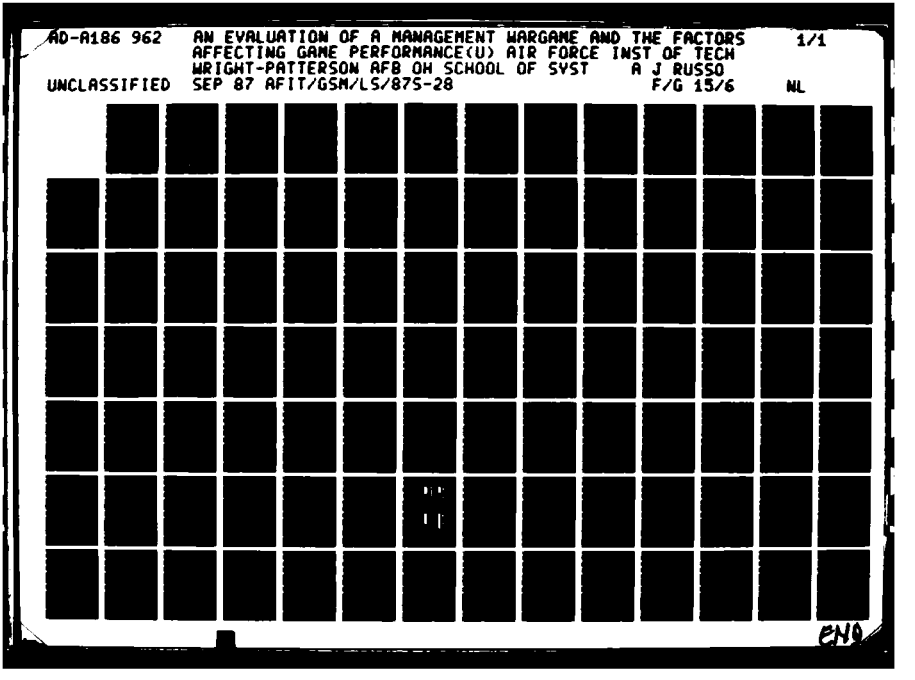
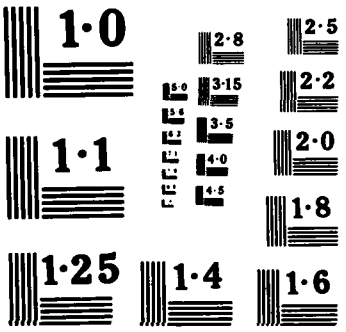

AD-A186 962 AN EVALUATION OF A MANAGEMENT WARGAME AND THE FACTORS AFFECTING GAME PERFORMANCE(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYST A J RUSSO
1/1
UNCLASSIFIED SEP 87 AFIT/GSM/LS/87S-28 F/G 15/6 NL
END


1·0
1·1
1·25
2·8
3·15
3·5
4·0
4·5
5·0
1·4
2·5
2·2
2·0
1·8
1·6

AN EVALUATION OF A MANAGEMENT WARGAME AND THE FACTORS AFFECTING GAME PERFORMANCE

THESIS

Anthony J. Russo
Captain, USAF

AFIT/GSM/LS/87S-28

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

2

AFIT/GSM/LS/87S-28

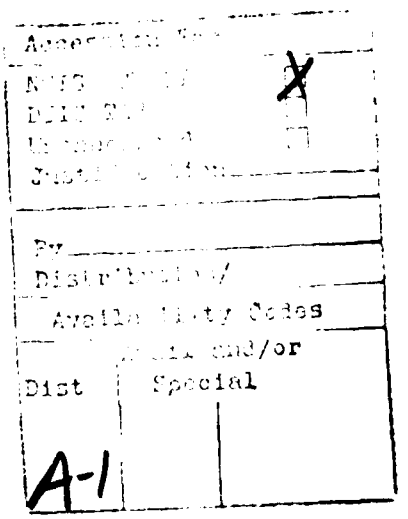
AN EVALUATION OF A MANAGEMENT WARGAME AND THE FACTORS AFFECTING GAME PERFORMANCE

THESIS

Anthony J. Russo
Captain, USAF

AFIT/GSM/LS/87S-28

The contents of the document are technically accurate, and no sensitive items, detrimental ideas, or deleterious information is contained therein. Furthermore, the views expressed in the document are those of the author and do not necessarily reflect the views of the School of Systems and Logistics, the Air University, the United States Air Force, or the Department of Defense.



AFIT/GSM/LS/87S-28

AN EVALUATION OF A MANAGEMENT WARGAME AND THE FACTORS AFFECTING GAME PERFORMANCE

THESIS

Presented to the Faculty of the School of Systems and Logistics of the Air Force Institute of Technology Air University In Partial Fulfillment of the Requirements for the Degree of Master of Science in Systems Management

Anthony J. Russo, B. S.

Captain, USAF

September 1987

Preface

The purpose of this thesis was to take one small step towards filling a serious gap in our knowledge of an important tool. The use of wargames in the development and training of officers is a long standing tradition in the military. With the advent of the computer, the use and development of wargames has increased at a dramatic pace. Wargames offer an inexpensive alternative to the escalating costs of other forms of training.

Yet, we know very little about how they teach, or even if they teach. Hopefully this thesis is one small step towards a better understanding of an important educational tool. Civilian researchers have recognized the value of these games at all levels of education, and have devoted serious academic attention to their development, use, and evaluation. This thesis is an attempt to demonstrate a systematic and quantitative method to evaluate a particular wargame, with the hopes that others will expand in this direction.

I would like to express my appreciation to my advisor, Dr Mauer, for his expertise and experience with TEMPO and with the research process in general. I especially appreciated the rapid turn-arounds, and the insightful comments which always improved the product, but never delayed it. In addition, I would like to thank Lt Col Dumond for his inspiration and support in the early stages, and all my TEMPO-AI playtesters, who made this thesis possible.

Anthony J. Russo

Table of Contents

List of Figures

List of Tables

Abstract

Despite the long association between military organizations and the instructional use of games, virtually no research has been done to validate the concept of a wargame. The objectives of this thesis were to: 1)identify a need for serious academic research on the validity of management wargames, 2)establish a methodology for the objective and quantitative analysis of of management wargames, and 3)apply that methodology to a specific management wargame.

