning outcomes. Second, since the idea here is program assessment, students may self-identify gaps in knowledge, learning, or training that can be addressed by program designers and curriculum development teams for the next time a program is taught. This last benefit is critically important to programmatic assessment and OBME as it completes the cycle of assessment and enhances the feedback-assessment loop.<sup>25</sup>

### *Reduce Student Test Anxiety*

Finally, many students struggle with anxiety during their academic careers, especially when facing exams and when worrying about grades. One recent study by the National Institute of Health reported that more than 75 percent of the students surveyed were stressed out before an exam.<sup>26</sup> A study by the New York State School Boards Association reports that 28 percent of school psychologists reported that one-half or more of the students they counseled displayed adverse symptoms prior to state exams.<sup>27</sup> A study by the University of Chicago found direct links between anxiety, emotion, and achievement across the globe.<sup>28</sup> Although there is much debate about whether this anxiety actually impairs retrieval of previously learned knowledge during exams, the anxiety is well

<sup>25</sup> Mulready-Stone, "A New Form of Accountability in JPME."

<sup>26</sup> SreeRam Thiriveedhi et al., "A Study on the Assessment of Anxiety and Its Effects on Students Taking the National Eligibility cum Entrance Test for Undergraduates (NEET-UG) 2020," *Cureus* 15, no. 8 (2023), <https://doi.org/10.7759/cureus.44240>.

<sup>27</sup> Paul Heiser et al., "Anxious for Success: High Anxiety in New York's Schools," New York State School Boards Association (NYSSBA) and New York Association of School Psychologists (NYASP), 2015.

<sup>28</sup> Carla Reiter, "Anxiety Affects Test Scores Even among Students Who Excel at Math," *UChicago News*, 10 March 2017.

documented and in many cases has been shown to be detrimental to performance.<sup>29</sup>

Even if adult learners in PME institutions may have less trouble with test anxiety, it is still desirable to create less performance anxiety during assessments. Simulation-based assessment is a group activity that eliminates the traditional quiet, individualized test-taking environment. Conducting assessment this way enables students to demonstrate their knowledge and skills through dialogue and allows instructors to dig deeper into student understanding.<sup>30</sup> Students then get the chance to explain their answers or discuss them with peers, often in an ungraded environment.<sup>31</sup> This can help reduce student anxiety as they will worry less about making silly mistakes or misunderstanding the task, and they can take some comfort in knowing that they will have ample opportunity to display their knowledge in a variety of ways.

### Disadvantages and Challenges of CAPEX-style Assessments *CAPEX-style Assessment Simulations Can Require Intensive Resources to Be Effective*

Simulations designed for instructional or experiential purposes can vary greatly in the time and resources they require. They can range from a short 10-minute exercise to a term-long immersive experi-

---

<sup>29</sup> Maria Theobald, Jasmin Breitwieser, and Garvin Brod, "Test Anxiety Does Not Predict Exam Performance When Knowledge Is Controlled For: Strong Evidence against the Interference Hypothesis of Test Anxiety," *Psychological Science* 33, no. 12 (2022): 2073–83, <https://doi.org/10.1177/09567976221119391>; John Jerrim, "Test Anxiety: Is It Associated with Performance in High-stakes Examinations?," *Oxford Review of Education* 49 no. 3 (2023): 321–41, <https://doi.org/10.1080/03054985.2022.2079616>. Jerrim, for example, finds no clear link between anxiety and performance among students age 15–16 years.

<sup>30</sup> Kollars and Rosen, "Simulations and Active Assessment," 153. Kollars and Rosen also point out that students are less likely to fail due to technicalities like poorly worded test questions or misunderstood directions.

<sup>31</sup> It is crucial to separate grading from assessment. Students might complete a graded assignment that is not used for program assessment; they may alternatively be asked to complete an assessment activity that does not result in a grade for a course. In some cases, an activity may be individually graded *and* used for assessment, but even in these cases, faculty and assessors will typically have different criteria and practices for these two different methods of evaluation. As Mulready-Stone, "A New Form of Accountability in JPME," shows, "grading is not outcomes assessment."

ence.<sup>32</sup> Assessment simulations, however, tend to be high-intensity experiences. First, they require all students (or a random sample who will not be angered by being asked to do more work than their peers) to participate, often at the end of an academic year when they are focusing on their next duty assignment. The simulation may require them to research roles or conduct other time-consuming, advanced preparation for the exercise. Second, they require a committee of faculty and administrators to design the exercise, create the materials, coordinate the logistics, sync the data with student information, and analyze the results. They also require extensive physical space to execute the exercise and virtual space to store the data. Finally, high numbers of faculty must pitch in to act as assessors at a time when many are grading final papers and projects, devoting time to training and observing the simulation. All of this can be extremely labor intensive, and such costs must be considered when adopting an assessment simulation.

