Year 2024, Vol. 21, Nº 1, pp. 117 - 126. Complementary Edition
DOI: https://doi.org/10.35628/resup.v16i2.140

Wargaming as a discipline that constantly develops. How to ensure that new formats are improving the practice?

Natalia Zwarts
https://orcid.org/0009-0009-8322-0167
Is a lecturer in the Information and Communication Security Management Program at the Hague University. She has specialized in wargame design and experimental formats for education and research. Her previous appointments include synchronization of advanced technologies within NATO and civil-military cooperation. She is currently developing educational wargames aimed at improvement of cybersecurity.
The Hague University of Applied Sciences
Email:N.H.Zwarts@hhs.nl

Abstract: This article presents the features of wargaming, which can be used to support learning, analyzing and developing within organizations. Despite being an ancient discipline recognized by practitioners, it is often criticized for a potential bias in design, subjectivity of players and producing mostly anecdotal evidence in its results. To counter these issues, wargames can be assessed through the scale of reliability, comparing their data sources, number of iterations, player pool and scientific maturity. Once the assessment is complete, it is possible to recognize how trustworthy the conclusions from the wargame are. If there is a need to upgrade the existing design, the author proposes steps which will aid in increasing reliability.

Keywords: Wargaming, training, reliability, design, education.

1. INTRODUCTION

Wargaming is a tool capable of the three main functions: training, analysis, and development. (Bourguilleau, Wojtowicz, Lépinard, 2020). The paradox of wargaming is that it is both ancient and innovative. Matthew Caffrey assembled the most complete wargaming history representing its multiple facets. He proposed a division into four generations of wargames: from (1) training aids focused on survival skills to (2) abstract strategy games; (3) combat simulations for optimization, and (4) global proliferation through diverse organizations, as well as evolving methods (Caffrey, 2019). If history points to any direction, it is a growing catalogue of wargaming solutions.

If the method to the solution is organized according to scientific principles, wargaming can act as a proxy laboratory to measure effectiveness of chosen plans (Liu, Ding, Hu and Wang, 2023). Wargaming has the advantage of immersive synthetic environment allowing interaction between players, scenario and mechanics of the system (Wojtowicz, 2020). The presence of consequences and ability to measure the outcomes provides an analytical advantage over other investigation methods (Lin-Greenberg, Pauly and Schneider, 2021).

John Curry dubbed wargaming as "a flawed, but useful tool" (Curry, 2020). His argument for critical look at the results of structured wargaming are: possibility of discrepancies, incorrect modelling and lack of real-life evidence to support the collected outcomes. This initial critique leads us to the topic of this article: how to ensure that wargames are true? How can the accuracy of results be checked? How can commanders trust wargaming as a training tool?

The urgency of this topics is bound to the point of wargaming practice: it can only be trusted if it conveys correct data, conclusions and learning objectives. If it perpetuates fiction, or even falsehood, the practice becomes more dangerous than useful. Stephen Downes-Martin phrased it as multitude of developed methods, with a shortage of evidence and logic (Downes-Martin, 2015). In many ways, wargaming practice is much more prolific than wargaming as a scientific discipline, especially within the military education domain.

The Naval Postgraduate School (NPS) in Newport cultivates the wargaming tradition with 35 active wargames used for training across different formats and objectives (Route, 2016).

FIGURE 1
NPS students during the Wargaming Applications course, 2021.

Source: https://nps.edu/-/game-on-nps-wargaming-week-ties-tactics-strategy-to-improve-defense-planning [se accedió el 29.11.2023]

They can certainly be counted as experience for students challenged with critical thinking, planning and anticipating enemy actions. Despite proven track of learning, questions can be posed about their conclusions. How can they be compared and deemed as credible in their conclusions? How to ensure that students are developing relevant skills? How to systematically evaluate the effectiveness of wargames? The following article provides a classification of wargaming interventions alongside ways of validating the results of the specific design.

Evaluation of wargames contributes to the credibility of results. It takes away the need to justify using this method, instead providing evidence of valid effects. It provides a way to critically assess whether a wargame meets its purpose, and how reliable is it as a tool. The scope of this article is delineated by current use of wargaming (2020-2023) and categorization by sample size, validity, reliability. In order to recognize these points with applicable recommendations, the author presents the Maryland Scientific Methods Scale with its translation into wargaming practice.

2. THEORETICAL FRAMEWORK

In 1997, Congress commissioned a team of six Maryland faculty members: Lawrence Sherman, Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter and Shawn Bushway, to conduct a thorough evaluation of crime prevention programs in the United States leading to the publication of a Research Report "Preventing Crime: What Works, What Doesn't, What's Promising". The goal of this research was to develop a way to distinguish between trustworthy work and studies which can't be seen as credible. This task delivered one of the crucial tools to discuss the worth of the evidence delivered by given research.

Originally developed to measure the methodological quality of crime prevention studies, the Maryland Scientific Methods Scale, proposed 5 levels of assessment, with 1 being the most basic and least reliable and 5 being the most advanced and most reliable.