The specific wargame evaluated in this thesis was TEMPO-AI, a computer version of the military force planning game used at Squadron Officer School resident program. The game was designed to provide the students in the correspondence course the same learning experience provided to those in residence.

This thesis evaluated the validity of TEMPO-AI as a learning instrument. TEMPO-AI was played over 60 times, and 25 students at the Air Force Institute of Technology (AFIT) played the game in a controlled experiment. The results of the experiment indicated that students that had nearly completed a Masters Degree program, scored significantly higher than those who were just entering the program. This indirectly supports the validity of TEMPO-AI since those students with greater exposure to advanced principles of

management would be expected to score higher in a game that rewarded correct managerial decisions.

Other factors, such as Technical Education, Wargame Experience, TEMPO Experience, Acquisition Experience, and Operational Experience were tested, but none were found to be significant at alpha = .05. In fact, those officers with more than two years operational experience scored considerably lower than the less experienced officers. However, a Test of Homogeneity revealed that a disproportionate number of the officers with Operational Experience were also among those just entering the program. This implies that the statistical correlation (-.25) is probably due to a lack of experience in some areas rather than any detrimental effect of operational experience.

In addition, the tests revealed numerous problems with the structure of the game, with the most serious of these being inadequate on-line documentation. To correct this and other problems, a number of revisions were proposed.

Perhaps an even greater impediment to implementing TEMPO-AI in a correspondence program is the reluctance of students to play the game on their own. A random sample of 54 AFIT officers yielded a 0.0 percent response rate, and test subjects were obtained only after financial incentives were offered. This is a cause for concern, since successful implementation in Squadron Officer School correspondence courses would depend on voluntary participation.

AN EVALUATION OF A MANAGEMENT WARGAME AND THE FACTORS AFFECTING GAME PERFORMANCE

I. Introduction

General Issue

In a recent article for Defense magazine, Lieutenant General Lawrence stressed the instructional use of wargames, and stated that:

> With the recent increased emphasis on wargames at all levels of the Department of Defense, they will play an increasingly important role in the training of our nation's leaders and in formulation of military strategies and doctrines. (7:29)

As the Department of Defense begins to increase its reliance on the wargame as an instructional tool, the need to study the external validity of this important tool becomes more acute.

Key Terms and Definitions

A **simulation** is any system or operation which has a relevant behavioral similarity to the original system or operation. Therefore, a simulation does not have to be, and usually is not, an exact reproduction, but must behave in a similar fashion. The simulated system is called the simuland (10:20).

A **game** is a simulation in which the human participants

assume the role of managing the simuland. In other words, the focus of the game is on the human element.

Validity is a representation of how well a game measures the variables as conceptualized.

External validity is an expression of how well a game corresponds to the real-world situations that occur in the simuland.

Background

The use of wargames to train military officers is not a new concept. Indeed, the first recorded reference to the utility of wargames appeared as early as the fifth century B.C., in the writings of the Chinese philosopher Sun Tzu (5:1). Since Sun Tzu, wargames have figured prominently in military history. For example, the Japanese did not make the decision to launch the attack at Pearl Harbor, until their political and military leaders had played an exhaustive series of wargames to determine the repercussions (1:133). After the war, Admiral Nimitz proposed that the Navy increase its emphasis on wargaming, because he had observed that every major engagement of World War II had been previously experienced in the context of a wargame at the Naval War College (10:292).

Since World War II, the use of military games in the United States has expanded dramatically, mostly due to the introduction of the computer. The computer offers a quantum leap forward for researchers and game designers because of the vast data management and computational tasks that can be

performed. In 1981, Harris and Nickelson reported that the Department of Defense had an 'official' catalog of 152 military games (5:7). At the rate they were being introduced, that number could now be well over 200. The Air Force, in particular, has recently increased the emphasis on the use of wargames for training purposes. Wargames have been incorporated in the curricula of the Air War College, Air Command and Staff College, and the Squadron Officer School. Recently, the Air Force Institute of Technology (AFIT) incorporated a wargame into the capstone course of the Systems Management program (3).

After World War II, the civilian sector began to apply the principles of military gaming to business problems. In the last 25 years, over 228 business games have been introduced for instructional purposes (13:251). According to Wolfe (14:350), 89 percent of the industrial firms surveyed reported the use of games in some capacity. As far back as 1975, Wolfe reported that 90 percent of prestigious academic institutions were using games as part of their curricula (14:350). Game performance represents as much as 55 percent of a student's grade in these courses (13:258-261).

The Problem

Despite the long association between military organizations and the instructional use of games, virtually no research has been done to validate the concept of a

wargame. General Lawrence, President of the National Defense University, has defined the specific role that wargames should fill.

> The purpose of wargaming, then, is not to teach students how to react to specific situations, but to make them aware of the factors that influence the outcome in conflict situations...By learning what is important, they will be better prepared to develop the real operational plans needed to secure strategic and military objectives. (7:25)

Given this role, the next step must be to determine if wargames do, in fact, teach students the factors that influence outcomes.

Research Hypothesis

Managerial comprehension and multi-level synthesis are nebulous concepts and difficult to measure. However, a purpose of the AFIT Masters Degree program is to develop these skills. Therefore, if a group of Air Force officers were to play a management wargame that was designed to reward these abstract skills, then those officers who have completed an AFIT Masters Degree would be expected to outperform those officers who have not completed the program.

If the two groups of officers are alike in all other respects, then a significant positive correlation between game performance and the AFIT degree would indicate that the game provides a measure of management ability. This implies that the game is analogous to situations that will be encountered in the real world.

Scope

There are many factors which could influence management ability. Experience, past performance, professional military education, and advanced management education are among the criteria the Air Force uses to promote officers to higher levels of managerial responsibility. Since past performance is difficult to quantify, this study concentrates on the examination of game performance as it relates to different types of job experience and the different levels and types of education.

Clearly, the Air Force devotes significant resources to sponsor officers through advanced management degree programs, including in-residence programs such as AFIT. The results of this study depend on the assumption that an AFIT Masters Degree program significantly develops managerial skills. By selecting officers assigned to AFIT across two different year groups, possible confounding variables were eliminated. Since the characteristics of the AFIT population do not vary significantly from year to year, any variation in game performance should be due to advanced education rather than some other factor.