### *Difficult to Observe and Assess the Desired Individual-level Behaviors*

Tests excel in identifying individual student performance. In group simulations, however, it can be very difficult to accurately observe and assess individuals. There are several issues here. First, there is the risk that some individuals will not participate extensively in the simulation. In any group setting, it is possible for certain members of the group to perform more dominantly than others. In simulations and exercises where group decision making and teamwork are expected, it will be possible for some members of the group to fade into the background—either because of incentives to free ride or because of the nature of the exercise. Team moves or decisions (the outcomes of gameplay and crisis action planning) that are the output of group work may not give insight to individual performance, knowledge, or skills.

This is compounded by the fact that even if members participate fulsomely, they may not demonstrate the specific learning outcomes desired. As an example, Kate Kuehn in a study of war-gaming assessments observed that “often, faculty formative feed-

---

<sup>32</sup> Rebecca A. Glazier, “Running Simulations without Ruining Your Life: Simple Ways to Incorporate Active Learning into Your Teaching,” *Journal of Political Science Education* 7, no. 4 (2011): 375–93, <https://doi.org/10.1080/15512169.2011.615188>.

back would focus on how well an individual contributed to their team rather than on mastery of particular knowledge and skills.”<sup>33</sup> It may be sufficient to observe group outcomes when assessing a program learning outcome, but if information is desired on individual performance, assessors must observe this directly from everyone. As Kuehn further elaborates, “If evaluating higher order thinking such as decision making at the individual level, one must see the thinking process of each participant or else make a contentious assumption that the final team decision and observed team conversation reflects each individual’s thinking skills.”<sup>34</sup> When participation rates are low that necessarily makes assessment more challenging.

For students who do participate, their motivations may lead them to exhibit behaviors that interfere with program assessment. Specifically, there is a risk with crisis simulations and wargames that students focus too much on winning and not enough on the outcomes or learning objectives at hand. Students of all types care about their grades and most people would also rather “win” than “lose.” Fielder explicitly highlights this in his work and stresses that the focus is on learning outcomes not winning.<sup>35</sup> Designers will need to be careful not to encourage the focus to shift toward winning and away from demonstrating knowledge, skills, or proficiency.

There is also a difficulty in observing and recording information and data in dynamic crisis simulation settings. Without robust tools to record the actions, words, ideas, and thoughts (those expressed out loud) or the ability to video and audio record the entire game/simulation, capturing relevant data is exceedingly difficult. Not only must the data be noticed and observed but it must also be accurately recorded in real time. It can be incredibly challenging to listen in on every sidebar and conversation, particularly for an observer who is supposed to keep their presence nonintrusive. This is something other assessment tools have an advantage in, as in the case of tests and writing assignments, where the students give you this data in the form of answers to questions or their thoughts writ-

<sup>33</sup> Kate Kuehn, “Assessment Strategies for Educational Wargames,” *Journal of Advanced Military Studies* 12, no. 2 (2021): 139–50, <https://doi.org/10.21140/mcu-j.20211202005>.

<sup>34</sup> Kuehn, “Assessment Strategies for Educational Wargames,” 148.

<sup>35</sup> Fielder, “Reflections on Teaching Wargame Design.”

ten in essays. Video and audio recording of simulations and games is not a feasible solution for all games all the time for a variety of reasons, including ethical and privacy concerns.

As a result, simulation assessments require extensive work to train assessors on what to look for, how to record it, and how to reliably apply rubrics for assessment and evaluation. The simulation and game must be designed with assessment in mind and the tools and materials must be built in to capture the data needed for quality assessment. Even when training and conditions are optimal, there will likely be many more cases of “not observed” in a simulation-style assessment.

### *It Is Challenging to Design a Simulation that Clearly Aligns with the Intent of Assessment*

The simulation design stage may not ensure that the varying demands on the exercise continue to align with assessment objectives. It is a challenging task to design crisis simulations and wargames involving groups of autonomous human beings that effectively emulate the real-world with all its complexity and nuance and at the same time drive the activity in a way that elucidates specific learning outcomes. This is why it takes years of practice and training to develop good wargames.<sup>36</sup> As Elizabeth Bartels highlights, “master designers throughout government and industry work for years to develop the knowledge needed to select the right approach for the problem at hand.”<sup>37</sup> Game and simulation materials, scenarios, and scripts must be carefully designed to yield the behaviors and outcomes that assessors are looking to see. If the scenario or materials are insufficient it is possible that the individuals taking part in the assessment may perform activities or solve problems unintended by the designers. If assessors must get involved during the scenario to correct this, it may bias the results and invalidate the assessment process by priming students to behave in a particular way.

For example, if an outcome under assessment focuses on ethical decision making, the scenario should set students up so they can demonstrate their abilities to engage in moral deliberation,

<sup>36</sup> John Compton, “The Obstacles on the Road to Better Analytical Wargaming,” *War on the Rocks*, 9 October 2019.

<sup>37</sup> Elizabeth Bartels, “Building a Pipeline of Wargaming Talent: A Two-track Solution,” *War on the Rocks*, 14 November 2018.

perspective taking, or their knowledge of ethical perspectives. If the scenario leans too much into something that is not being assessed—such as role-playing as specific actors in the National Security Council—the students may completely ignore the area being assessed. Evaluators then must decide whether to step in and prime the students to consider ethics—which will ruin the simulation as an assessment practice—or record a high number of “not observed” ratings. Those analyzing the data later will not know if ethical decision making was not observed because students have not learned to internalize those approaches, or because the simulation was designed in such a way as to lead students to focus on a completely different set of skills and knowledge.