- Level 1: Correlation between a prevention program and a measure of crime at one point.
- Level 2: Measures of crime before and after the program, with no comparable control condition.
- Level 3: Measures of crime before and after the program in experimental and comparable control units, controlling for other variables that influence crime.
- Level 4: Measures of crime before and after the program in multiple experimental and control units, controlling for other variables that influence crime.
- Level 5: Random assignment of program and control conditions to units.

Authors of the scale took the Randomized Control Trials (RCTs) as their point of reference for excellence. RCTs are well-known from clinical trials that require isolating the effects of a given drug from other factors. This means that the researcher in charge looks closely at every patient ensuring they fit the requirements in terms of their status (for example not having other diseases than the one that is the subject of the study). Within RCTs, the population is divided into groups, with some of them receiving the treatment and some not. This allows researchers to separate the causes of the intervention from the placebo effect.

In case of wargames, often the implementation takes precedence over scientific analysis. The higher the need, the lower possibility of collecting data and setting up multiple groups. Some of the prescribed concepts could be included without

loss of time, for example with short surveys measuring how the knowledge of the players is changing.

The division and measurement give a definitive result in the effect of the treatment, including adverse symptoms and improvements from the baseline. The effects are measured for both positive change and negative reactions. This is a more detailed view into change of situation based on the data points.

The scale allows for classification of data collection, sample size, randomization and experiment design to recognize the realistic outlook on scientific efforts: always being the current perspective on the given problem. This limitation is also a solution: viewing the study through its position in the scale of reliability.

3. CONTEXT

In the last 10 years, science went through confidence crisis due to evidence of low replicability of previous studies. The critical question arose, mainly, whether evidence of experiment reported in reputable literature, is correct. The efforts to duplicate famous experiments came to negative results, disproving much of the knowledge recognized as standard in disciplines such as psychology, economy, and even law (Świątkowski and Dompnier, 2017).

Tools aimed at better revision of proclaimed results started to be proposed. The core of the problem can be summarized as "underpowered designs and exaggerated results" (Ferraro and Shukla, 2020). The following section focuses on the wargaming-specific considerations regarding increasing the accuracy of results and reliability of the studies.

Central Argumentation

Wargaming is susceptible to subjectivity due to high degree of human interaction. Many claims have been made about the high dependence of wargames on multiple factors: first, the narrative used to present its challenge, second mechanics that steer gameplay, third the materials used to build the game, and most importantly, the players involved in each iteration.

Wargames need to be viewed as a two-way street, also accounting for the influence of facilitators on the results. Given a set of the same scenario, the same players, and different facilitators, it is certain that there will be a change in the given feedback. The experience of the player is certainly influenced by the way in which instructions are given, rules are explained and objects are moved. It can be viewed as a strength of different expertise or as a problem due to consistently varying results.

The effects of a wargame are often changed by the perception of the player's experience - for example not enjoying the scenario or rules might translate into assessing the wargame as a bad tool. The key in terms of assessing wargames as scientific methods is looking into the levels of objectivity that can be achieved.

If we would attempt to provide a scale of reliability for wargaming, a simplified version could be proposed to allow for the comparison of different formats according to their credibility. As mentioned above it is not meant to diminish the value of experimentation that leads to discovery, but rather a scale showing advancement of reliability depending on the stage of development.

- Level 1: It works (individual iteration, expert opinions, and anecdotal evidence).
- Level 2: It was tested on multiple groups and results can be systematically analyzed.
- Level 3: There is a measurement of baseline values and a measurement of change after the wargame.
- Level 4: There are control groups, showing results with or without wargame.
- Level 5: There are multiple formats tested, compared, and analyzed. The most effective solution is chosen and data collection can be conducted over an extended period of time to inform real-life improvements.

In the drive towards objectivity, it must not be forgotten that the goal is not to reach a sterile environment. It is important to capture context, additional observations, and falsified hypotheses. It is the equivalent of finding out that the support system has equal importance to supplements in clinical trials. Within the pursuit of increased accuracy, studies should not be evaluated, but rather positioned at the level that represents their possible conclusions. It allows for experimentation on all levels and a comprehensive development within the discipline.

Analysis/Discussion

It is difficult to self-assess (Vatne, Guttelvik, Hennum and Malerus, 2022) the level to which the results can be trusted, but it is possible to recognize at which step of the scale the wargame is. To achieve that, indicators of differences and potential movement up the scale should be pointed out. The following section illustrates an example of increasing the scientific value of the conclusions through improvement of the experiment design.

How to apply the Wargaming Scale of Reliability to an existing design?

Wargaming has an extended practice, but a scarce scientific representation. New studies are proposing ways to improve the traditional flaws of less structured formats. For example, comparing existing designs and testing their accuracy to portray the reality of a specific battle (Burden, 2023). There is not yet a framework to systematically level up from base approaches to higher reliability. In the following table, steps proposed to build up towards the highest level of reliability in wargaming are proposed.

As an example, a singular edition of a wargame will likely have effect on training, analysis and development within an organization. But to be able to claim that it raises the level of knowledge about modern technology, the initial and later understanding to check if there is a progress. At this point, it might be shown that 60% of the participants do acquire knowledge during the wargame. To compare it with a lecture or other educational intervention would further demonstrate that wargame is actually the best solution to our problem of insufficient preparation to a given task. It could be extended to more groups to see whether a specific wargame works better and what makes it useful to the group.