The focus of this research is on the examination of individual performance. Unlike most wargames, which involve multiple participants, performance was not affected by player interaction. Although human-human relationships are certainly important in management situations, attempting to incorporate them into this particular study would have

introduced numerous subjective and uncontrolled variables beyond the scope of this research.

II. Discussion of the Literature

The Scope of Military Gaming

If, as General Lawrence stated, the goal of wargames is to prepare future leaders for the day when they will have to make operational plans, then most of the wargames are flawed. Shubik (10:293) surveyed the inventory of military games, and found that virtually all games used by the Department of Defense are dedicated to the evaluation of weapon systems. Many of these games deal with missile kill probabilities, aerial dogfight tactics, perimeter defense, and similar issues. While these factors are all important, and worthy of study, the 'big picture' has been missed. How many of the students that these games are targeted for will see action in the cockpit? And yet a factor like logistics, a classic and universal military problem, has been ignored in every single one of these combat-oriented games. If the purpose of these games is to assess the effectiveness of weapon systems, then they may be sufficient. But, if the purpose is to educate the leaders of tomorrow, then the scope must be broadened to encompass the factors that these students are going to be expected to deal with in a crisis situation.

One attempt to address the element of logistics was made by the introduction of the simulation called LOG-PLAN-X. The computer-assisted simulation is a management exercise which tasks the players with providing logistical support for one reparable item (5:1-34). The

scope of this game is also too narrow because it again deals with only one of the main elements of defense planning, just as the class of games discussed previously examined only weapon effectiveness. The goal of a game targeted for management students should be to teach these students to comprehend and manipulate interrelating variables. As Harris and Nickelson said:

> Nothing else short of real world experience allows one to consider a total management situation with its attendant variables, interactions, and relationships (5:3)

While the 'total management situation' cannot ignore logistics, logistics cannot be an end in itself. The narrow scope limits the effectiveness of these games as training tools, and makes any discussion of the external validity of these games extremely difficult.

Emphasis on Team Play

Another factor impeding an academic study of the validity of wargames is the emphasis on 'team play.' Military games are almost universally multi-player in nature. Therefore, any attempt to find a correlation between game performance and real-world performance is confused by a series of subjective factors which are difficult to control in a research setting. Successful game play should represent an individual student's comprehension of the simuland, if the game has external validity. Yet, in nearly all of these games, game performance was much more a factor of the team selection and whims of the particular

game director. According to Harris and Nickelson:

> A primary determinant of the success or failure in playing the simulation was the initial selection of team members. LOG-PLAN-X does not provide for a clear-cut chain of command authority structure within the simulated environment. As such, the personalities involved may be the driving success/failure during play...the course director is afforded the opportunity to create power vacuums and conflicts at will. (5:20-21)

Although Harris and Nickelson considered this variability to be a desirable feature, most civilian researchers disagree. In studies examining the validity of business games, researchers are relying on games that emphasize the single-person firm (4:290, 9:266, 11:27-37, 13:255, 15:57-58).

Validity--Military Sector

Despite the increased funding and the elevated emphasis on wargames by Department of Defense leaders (10:292; 7:25), very little attention has been given to the subject of validity. The use of games is routinely justified on the basis of subjective surveys of the participants. In general, no attempt is ever made to make a quantitative comparison of the game to the simuland. For example, the game LOG-PLAN-X, was never tested by the researchers because "Time considerations mitigated against our playtesting of the proposed versions" (5:34). Additionally, TEMPO-AI, a resource allocation game, is being incorporated into the Squadron Officer School (12) program despite inadequate playtesting.

These examples are the rule rather than exception. A great deal of effort (sometimes years) is invested in

developing these wargames, but the products are never tested. A search of the literature on military gaming revealed only subjective descriptions of the wargames, with an occasional survey of the subjective perceptions of the participants. No quantitative studies were found.

Validity--Civilian Sector

Researchers attempting to validate the civilian counterparts to wargames have discarded the use of participant surveys as the primary research instrument. Previous subjective studies contradicted each other as the surveys alternately indicated that business games were or were not valid depending on the prejudices of the researchers. The only consistent finding was that those students who performed well in the game were more likely to think that the game was valid. Students who did poorly were less inclined (11:36). The need for 'hard' data has been recognized by many authors (1:166; 4:291-292; 11:27-37; 13:274-282; 14:349-365).

Despite the relatively short history of business games, civilian researchers have made considerable progress in the examination of validity. In 1967, Vance and Grey found a high degree of trait similarities between successful business executives and students who performed well in business games. They also found a positive correlation between game performance and academic performance, as measured by overall grade point average (11).

In 1975, Wolfe and Guth (14) decided to directly test the effectiveness of games by contrasting them to traditional methods of instruction. Students were given a comprehensive examination, and were then split into two groups. One group played business games, and the control group read and discussed case studies. At the end of the course, both groups were again given a comprehensive examination. The scores of both groups improved, but the scores of the students who were taught the game were significantly higher than students who were taught using the traditional method of case studies.

Pierfy expanded on this line of research by examining 22 studies that compared case studies with games. Although his findings did not show a significant increased performance immediately following the games, most of the studies did show that those students that played the games had a greater retention of factual material over a period of time (9).

Recently, Wolfe conducted a follow-up study of the participants in his previous study (15). Surveys were sent to groups of students who had participated in the business games. After five years, the students that had played the game had significantly higher salaries than those that had been taught with case studies. This is particularly remarkable since the game was only part of one course in a 42 course program. Again there was a strong correlation between the overall grade point average of the students, and

the game performance of these students. Wolfe cautiously avoided claiming that his research proved validation of business games, because some other measures of business success, like the subject's proximity to his Chief Executive Officer, did not show any significant correlation. More research will have to be done before anyone can conclude that games are superior to other instructional tools. However, all evidence seems to indicate that games are at least a viable alternative to case studies.

Summary

Military gaming is extensive and increasing. Wargames are a practical and inexpensive way for leaders and managers to experience crisis situations without suffering the consequences of their actions. This experience is especially appreciated in a field where there may not be any "second chances."