The specific outcomes you are hoping for must drive everything from the design of the game to the assessment tools used to evaluate performance during and after execution. This is what Fielder refers to as the “Primacy of the Objective.”<sup>38</sup> But if being considered as a tool to evaluate an entire program, there is a clear risk that any specific scenario and simulation is asked to do too many things at once. It is likely that program learning outcomes are robust requirements and encompass broad, complex, high-level tasks/skills/knowledge. Single scenarios may not be able to capture all of these at once, and therefore as a programmatic assessment tool may need to be combined with other elements to get a complete picture.<sup>39</sup> Asking too much of the simulation is a potential hazard in utilizing these tools for programmatic assessment.

Another challenge is designing a scenario with the right balance of realism and abstraction. No simulation can avoid some level of abstraction, but too much can reduce the immersion, while too little can paralyze students with an excess of information and choice. In a JPME setting, there is a real need to maintain as much realism as possible, given the Joint Chiefs of Staff’s (JCS) desire to develop leaders who are better prepared for war and Joint warf-

<sup>38</sup> Fielder, “Reflections on Teaching Wargame Design.”

<sup>39</sup> *Outcomes-Based Military Education Procedures for Officer Professional Military Education* discusses a holistic and multifaceted approach to program level assessment, likely requiring multiple forms of assessment to cover the entire list of PLOs.

ighting.<sup>40</sup> As the JCS wrote, “the driving mindset behind our reforms must be that we are preparing for war.”<sup>41</sup> As practitioners will be the first to say, however, they cannot adequately simulate war in a classroom. Despite that fact, it may be possible, as discussed above, to stimulate emotional, moral, ethical, and intellectual engagement to such an extent that wargaming and crisis simulation may indeed become a visceral experience with lasting impact. This requires expert design and facilitation, and there is no guarantee that any given exercise will feel realistic enough to participants to fully demonstrate their skill and knowledge in matters of warfighting and strategic thinking. Wargames necessarily make important abstractions so they may focus on more important objectives and factors. This balancing act will require practice and fine-tuning and deserves consideration.

Lastly, evaluating student performance to assess learning outcomes is inherently difficult. Ultimately, the process relies on subjective assessments made by the observers, instructors, and/or evaluators. Since these events are explicitly not tests or written assignments with clear rubrics and direction, and since they are meant to assess skills learned throughout an entire program, evaluation and assessment may be complicated by a lack of standard rubrics or grading criteria. Someone must observe the game and interpret what they see. Assessors must be trained at least to some degree in how to do this task, but when evaluating entire programs, this is going to require many participants in the evaluator/assessor role. Standardizing their interpretations and understandings of what counts as good/bad or pass/fail performance will be challenging. Additionally, program-level outcomes, as mentioned before, are likely to be broad and complex, which adds to the difficulty of creating and implementing standardized and consistent assessment tools.

In summary, crisis action planning exercises may be useful in assessing program learning outcomes but there are numerous challenges to doing so. In the best case, the PLOs themselves will

---

<sup>40</sup> See *Outcomes-Based Military Education Procedures for Officer Professional Military Education*; and *Developing Today's Joint Officers for Tomorrow's Ways of War: The Joint Chiefs of Staff Vision and Guidance for Professional Military Education and Talent Management* (Washington, DC: Joint Chiefs of Staff, 2020).

<sup>41</sup> *Developing Today's Joint Officers for Tomorrow's Ways of War*, 6.



**Table 2.** Questions to ask when considering a crisis simulation or wargame for program assessment

Crisis simulation as assessment decision tool	
1	Is a CAPEX-style simulation a good fit for achieving program assessment, or are other methods available that would work better?
2	Do the specific PLOs lend themselves to a CAPEX-style simulation?
3	What challenges inherent to crisis simulation are likely to occur?
4	Are the resources available to overcome the likely challenges?
5	Will the information gained from a CAPEX be useful in informing curricular decisions?

Source: courtesy of the authors, adapted by MCUP.

be well suited to this type of assessment and experienced game designers will be able to craft scenarios that adequately solicit the desired behaviors and outputs as well as incorporate appropriate assessment tools. In the worst case, PLOs will not be amenable to crisis action planning exercises and will therefore be ill-suited to good scenario design and measurable outputs. If assessors push too hard to institute crisis action planning exercises in cases where they are not appropriate, or build exercises that are not well designed, it will be a huge waste of everyone’s time and will not contribute to the cycle of learning or a productive assessment-feedback loop. Some guidance is needed to determine whether the benefits are worth the costs.

### A Decision Tool for Adopting CAPEX-style Assessments

As a preliminary effort to provide guidelines for when a crisis action-planning exercise (CAPEX)-style simulation is most appropriate, the following list of questions serve as a tool to assist in decision making (table 2). These questions may be applied to any setting in which a programmatic assessment is required, and faculty are considering the use of a crisis simulation for this purpose.