Table 1. Rungs of the reliability scale.

Level	Existing design	Steps to level up
1: Initial design	It works (individual iteration, expert opinions, and anecdotal evidence).	• Planning a timeline for re-playing the wargame • Comparing insights of experts from policy and field • Evidence collected from different sources (literature, case studies, available statistical data).
2: Structured design	It was tested on multiple groups and results can be systematically analyzed.	• Structure that pinpoints how the results will be logged and compared • Planning multiple groups that provide feedback • Analytical approach to processing the results
3: Measurable effect design	There is a measurement of baseline values and a measurement of change after the wargame.	• Method to establish values before and after the wargame • Including measurements during the wargame to recognize the effects
4: Transferable design	There are control groups, showing results with or without wargame.	• Parallel wargames scheduled with intervention and control group.

5: Comparative advantage design	There are multiple formats tested, compared, and analyzed. The most effective solution is chosen and data collection can be conducted over an extended period of time to inform real-life improvements.	• Choice of multiple formats to perform structured analysis • Timeline allowing long-term data collection

Own elaboration.

4. CONCLUSIONS

Using wargaming as a tool brings many benefits from asking exploratory questions, through collection of the results, to having impact on reality. The introduction of human players supplies a significant part of the logic in the decision-making process, weighting the information according to its relevance. There are certain limitations that need to be accounted for to separate the mental model of the wargame designers and the reality of the situation (Wojtowicz, 2020b). As a counterpoint, the suspension of disbelief and the experimental environment can foster innovative ways of thinking, which would not arise without hypothetical scenarios. The key to development is to know what the wargame can achieve and what remains on the level of limited effect.

Wargaming cycles through periods of high activity, decline of practice, and the renaissance of returning interest (Bae and Brown, 2021). Working through the base levels of designing wargames and using them in the organization is a formidable step to establishing a long-term development. The scale of reliability presented in this chapter aims at growing the ambition to compare, extend the audiences of existing designs and raising the confidence in results.

References

Bae, S.J., & Brown, I.T., USMC (2021). Promise Unfulfilled: A Brief History of Educational Wargaming in the Marine Corps. Journal of Advanced Military Studies 12(2), 45-80. https://www.muse.jhu.edu/article/805918.

Bourguilleau, A., Wojtowicz, N., & Lépinard, P. (Déc 2020). Wargames for training future managers. Management et Datascience, 5(1). https://doi.org/10.36863/mds.a.14547.

Burden, David J.H.. "The Battles of Hue: Understanding Urban Conflicts

through Wargaming." Journal of Strategic Security 16, no. 3 (2023) :

128-140. DOI: https://doi.org/10.5038/1944-0472.16.3.2138

Available at: https://digitalcommons.usf.edu/jss/vol16/iss3/9

Caffrey, M. (2019). on Wargaming: How Wargames Have Shaped History and How They May Shape the Future, Naval War College, Newport Papers, 2019.

Curry, J. (2020). Professional Wargaming: A Flawed but Useful Tool. Simulation & Gaming, 51(5), 612-631. https://doi.org/10.1177/1046878120901852

Downes-Martin (2015). Wargaming as a Catalyst for Innovation. Connections 2015 Conference, National Defense University.

Ferraro, P., Shukla, P. (2020) Is a Replicability Crisis on the Horizon for Environmental and Resource Economics? Review of Environmental Economics and Policy, 14(2) 2022.

Lin-Greenberg, E., Pauly, R. B. C., & Schneider, J. G. (2022). Wargaming for International Relations research. European Journal of International Relations, 28(1), 83-109. https://doi.org/10.1177/13540661211064090

Liu, H., Ding, Y., Hu, H., Ou, H., Wang, J. (2023). Thinking of War Game Lab Construction. In: Long, S., Dhillon, B.S. (eds) Man-Machine-Environment System Engineering. MMESE 2023. Lecture Notes in Electrical Engineering, vol 1069. Springer, Singapore. https://doi.org/10.1007/978-981-99-4882-6_90

Route, R. (2016). Wargaming is Alive and Well at Naval Postgraduate School, The NPS Institutional Archive. Monterey, California.

Świątkowski, W. and Dompnier, B. (2017). Replicability Crisis in Social Psychology: Looking at the Past to Find New Pathways for the Future. International Review of Social Psychology, 30(1), 111–124, DOI: https://doi.org/10.5334/irsp.66

Wojtowicz, N. (2020a) Wargaming Experiences I: Soldiers, Scientists and Civilians. J10 Gaming, Delft.

Wojtowicz, N. (2020b) Wargaming of the Black Sea Security: Exploratory Study into the Strategies for the Region. Ukraine Analytica, 2020 (19).

Vatne, D., Guttelvik, M., Hennum, A.C., Malerud, S., Wargaming for the Purpose of Knowledge Development: Lessons Learned from Studying Allied Courses Of Action, Scandinavian Journal of Military Studies, 2022 (5), 297-308.