The importance of wargames makes it imperative that serious academic attention be focused on the validity of the instructional tools. Every effort must be made to ensure that the experiental learning corresponds to the simuland. To date, this has not been done. After reviewing 132 military games, Shubik concluded that "Documentation is, in general, poor and there appears to be a lack of scientific standards in data collection and validation...The level of professional communication appears dangerously low" (10:293).

As more resources, both in terms of money and manhours, are devoted to wargaming, the need for justification of

these resources becomes critical. The methodology used by the civilian sector to successfully validate business management games will work equally well with military management games. Games must be developed which encompass the full range of factors that the leaders will be expected to manage in the operational environment. These games must then be extensively playtested to work out any bugs or 'glitches.' Wherever possible, games should be designed for a single manager, and performance should be measured by quantitative, objective standards rather than the subjective opinion of the course director. Finally, the longitudinal studies such as those done by Wolfe (15), should be conducted to ascertain the external validity of the game model.

III. TEMPO and TEMPO-AI

TEMPO Background

TEMPO is a military planning game designed exclusively as an instructional tool (6). The use of TEMPO as an educational wargame goes back as far as 1962. The game has been well received by both instructors and students.

TEMPO is intended for students in management positions, and is currently in use at both the Air Force Squadron Officer School and the Air Force Institute of Technology. The game has been played in well over a dozen countries among groups as diverse as Pentagon generals and elementary school teachers (8).

Because TEMPO has been extensively playtested over a number of years, it is an attractive subject for research. The game already has considerable 'face validity.' That is, knowledgeable experts accept that the game does what it is supposed to do.

TEMPO Game Description

TEMPO can best be described as a resource allocation game between two equal, but competing, teams. A detailed description of the rules for TEMPO is included as Appendix A. The team size varies depending on where the game is played, but five or six players seem optimal. Regardless of team size, each team attempts to acquire and operate weapon systems selected from a menu of possible choices. Players have to make cost/benefit trade-offs based on the following

factors:

1) Cost
2) Schedule
3) Performance
4) Logistics

Constraints are placed on the available resources, and performance must be considered relative to the enemy's capabilities. For example, a defense system which is not defending against an opposing offensive system is devalued.

Just as Congress sets a fixed limit on Department of Defense spending, TEMPO sets a fixed limit on the amount of resources that can be spent in a given game-turn. The task of the students is then to get the biggest 'bang-for-the-buck.' In other words, the students are expected to get the most utility out of each dollar spent.

The effectiveness of each weapon system, both offensive or defensive, is predetermined and a 'util' value assigned to each system. This util rating is a measure of a given systems performance, but this rating must be kept in the context of what the enemy is doing. Utils can be in each of four categories: OA (Offensive A), OB (Offensive B), DA (Defensive A), or DB (Defensive B).

As the labels imply, DA utils counteract opposing OA utils, and DB utils counteract opposing OB utils. Defensive utils in excess of the corresponding Offensive category are not counted (except that 20% of the excess defense can be applied to the other defensive category).

The util rating of a system provides some indication of the benefit of purchasing or maintaining that system. But

the students must also consider the costs of that system in their decision process. There are three types of costs that must be considered:

1) Research and Development
2) Acquisition (New Systems) or Modification (Old Systems).
3) Operation and Maintenance costs

Since this is a competitive game, the opponent must be considered. Therefore, TEMPO allows resources to be applied to intelligence activities or counterintelligence.

Utils are only credited to systems that are operated in a war year. Therefore only resources expended on Operation yield a return on the investment. However other expenditures can provide future capabilities, information on new systems, information on the opponent's systems, or deny information to the opponent.

Each unit of a system contributes its util rating to the team's total until a saturation point is reached. This saturation point is defined as 2000 utils for any one weapon system. Units in excess of 2000 utils still contribute to the total, but at a decreasing value. This simulates the effect of diminishing returns when any one type of weapon system is produced to the the exclusion of others. A graphic representation of how the concept of diminishing returns works in TEMPO is presented in Figure 1. The natural consequence of this rule forces astute players to diversify.

Time plays an important role in the game, just as it

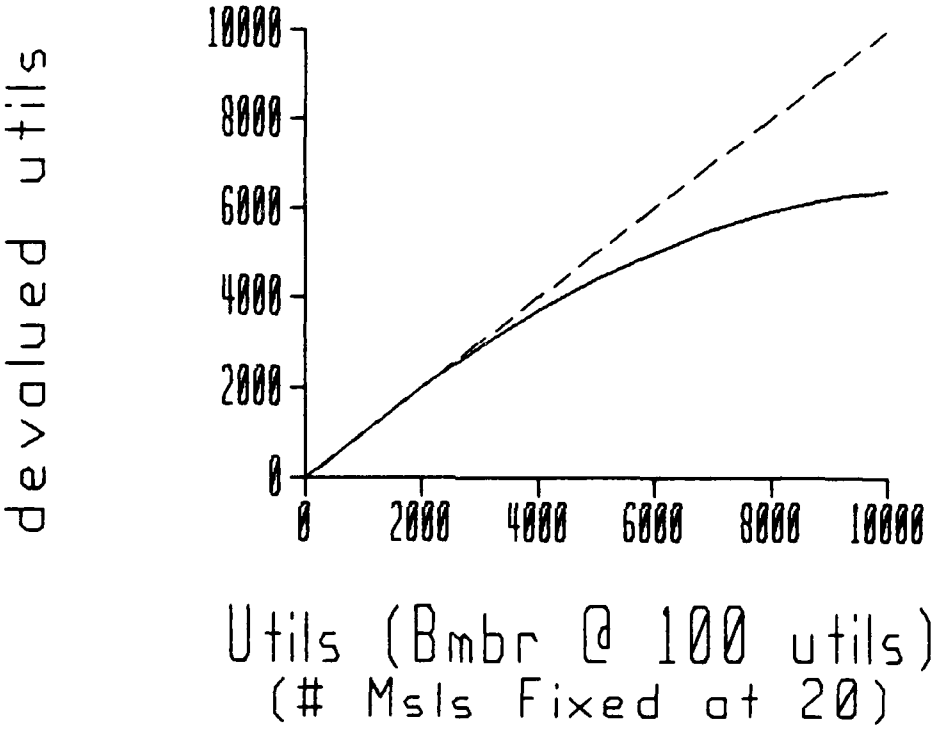


Figure 1. Diminishing Returns of a TEMPO System (6)

does in the real world. The total utility of a system is dependent on the number of years it is operated. The total cost per unit is also dependent not only on the number of units purchased, but on the number of years they are kept. In addition, the students themselves are limited in the amount of time that can be spent on any one Game-turn. Therefore, time is a resource, and TEMPO places constraints on that resource.