First, is a CAPEX-style simulation a good fit for achieving as-

assessment objectives? It is possible that there are some learning outcomes that can only be assessed this way. This arguably is true if outcomes require group-based performance in time-critical tasks or in studying decision-making processes, for example. In addition, there may be other factors pushing for a capstone or war-gaming experience that could benefit from being combined with assessment. It is also possible that individual courses and existing experiences do not provide adequate program assessment opportunities, and some kind of new event is needed. In such cases, a CAPEX may prove a useful method of program assessment. If, however, existing courses or experiences provide assessment opportunities, or there is no available time in the academic year to insert a multiday event, CAPEX may be a bad fit.

Second, do the required PLOs lend themselves to a CAPEX-style simulation? If the PLOs are not well suited to the advantages of crisis simulations, the most important of which are listed above, alternative measures might be better. For example, PLOs that are purely knowledge based might be better assessed through an exam or portfolio analysis. If some PLOS are mapped to specific courses, it might be better to design course-level assessments, which can still include simulations.

Third, what challenges inherent to crisis simulations are likely to occur? Table 1 provides a good starting point but may not be inclusive, and every assessor should ask how likely it is that these risks and challenges will be detrimental to their assessment. These challenges might include monetary costs, time, risks of burnout, onboarding or detailing well-trained simulation designers who also understand assessment, logistics, student resistance, lack of faculty or administrator buy-in. The list of potential challenges is long, and an honest assessment of the institution's challenges is needed before investing the energy required in building a CAPEX assessment.

Fourth, are the resources available to overcome the challenges of running crisis simulations? Not all good ideas are implementable given resource constraints and if it is known early on that there will be limited time or resources available, alternate means should be selected. Institutions must be willing to invest time, money, administrative support, and faculty and student energy in the CAPEX; otherwise, it is doomed to failure as few will think it is a good use

of time. Resource-strapped institutions interested in the benefits of simulations might consider using them first as instructional activities, rather than as program assessments. If PLOs can be assessed in less costly and resource intensive manner, institutions will benefit from being able to devote that time and those resources elsewhere.

Last, will the information gained from a simulation be useful? The entire purpose of OBME is ensuring that outcomes are met, and if not, to adjust curriculum appropriately to ensure that they are. Assessment data should directly inform decisions about curriculum, and a CAPEX-style simulation is only valuable to the extent that it will be supported as a source of such data. If faculty or administrators are likely to be skeptical of the data, resources constrain the data that will be collected, or the institution lacks robust data management, analysis, and feedback mechanisms, then the data from a CAPEX is likely to wilt unused. Likewise, if assessors cannot design and train on effective rubrics that evaluate observed behavior, there will be little high-quality data available to inform decision making. In such cases, a lot of effort would be saved by designing a less resource-intensive assessment mechanism. Finally, an end-of-program CAPEX assessment alone cannot tell assessors whether any observed knowledge, skills, or behaviors is due to the program; it is possible that students either entered the program with these outcomes or developed them concurrently with their program (a particular risk with students who complete their JPME requirements on top of a day job). At a minimum, some kind of baseline-establishing pretest is needed to compare to CAPEX results to be able to use the data confidently as evidence of program effectiveness.<sup>42</sup>

These five questions serve as an initial tool for assessors to guide decision making on whether to adopt a CAPEX-style assessment simulation. To illustrate their use, the authors turn to some preliminary data from the U.S. Naval War College CAPEX beta tests from AY 2021–22 and AY 2022–23. In general, they found that the advantages of the CAPEX are currently outweighed by the challenges and recommend pursuing other avenues of assessment.

---

<sup>42</sup> The gold standard from a research perspective would also include random selection and control groups, but those are not generally possible in educational assessment.

### CAPEX-style Assessment at USNWC: A Case Study

In line with the United States' Chairman of the Joint Chiefs of Staff's (CJCS) mandate that all Joint professional military education (JPME) schoolhouses become certified as outcomes-based military education (OBME) institutions, the U.S. Naval War College adopted a CAPEX-style program assessment in 2022.<sup>43</sup> The Naval War College is a graduate-level institution and this CAPEX assessment was designed to assess two specific programs: the master's degree programs for defense and strategic studies (College of Naval Command and Staff) and for national security and strategic studies (College of Naval Warfare). The program-level outcomes and the assessment methods discussed in this section refer to the specific PLOs for these programs.

The format of the CAPEX is a tabletop crisis action planning simulation where groups (preexisting seminars of approximately 11–13 students) are provided a brief scenario and are asked to develop and provide potential response options to the national security advisor (NSA). Following the group exercise, a short, individual writing exercise is assigned, followed by a smaller group oral board (usually four students per group). The group work prompt consists of a brief description of the situation, which is set in the real world so that all information not provided can be drawn from their knowledge of the world as it exists today. This is followed by the fictional NSA's guidance for developing courses of action. The final product of this group work is a memorandum of no more than 750 words describing the potential options. The individual writing task is a 500-word product in which students further defend one of the three options presented in the group work memorandum. The oral panel is a series of eight questions that all four students answer, and they take turns with who answers the question first.