Uncertainty is another important aspect of TEMPO. During the Research and Development phase, costs and utils of a system are only estimates. Like real estimates, they are seldom exact. Some systems improve and some decline. The occurrence of war is also uncertain. Since neither team knows when a war will occur, it is impossible to apply some kind of deterministic optimizing strategy. For example, a strategy that optimizes utils in Year 4, does so by sacrificing optimality in other years.

Many other forms of uncertainty exist in TEMPO, and some are considerably more subtle than those already mentioned. For example, first time players will not know how many new research proposals will be introduced or whether they are worth waiting for. Some systems will come out with a modification; others do not. Also, the other team must always be considered. Defending against a non-existent attack decreases the effectiveness of a team's forces.

Certain events can result in a penalty being applied to

one or both of the teams. The most common of these occurs when a war occurs. Both sides immediately lose $400 from the next budget. In addition, the losing side also loses money equal to the margin of his loss as measured in utils. These penalties are an abstract representation of combat losses. Lesser penalties are applied to teams that do not submit their orders on time or overspend their budget. In the latter case, twice the difference is subtracted from the next annual budget. Except for an extreme loss in a war, none of these penalties represent a sizable portion of the budget.

Selection of TEMPO-AI

From an instructor's point-of-view, TEMPO is the ideal educational wargame. There are less than five pages of simple rules, yet the decisions are complex. The mechanics of running the game are negligible, but the lessons learned are profound. Although there may not be any one 'right' way to play TEMPO, it is easy for an experienced observer to spot many 'wrong' ways.

Despite the many attractive features TEMPO offers, it is a difficult game to study quantitatively. As previously mentioned, the game is designed as a multi-player exercise, and the effects of group dynamics are influential. There is no easy way to score the game, since the score is necessarily a relative measure and not an absolute one. Therefore, performance cannot be measured quantitatively.

In addition, TEMPO requires many subjects and at least

two referees for a full day. This limits the sample size, and therefore limits the conclusion of any results obtained. TEMPO may also be partially dependent on the experience and motives of the game director. Therefore, there may not be any meaningful comparison between results obtained at different locations and times.

Recently, however, a single-player version of TEMPO, known as TEMPO-AI (12) was developed by Captain White at the Air Force Institute of Technology (AFIT), School of Engineering. The game is designed to provide students taking Squadron Officer School with the same opportunity to learn from TEMPO as those officers that attend in residence. TEMPO-AI is a computer program that runs from a floppy disk on any computer compatible with an IBM operating system.

TEMPO-AI places one student against a computer program that was designed with the principles of Artificial Intelligence. The computer uses a simple, but consistent, algorithm to choose its moves. Therefore, any score is a function of individual performance against a fixed opponent. The use of a computer program should standardize the game such that results could be replicated and compared at different locations and different times. More significantly, the nature of the computer version permits a great deal of flexibility in testing subjects and accumulating a significant sample size.

TEMPO-AI Game Description

A detailed description of the rules for TEMPO-AI are included as Appendix B to this thesis. This section focuses on the relevant differences between TEMPO and TEMPO-AI.

The most notable feature of TEMPO-AI is that it is designed for individual use without the benefit of an instructor. Therefore, the game is smaller in scale. TEMPO-AI covers only eight years instead of ten, and there are fewer systems to choose from. By the third game-turn, the last new system has been introduced and the remainder of the game plays fairly quickly.

In TEMPO-AI new weapon systems can be operated the same year that they are acquired, instead of the delay imposed in TEMPO. Also, the research and development time tends to be shorter, but relatively more expensive.

The game still divides weapons into the same four categories, but uses 'Air Force' labels (Bomber, Fighter, Missile, Anti-Ballistic Missile) rather than the generic labels of TEMPO. Students are required to balance weapon types in TEMPO-AI, replacing the TEMPO rules for giving diminishing returns on the operation of any one particular system. The balance is computed based on numbers of systems rather than the utils of those systems. A given system provides full utils until the total number of units in that category exceed twice the number of systems in the complementary category. Unlike TEMPO, there is no gradual diminishing returns curve applied to the excess systems. A

graphic representation of this 'Balance of Forces' rule is presented as Figure 2.

TEMPO-AI places more severe penalties on students that overspend their budgets, but less severe penalties on those with more 'real-time' feedback than TEMPO where the feedback takes place in a debriefing at the end of the game. In TEMPO-AI, the players receive an annual report of whether they were ahead or behind, and the intelligence is signif icantly more accurate than its TEMPO counterpart. Also TEMPO-AI does not permit counterintelligence to reduce the accuracy of the intelligence.

TEMPO-AI has somewhat more uncertainty than TEMPO, as systems in the research and development phase sometimes fluctuate wildly. Unlike TEMPO, the percentage chance of war is a function in the difference between offensive forces. This also tends to make the chance of war more uncertain.

The rationale for most of these changes are included in Captain White's thesis (12). Some of the changes are cosmetic, others could potentially affect the way the game is played. A summary of the differences between TEMPO and TEMPO-AI is presented as Table 1.

Suitability

Despite the differences, TEMPO-AI attempts to teach the same principles as its predecessor. Students must still go through the same problem-solving tasks, and their decisions must be based on the same types of considerations.

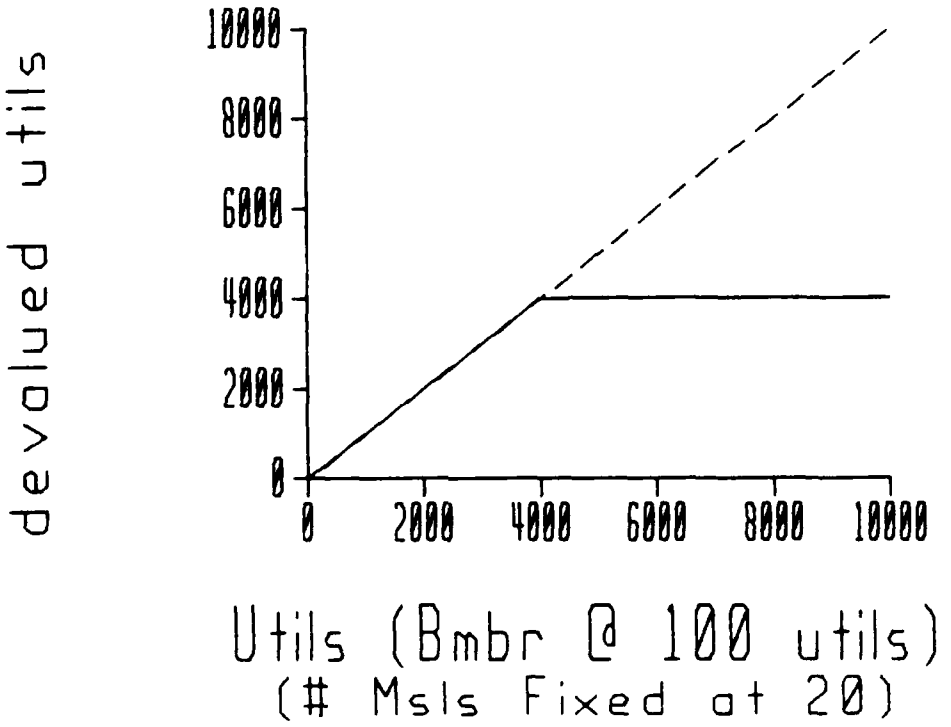


Figure 2. Balance of Forces Penalty on a TEMPO-AI System

Table 1

TEMPO versus TEMPO-AI

Characteristic	TEMPO	TEMPO-AI
Duration	8 hours	1.5 - 2 hours
# of players	10 - 24	1
# of referrees	min = 2	0
Year Published	1962	1986
# Game-Turns	10 (inc 5-yr plan)	8
# weapon system categories	4	4
# possible weapon systems	20	15
Effectiveness of weapon systems	'utils'	'utils'
Acquisition	One Turn Delay	No Delay; Cost includes Op Cost
'moth-balling'	Only for units in modification	Not allowed
diversification encouraged thru:	Diminishing Returns (Figure 1)	Balance of Forces (Figure 2)
% chance of war	predetermined	Influenced by disparity of forces
Intelligence	Changes in force structure/R&D	R&D and/or Range of forces
Counterintelligence	Reduces accuracy	Not allowed
Overdefense	Counts 20% vs other category	Does not count
Overspending	Penalized in next year's budget	Loss of current system + budget
Feedback	Debriefing	Final score
Victory	No explicit victory condition	Total net offensive utils (turns 1-8)

Unfortunately, the author of TEMPO-AI (12) could not adequately playtest the game. Therefore, there is no assurance that TEMPO-AI will accomplish its objective of providing correspondence students with the same learning experience afforded those students who attend Squadron Officer School in residence. This is not a criticism of the author, but rather a systematic flaw in game development in general.

Therefore, TEMPO-AI is an excellent game to test in a research setting. It is a single player version of an established educational wargame which also provides a quantitative measure of performance. If the game is a faithful recreation of its predecessor, than those students with superior managerial skills would be expected to perform better than those that do not. In any event, extensive playtesting by a group of students at the same level as the intended users should reveal imperfections which can be corrected before the game is implemented in a mass education program.

IV. Experimental Design

Overview

One of the first steps taken to establish the validity of business games appeared in the 1967 article by Vance and Grey (11). In that study, Vance and Grey established that successful business executives and successful game players exhibited similar traits. As previously mentioned, that study also found a high correlation between game performance and other measures of student performance.

The experimental procedure used to examine the validity of TEMPO-AI was similar to that done by Vance and Grey. The purpose of the experiment was to test whether successful game performance correlates with traits considered important, by the Air Force, for advancement to positions of greater managerial responsibility. This chapter details the steps that were taken to accomplish this.

The first topic that will be discussed concerns a series of pretests that were done to establish a specific test procedure and a reliable measure of performance. This will be followed by a definition of the population, and a description of the sampling method used. Then the specific test procedure will be explained, including a discussion of the rationale for the selection of the factors under examination. This chapter will then conclude with a description of the statistical test and hypotheses used to analyze the data.

Reliability

Wargames are often criticized because some people believe that it is possible to achieve a high score by taking advantage of some quirk in the rules. In other words, the subject is able to solve a problem in a manner not available in the real world because of the nature of the game. To address this problem TEMPO-AI was rigorously pretested prior to the actual experiment.

The first phase of this pretesting involved soliciting comments from three TEMPO referees who had played TEMPO-AI. This first phase was uncontrolled; there were no time limits, direct observations or performance measurements. The comments indicated that two games were not identical, but also encouraged further study. Working from the other direction, an experienced TEMPO-AI player observed three games of TEMPO. This initial experience identified the major differences between the two games, but also seemed to indicate that the basic decision process was the same for both games.

The second phase of pretesting involved six professional wargame players/designers. The group consisted of three civilians and three military members, with a representative from the Army, Navy and Air Force. This group played TEMPO-AI over 30 times in an effort to discover flaws in the programming logic. Different time limits were used, and different amounts of information were given to the test subjects. No scores were collected in this phase, but all

of the games were directly observed and participant feedback was encouraged.

The pretesting revealed that the computer program ran both quickly and accurately. There did not appear to be any unintentional loopholes in the programming logic, but the on-line documentation proved to be inadequate. Several key rules were not listed at all, and others were difficult to understand. No first-time player was able to interpret the on-line documentation correctly. Every player understood the rules after they were explained verbally.

Based on the results of the second pretest, a specific test procedure was determined and tested on a third group of people. This group consisted of five people, all of whom had completed or were in a graduate program. The group contained two doctoral candidates in Electrical Engineering, a business executive, and two officers' spouses. All five were tested under the same controlled conditions that were later used to conduct the actual experiment.