The purpose of this CAPEX is to assess existing PLOs that have been derived from the Joint learning areas (JLAs) designated by the CJCS.<sup>44</sup> All PLOs are meant to align with the JLAs in a significant way, and all PLOs are to be assessed as part of a full OBME

<sup>43</sup> See *Outcomes-Based Military Education Procedures for Officer Professional Military Education*.

<sup>44</sup> The Joint learning areas can be found in *Officer Professional Military Education Policy*, CJCSI 1800.01G (Washington, DC: Joint Chiefs of Staff, 2024).

Table 3. Naval War College program learning outcomes, 2023–24

College of Naval Command and Staff PLOs; intermediate-level course (ILC) JPME Phase I		College of Naval Warfare PLOs; senior-level course (SLC) JPME Phase II
PLO 1	Demonstrate Joint planning and warfighting ability in military operations and campaigns across the continuum of competition	Demonstrate Joint warfighting leadership when integrating the instruments of national power across the continuum of competition
PLO 2	Create theater and national military strategies designed for contemporary and future security environments	Create national security strategies designed for contemporary and future security environments
PLO 3	Apply the organizational and ethical concepts integral to the profession of arms to decision making in theater-level, Joint, and multinational operations	Apply the organizational and ethical concepts integral to the profession of arms to decision making in theater-level, Joint, and multinational operations
PLO 4	Apply theory, history, doctrine, and seapower through critical, strategic thought in professional, written communication	Apply theory, history, doctrine, and seapower through critical, structured thought in professional, written communication

Note: these are the latest versions of the PLOs that are under revision.  
Source: internal memorandum, “Requirements for NWC JPME-I/II Curriculum, as of 22 Oct 2023,” authors’ files.

certification process.<sup>45</sup> The PLOs for the Naval War College are listed in table 3. Each have been painstakingly crafted through a diverse and inclusive process that solicited input from multiple stakeholders throughout the institution and the broader PME community. They are designed to be broad and comprehensive, and therefore leave ample room for flexibility in when and how they

<sup>45</sup> Importantly, the Naval War College is still in the process of achieving full OBME certification and has met all required milestones to date. The CAPEX, as designed, is in development and does not represent the sum total of USNWC assessment efforts. It is only a part of NWC’s assessment plan.

are taught, practiced, and assessed. Additionally, each PLO is supported at the course level by course learning outcomes (CLOs) that achieve specific course goals designed to build and contribute to the overall PLOs for each program.

It is easy to see that these PLOs are quite comprehensive and demand that the Naval War College create programs that achieve challenging learning outcomes. The comprehensive nature of these outcomes is one reason why the Naval War College turned to a crisis simulation-type exercise to evaluate their program level outcomes. In conversations with those in charge of creating and executing the CAPEX, it was clear that an end-of-course simulation-type exercise was deemed necessary because the original PLOs (different than those listed above) required demonstrating knowledge and skills that could only be gained from completing the entire course of study. This meant that a continuous assessment process that measured PLOs at different points in the curriculum would not work. However, many in the college disagreed on whether an end-of-year simulation would be required, and the debate continues today. Using the assessment decision tool, the authors' conclusion is that this CAPEX as it is currently configured does not meet the assessment needs of the Naval War College.<sup>46</sup>

Overall, after action reports from the last round of CAPEX beta testing show that the exercise still requires extensive improvement to meet the needs of program assessment. The CAPEX demonstrates both the advantages of crisis simulations but also falls victim to some of the risks and challenges inherent in these types of exercises. Applying the five questions of the tool provides guidance as to how to move forward.

As to the first question in the tool—Is a crisis action planning exercise a good fit?—the authors believe the answer is mostly no. First, nothing in the CJCS instructions requires PME institutions to use this type of assessment. PME institutions are expressly given the freedom to decide when and how to assess PLOs if the institution effectively meets the OBME requirements. In other words, if the program is assessed by considering outcomes rather than content

---

<sup>46</sup> The authors have gathered some after action reports and engaged in numerous conversations with the designers and facilitators of the CAPEX. This evaluation of using CAPEX for program assessment is based on these reports, informal conversations, and analysis from participating as assessors.

and time spent (as previously used), the assessment committees are free to assess the program in a piecemeal fashion, throughout the individual courses for example, rather than relying on a single capstone event to assess PLOs. Additionally, the existing PLOs can be assessed during the constituent courses with appropriately targeted assignments, which is much less resource intensive and adds less to overall institutional workload than a stand-alone, end-of-program exercise. For example, PLO 4 aligns very well with assignments conducted in the Strategy and Policy Department and should be well suited to assessing these outcomes.<sup>47</sup> The PLOs are *not* of a nature that requires a student has completed the entire course prior to assessment of any single PLO, which negates any need for a specifically end-of-term assessment event. While ending the course of study with an experiential capstone does fit the Naval War College's overall educational ethos, structural factors impede CAPEX from being a true capstone experience. Unlike most U.S. PME institutions, NWC accepts off-cycle students who begin their program midway through the academic year. These students still participate in CAPEX, despite having not finished their program; for them, it is not a final capstone experience. In many ways, the desire to have an end-of-program capstone event as an experience for students became conflated with the need to conduct program assessment, but trying to achieve both purposes at once constrained the ability of the CAPEX to serve either purpose particularly well.