The results of this final pretest indicated that the test procedure was sufficient. All test subjects were able to comprehend the rules, and completed the game within the time constraints. Although the five test subjects were selected as a sample of convenience, all other aspects of the pretest were identical to the actual experiment.

Population

Test subjects were drawn from two different populations. Population A consisted of students assigned to class

87S or 87D in the Air Force Institute of Technology (AFIT), School of Systems and Logistics. These students were tested in their final quarter, and therefore had completed almost all requirements for a Masters Degree. Population B consisted of students assigned to the AFIT class 88S or 88D. These students were tested prior to taking any graded course work in their Masters program.

Since selection criteria for AFIT do not vary significantly from year-to-year, this thesis assumes that the two populations are similar in all respects except the advanced management education received at AFIT by Population A.

Sampling

There were 159 students in Population A. Population B contained 162 students. In order to obtain the initial test subjects, every third student in Population B was sent an invitation to play TEMPO-AI and a very brief description of the game. Of the 54 students in the sample, zero accepted the invitation which corresponds to 0.0 percent response rate.

Since TEMPO-AI is designed for voluntary use by students enrolled in a correspondence course, the lack of response is discouraging. Clearly, TEMPO-AI cannot be of any use unless the target population can be induced to play the game. This is a significant finding of this research which will be discussed in greater detail in Chapter VI.

Although significant, the non-response did not fulfill the purpose of this study or address the research

hypothesis. Therefore, a different approach was taken to obtain a sample and test TEMPO-AI under controlled conditions. The second sample was obtained by providing an incentive to participate. In this case, the incentive was a financial reward to be presented for the best three scores.

In other words the experiment was advertised as a competitive tournament with prizes to winners. The competition was open to all members of both populations, and the game was profusely advertised with flyers, posters, and letters. Interested students were asked to put their name on a sign-up roster to obtain a description of TEMPO-AI.

Fourteen students from Population A signed the roster; they were contacted individually and provided a detailed description of TEMPO-AI. This description is included as Appendix B. Thirteen students from Population B signed the roster and received the game description. These students comprised Sample A and Sample B respectively.

Sample A and Sample B were not true random samples of the population since they consisted of volunteers enticed by the possibility of financial reward or simply by the competition itself. However, a Post-Test Questionnaire was used to ensure that the samples were representative of their populations. A demographic representation of these samples is provided in Chapter V. This Post-Test Questionnaire is included as Appendix C to this thesis.

Test Procedure

Play of the Game. As previously mentioned, all test

subjects were given a copy of the TEMPO-AI rules in advance. These rules are included as Appendix B. Additionally, the game director reviewed these rules with each subject at the beginning of the test.

All tests were conducted in the same terminal room on Z-248 microcomputers. None of the subjects had any experience with TEMPO-AI, although many had played versions of TEMPO. Subjects who forgot to bring scrap paper, a calculator, or their copy of the rules were provided these materials. As many as five subjects were playing the game at any one time. All subjects were free to ask questions concerning rules, but were not given advice concerning strategy.

Each Game-Turn had a time limit, and there was a two-hour limit for the entire game. The two hour time limit was primarily imposed due to a shortage of terminals, but did not appear to be a severe constraint. The specific time limits were:

```
Game-Turn 1   ::  30 minutes
Game-Turn 2   ::  25 minutes
Game-Turn 3   ::  20 minutes
Game-Turn 4   ::  15 minutes
Game-Turn 5-8::  10 minutes each
```

More time was allowed for the early turns because new R & D (Research and Development) and acquisition proposals had to be evaluated. Additionally, the players needed to become familiar with the computer program. The later turns involved less thinking, because no new programs were emerging.

Scoring. A major problem with military gaming, particularly management wargames, is the determination of a consistent measure of student performance. In most business games, the objective is to make the largest profit or the best return on investment. The objective of a military manager is less clear.

TEMPO-AI provides a score based on the sum of net offensive utils acquired by a player throughout the game. This score is compared to the computer's total to determine an overall 'winner.' This method is inadequate because it fails to differentiate between war years and non-war years. Timing is an important aspect of TEMPO. This may be the reason the name TEMPO was chosen for the game. In both TEMPO and TEMPO-AI students are tasked with dealing with time and the uncertainty of war. Therefore, only the game-turns that involved a war were counted in this experiment. Since the number of wars in any one playing varied, the total number of net offensive utils was divided by the number of wars to determine an average. The formula for determining score was:

$$\text{Score} = \frac{(\text{Player's Total}) - (\text{Computer's Total})}{\text{\# of wars}}$$

where the totals are expressed in net offensive utils.

Post-Test Questionnaire. Immediately upon conclusion of the test, participants were asked to fill out a short Questionnaire. This is included as Appendix C to this thesis. The Questionnaire was included to help ensure that

the Samples contained a fair representation of the two populations and that the populations were similar in all respects except for advanced management education.

Class. Question 1 asked students to identify their academic class. The answers to this question were used to sort students into Populations A and B for purposes of answering the research hypothesis (H1).

Education. There were two purposes to question 2. The first was an attempt to establish that neither sample contained students with previous advanced management education. The second purpose involves a secondary hypothesis. Recently Air Force Systems Command has implemented a policy of increasing the numbers of engineers in the Acquisition Program Management career field. Experienced program managers were moved out of the career field to allow entrance of engineers on the assumption that their technical expertise would improve management competence. A secondary hypothesis (H2) explores this issue.

Degree. The purpose of this question was to ensure that neither sample contained students with previous advanced management education from other institutions. This was a potentially confounding variable.

Military Experience. This question ensured that the Samples represented students with about the same level of military experience as those students that TEMPO-AI was originally designed for.

Operational Experience. Officers in the Acquisi-

tion career field are encouraged to seek operational experience. The operational perspective is seen as a desirable aid to managerial decision making. Certain positions are restricted to those who have this experience, and certain number of Air Force acquisition officers are rotated through operational assignments. The affect of operational experience on game performance was explored as a secondary hypothesis (H3).

Acquisition Experience. Since a major part of TEMPO-AI involved the acquisition of new weapon systems, the issue of previous acquisition experience appeared relevant. This question supports another secondary hypothesis (H4).