Additionally, there are alternative methods available to assess these PLOs. Within specific courses, research assignments, group exercises, Joint warfighting crisis simulations, strategic assessments and analysis, written assignments and other faculty graded or assessed events already exist in the core courses and departments that could be used to assess the existing PLOs. The primary rationale for an end-of-program CAPEX was to account for skills that required completion of all the core courses before assessment. This is no longer the case. As just one example, essay assignments in the strategy and policy courses require students to apply theory, history, doctrine, and seapower through critical, strategic thought to a series of case studies throughout the course. These assign-

---

<sup>47</sup> The current PLOs appear to have been better aligned with the individual core courses taught as part of the JPME curricula at the NWC. This may facilitate a shift away from the CAPEX requirement and is actively under consideration.

ments can serve as written communication in support of PLO 4, with perhaps very minor adjustments to drive more focus toward the maritime (seapower) components or cases.

The second question—Do the specific PLOs lend themselves to taking advantage of crisis simulation?—is more mixed. PLOs 1 and 3, which focus on Joint planning and decision making, respectively, do lend themselves to many of the advantages and reasons to use a crisis simulation. PLO 1 might best be assessed in a group setting in which multiple actors and capabilities are drawn on to create an operational plan that addresses many aspects across a continuum of conflict. Depending on what aspects of decision making the assessment team is focused on, PLO 3 may also be well-suited to a crisis action planning type event. However, both PLOs may also be addressed throughout different courses, particularly in the JMO course (PLO 1), and in the leadership in the profession of arms course (PLO 3) with existing or slightly modified assessment tools. PLOs 2 and 4 are exceptionally well adapted to assessment during existing course assignments in the theater/national security decision-making courses and the strategy and war/policy courses and do not consist of tasks that are best assessed through a simulation.

Regarding question three, a CAPEX-style event in this setting is likely to encounter almost all the listed challenges and risks, resulting in a resource intensive and highly complex process with the exception that there is almost no risk of students focusing too much on winning due to the specific design of this CAPEX. While this alone should not rule out a CAPEX-style event, the administrative burdens of conducting the exercise might. The NWC CAPEX faced issues with observing individual performances: during 2023–24, almost 50 percent received “not rated” entries from assessors. For some respondents, it was because they were off-cycle students or unable to attend the entire event, but free-riding cannot be eliminated as a factor, as many students were in the final stages of preparing to move to execute their next set of military orders.<sup>48</sup> The entire student body participated, spending a week of their time despite end-of-term pressures. This is a considerable investment of time to evaluate a single PLO, which should prompt considerations of whether these challenges can and should be overcome.

<sup>48</sup> *Crisis Action Planning Exercise (CAPEX)—June 2024: Preliminary Report* (Newport, RI: Office of Institutional Effectiveness, Naval War College, 2024).



Question four on the availability of resources is more difficult to answer. Leadership did invest in the CAPEX, but some costs are not monetary, notably those to the faculty who were asked to devote time to supporting CAPEX as moderators and assessors. This represents real opportunity costs due to devoting the necessary time and effort to designing these scenarios, simulations, grading rubrics, and faculty training. As one small piece of evidence, CAPEX required almost 100 fully qualified faculty members during three days to train, facilitate, and debrief the exercise during the 2023–24 academic year.

Question five asks whether the information gained from a CAPEX will be useful in informing curricular decisions, and so far at NWC the answer is largely no. Challenges in accurately observing and recording data on students has resulted in very high “not recorded” rates, making it difficult to glean useful insights from the data. Furthermore, the lack of any kind of baseline-establishing pretesting or control groups makes it difficult to determine whether behaviors that are observed are due to learning at NWC or representative of prior training and education. This renders the CAPEX data of limited utility in assessing whether those demonstrating mastery developed their skills and knowledge at NWC or elsewhere.

In addition, to date, the impact of CAPEX on curricular decision making has been limited. In the most recent report, 91 percent of students rated were deemed to have passed the mastery threshold for PLO 4.<sup>49</sup> While after action reports are plentiful, there is little evidence to suggest that the PLO analysis from CAPEX led to any kind of curricular revision. Instead, decision makers have changed the wording of PLOs and requested curricular innovations that respond to higher-level pressures, such as the creation of the new Perspectives on Modern War Course, an innovation that was requested by the president and provost at NWC in light of external pressures on modernizing the curriculum and internal desires to integrate the existing core curriculum and ensure students have time to discuss the remarks of distinguished visitors. At no time during the creation of this new two-credit course, which required stripping credit hours from existing courses, was CAPEX mentioned as a

---

<sup>49</sup> *Crisis Action Planning Exercise (CAPEX)—June 2024: Preliminary Report.*

source justifying the change.<sup>50</sup> If the only place CAPEX finds a role in NWC operations is to respond to accreditation reports to show that assessment is taking place, this suggests that the exercise has minimal value as an assessment tool, even if it serves as a potentially valuable capstone experience for students.

Overall, the result of this tool-based analysis is a healthy skepticism about the value of the current CAPEX at the Naval War College. This should not take away from the excellent and thoughtful work that many faculty and staff put into instituting the CAPEX to improve assessment at NWC. However, if these are indeed useful questions to ask, then an honest consideration should lead decision makers to seriously question the utility of an end-of-program CAPEX-style assessment. It may, however, be useful to also analyze CAPEX using the advantages and disadvantages outlined in the article to provide further insight into this case study of crisis simulation assessment.

NWC's CAPEX excelled in its authenticity, immersive engagement, and capstone setting. As designed, the crisis exercise created a planning environment where students used their existing knowledge and tools to work as teams to develop strategies to address a problem. They had to quickly assess the situation, develop courses of action, present their options, defend them in a written assignment, and then answer questions before an expert panel. All these activities are ones that military officers and their interagency counterparts can expect to do in their professional work. There was also a high degree of both student and faculty engagement. As an internal memorandum highlights, the assessment committee was "extremely pleased" to report that it was "quite clear" that faculty and students were "professionally engaged in the event."<sup>51</sup> The CAPEX has also yielded some positive data regarding the proficiency levels of those students who were assessed, demonstrating the value of having a capstone event to capture such data. In the senior-level course (SLC) CAPEX last year, for example, 90.67 percent of students observed received the top two box scores (proficient or exemplary) when assessed for PLO 1. Encouragingly, the

<sup>50</sup> The authors are one of two CAPEX directors and one of the members of the team that created the new course; this finding is based on their experiences in these roles.

<sup>51</sup> "Memorandum to Assessment Committee from Director of Institutional Effectiveness," dated 23 October 2023, authors' files, hereafter October 2023 memo.

assessors in the SLC course only reported a “not observed” response in 10.71 percent of students, a testament to faculty and student engagement and focus during the assessment.<sup>52</sup> In the case of the intermediate-level course (ILC), 63.34 percent of those observed were in the top two boxes for PLO 1C (a subset of PLO 1).<sup>53</sup>

Yet, challenges persisted. While simulations can allow assessors to observe skills and processes, the challenge in observing individual-level behaviors meant there was a high number of “not observed” ratings. In one iteration in the ILC, more than 78 percent of students were “not observed” in the sub-PLO outcome “demonstrate joint warfighting in military operations.” It is possible this sub-PLO may simply not be observable in a formal classroom setting. Further, even if some nonliteral interpretation of the sub-PLO is used, it was very challenging for assessors to observe individual behavior while students were working together in small groups. Whatever the reason, that only 22 percent of students had recorded observations of any rating suggests an issue in the design of the simulation.

Design alignment was also an issue. The after action reports noted a mismatch between the simulation materials (the situation, the scripts, prompts, etc.) and the PLO subcriteria defined on the rubrics that assessors used to evaluate student performance. Assessors reported “that certain scripts, prompts, and other material prepared before the exercise did not align well with some of the PLO sub-criteria defined on the rubrics.” This may be a primary source of the “not rated” observations mentioned above. This aptly demonstrates the challenge of crafting an appropriate scenario to capture the specific PLO of interest. Other CAPEX assessors (NWC faculty selected by their departments to participate) lamented that the assigned scenario in the SLC CAPEX did not actually consist of a “crisis” and so the option for students to “do nothing” was too attractive, essentially invalidating the entire purpose of a crisis action planning exercise. This further demonstrates the difficulty of crafting scenarios and scripts that allow for observable and assessable data.

Another challenge identified was that of data collection, for-

<sup>52</sup> October memo.

<sup>53</sup> Without an experiment with a control group that does not participate in CAPEX, and with no alternative capstone to measure, there is no way to know whether CAPEX reduced student anxiety compared to a nonexistent test.

matting, and analysis. Reports identified that “more robust” data would be required to “allow the type of analysis required for PLO assessment.”<sup>54</sup> These needs were mostly technical in nature, having to do with the scale used for proficiency measures (such as using a three- or four-point scale), conventions for assessor comments, consistent formatting for PLO assessment rubrics, and similar concerns. Additionally, since all students were evaluated by two assessors (who individually reported their own results), there was not enough data available yet to assess inter-rater reliability, something that also highlighted concerns about standardizing faculty understanding of the criterion measures and the assessed mastery level assigned. While training was provided to faculty, none of it centered on generating consistent interpretations of the rubrics, leaving faculty free to use their individual interpretations of what they saw.

Lastly, the assessment committee identified that the CAPEX in its current form may in fact be attempting to do too much with too few resources. The memorandum noted that the committee is investigating what “embedded course assessments/assignments can be used to gauge student PLO or sub-PLO mastery.” This is being considered to explicitly hedge against trying “to do too much; something the J7 has warned us about.”<sup>55</sup> With hundreds of faculty and all students involved at the end of the academic year with many other competing priorities, it may be that the investment of time and energy is too great, when there are possible lower-cost alternatives available to achieve the assessment purpose.

In sum, the authors’ experiences and what little formal evidence is available from early beta testing of the NWC CAPEX demonstrates that the challenges of producing a well-aligned, sufficiently resourced assessment capstone experience are great. The CAPEX designers and facilitators have done an admirable job designing an assessment tool that can do the job. They deserve accolades for their superb efforts to overcome the challenges inherent in this kind of tool and to benefit from the advantages of such tools. But if there are other ways to complete program assessment with much less cost in terms of faculty resources and time required, then those

---

<sup>54</sup> October 2023 memo.

<sup>55</sup> October 2023 memo.

require exploration, especially when the data has not yet shown itself to be of extensive use for curricular revision.

## Conclusion

This article argues that there is a lack of data and research focused on when crisis simulations and wargames are useful as *programmatic assessment tools*. There is little doubt from existing literature that CAPEX-style exercises are valuable as instructional and experiential tools, and that students get great value from participating in them. As capstones, they can excel. But combining an experiential capstone event that aims at increasing student learning with an assessment event that evaluates learning is a challenging task that requires analysis and study.

In this essay, the authors sketched out some preliminary answers to several questions related to this clear gap. They offer a simple framework for considering when and if a crisis simulation or CAPEX-style event is the most appropriate tool for assessment, and examine the key advantages and challenges of using crisis simulations as assessment tools. It is by no means clear if these lists are all inclusive or exhaustive. Future research should continue in this vein to identify the most important strengths and weaknesses of these tools for assessment purposes. This should include empirical research that aims to show the specific informational and data-analytic advantages to using these tools instead of more traditional assessment methods.

The authors also presented a new decision tool, asking decision makers to consider five key questions when examining CAPEX-style simulations for use as assessment vehicles. Future research should refine this list and explore simple but powerful frameworks for helping institutional assessment bodies make informed decisions about utilizing crisis simulations and wargame-type events. CAPEX-type events will not be the best option for all PLOs and will very rarely be the only option. With the current growth of crisis simulations and gaming techniques as pedagogical and research methodologies, it will be wise for those who use them to develop best practices and efficiencies to make the best use of the institution's limited and valuable resources.

The case study of the Naval War College experience from 2022

to 2024 demonstrates that more work is needed to design capstone assessment experiences that can maintain the benefits of simulations while meeting the challenges of resourcing, observing desired PLO outcomes, and aligning design with assessment needs. OBME requires innovation, and the NWC CAPEX exemplifies that approach. Its current inability to maximize advantage and minimize costs could be addressed through two paths: first, the assessment team, armed with the knowledge of the last two years, can revise the CAPEX to improve its resourcing, observational needs, and design alignment. Second, the college can look to noncapstone measures of assessing program learning outcomes. In the latter case, it may be possible to preserve the CAPEX as a valuable experiential learning capstone for the students. Removed from the need to assess students and PLOs, CAPEX would be restored to all the valuable gains from using simulations in the classroom established by the extensive literature on that subject. Regardless, the mismatch between experience and assessment tool must be resolved if CAPEX is to achieve its goals and bring its projected value to the college.

Between the analysis of advantages and disadvantages, the decision-making tool, and the experience of the Naval War College, the authors hope that other military institutions faced with meeting the needs of OBME are able to find a viable path forward as they consider how to construct valuable experiences that provide real benefits to the institution from the instructional, experiential, and assessment potential of CAPEX-style simulations.

---

### **About the Authors**

Cdr Daniel Post (USN), PhD, is a professor of strategy and policy at the U.S. Naval War College in Newport, RI. He received a BS in mathematics from the U.S. Naval Academy, an MA in national security and strategic studies from the Naval War College, an MA in political science from Brown University, and a PhD in international relations from Brown University. His research focuses on nuclear deterrence, escalation dynamics, limited nuclear war, and conflict termination. He is also the codirector of the Perspectives on Modern War Course at the Naval War College.  
<https://orcid.org/0009-0001-3107-3080>

Dr. Amanda M. Rosen is professor of teaching and learning and interim

director of the Writing and Teaching Excellence Centers at the U.S. Naval War College, where she conducts faculty development, research, curricular development, and assessment on educational gaming and instructional strategies in the security studies and professional military education contexts. She is the author of the 2024 book *Teaching Political Science: A Practical Guide for Instructors* and is the cofounder of the award-winning *Active Learning in Political Science* blog. Her work has appeared in the *Journal of Political Science Education*, *PS: Political Science & Politics*, *International Studies Perspectives*, *Politics & Policy*, and multiple edited volumes.

<https://orcid.org/0000-0003-1858-481X>

The views expressed in this article are solely those of the authors. They do not necessarily reflect the opinions of the U.S. Naval War College, Marine Corps University, the U.S. Marine Corps, the Department of the Navy, the Department of Defense, or the U.S. government.