Wargame Experience. This question was included to provide a check on another potentially confounding variable. Theoretically, those students experienced in other wargames might be more comfortable with this type of test.

TEMPO Experience. TEMPO game directors have not seen any evidence that previous experience with one version improves performance in another version (8). In fact, there may be a decrease in performance due to misconceptions (3). This question examined the possible influence of familiarity with other versions on TEMPO-AI performance.

Enjoyment. Question 9 on the Questionnaire asks the participants how well they enjoyed the game. This is not a frivolous question because the success of TEMPO-AI in a correspondence course depends on the students completing the game. This is more likely to occur if the students

enjoy what they are doing.

A secondary purpose of this question, and all other questions that followed it, was to test whether the two populations perceived the game in the same way.

Performance (Perceived). Although the computer program provides the user with a score, the score has no meaning except in comparison to others. All but two of the test subjects 'lost' to the computer, but they were informed in advance that the computer was likely to win. Therefore, players who scored poorly would not necessarily know that they had done anything incorrectly. The converse was not true; players who beat the computer knew they had done well.

Thought. This question provided a check on whether or not the students took the game seriously. It also provided some indication as to whether the required decisions were challenging.

Understanding. Both this question and the next question provided an indication of whether or not the students thought they knew what was going on. Regardless of performance, students that end a game completely bewildered may not have learned much.

Confidence. Like the previous question, this question sought to determine whether the students were following some kind of plan (however flawed) or merely guessing.

Realism. TEMPO-AI is a deliberately abstract game; the simuland being represented is problem-solving not nuclear war. However, test subjects frequently discount a

game if they feel it isn't 'realistic.'

Utility. In some ways, this was the most important question in the Questionnaire, since the perceived utility may determine whether or not the students felt that there were lessons to be learned from playing TEMPO-AI. As stated earlier, this thesis did not rely solely on student's perceptions. If the game was designed properly, they might learn important lessons from the game without fully realizing it. However, if students felt that the game was useful, they might be more successful in applying the lessons learned to real situations.

Comments. The final question was open-ended and simply asked students for any comments they might have on TEMPO-AI. Although not required, many students did take the time to write many suggestions for improvement.

Statistical Test

Test Statistic. For the primary research hypothesis scores were sorted by the student's answers to Question 1 on the Post-Test Questionnaire. For the secondary hypotheses, the scores were sorted into groups based on the student's answer to the relevant question on the Questionnaire.

Two-Samples t tests were then conducted to determine whether there were significant differences in the computed mean scores. The use of a t test will produce a valid test if two main assumptions are made. The first assumption is that the sample is drawn from a normal distribution; the second assumption is that the population variances are

approximately equal (even if unknown).

The latter assumption is easily satisfied by a visual inspection of the data. However, the assumption of normality is difficult to substantiate with the small sample obtained. Fortunately, the t test works well even with approximately normal distributions. According to Devore:

> In fact, it has been shown that if the distributions being sampled are not too nonnormal and/or the two variances are not too different from one another, the t test works reasonably well in the sense that the actual level of significance is approximately the specified alpha...the t test is robust in the presence of mild departures from assumptions (2:292).

Therefore, the t test should produce a valid test for the hypotheses involving comparison of game performance between two groups.

The statistical test will be a right-tailed test conducted at alpha = .05 (t_{crit} = +1.717). The null hypothesis will assume that the mean score for Population A will not be significantly higher than the mean score for Population B. If T is greater than t_{crit}, then the null hypothesis will be rejected in favor of the alternate (that the mean of Population A is significantly higher than the mean of Population B).

H_o: $u_a - u_b <= 0$

H_a: $u_a - u_b > 0$ (Right Tail Test)

The small sample test statistic is:

$$T = \frac{A - B - 0}{S_p * (1/m + 1/n)^{1/2}}$$

where A and B are, respectively, the sample averages for Populations A and B, and S_p is the pooled estimator of the sample standard error. The sample sizes are denoted m and n, respectively.

Hypotheses. The primary Hypothesis (H1) examines the affect of advanced management education on game performance. Sample A contains students that have nearly completed a Masters degree at AFIT, and Sample B contains students that have not.

The affect of technical education on game performance is the subject of the second Hypothesis (H2). Sample A is redefined to include students from both year groups that identified themselves as engineers. Sample B then contains those students that have not had engineering.

Operational experience is the focus of H3. Sample A contains those officers with two or more years of operational experience. Sample B contains those that indicated that they had less than two years experience in operational assignments.

The next hypothesis (H4) investigates the impact of acquisition management experience on performance in TEMPO-AI. Sample A contains those students with two or more years of experience in an acquisition-related career field, and Sample B is then defined as those with less than two years experience.

Previous familiarity with wargames is screened in another hypothesis (H5). Those students that indicated that

they had at least some previous experience with other war-games are defined as Sample A, and the remaining students formed Sample B.

Finally, the affect of previous experience with TEMPO on TEMPO-AI performance is examined (H6). Those with previous experience with TEMPO are defined as Sample A, and the officers with no previous TEMPO experience are defined as Sample B.

Other Tests. Although the emphasis of this experiment is on the factors that affect game performance, a Post-Test Questionnaire was used to obtain student's perceptions about the game. In Chapter V, the answers for each of the samples are compared with respect to the primary and all secondary hypotheses. The results are presented using two-way contingency tables with a chi-squared test for homogeneity as described by Devore (2:541-543).

Additionally, pairs of questions were compared using a sample correlation coefficient (2:444). For example, students were ranked according to their own perceived performance, and this was compared to a true ranking according to actual performance. Due to time constraints, all of the possible combinations were not examined. The correlation coefficients are reported for those combinations that were particular interesting or that were particularly surprising.

V. Results and Discussion

Response Rate

As discussed in Chapter IV, 14 students from Population A signed the roster and received the game description. Of these, 13 played the game and 12 finished with one player exceeding the two-hour time limit. Therefore, only 12 of the 13 scores counted in measures of game performance. However, the student did complete the Questionnaire. This yielded a Sample response rate of 85.7 percent for the game performance measurement and 92.9 percent response for the Questionnaire. The sample constituted 8.8 percent of the population.

Of the 13 students who signed the roster from Population B, 12 played and all completed the game. One student did not answer the last question on the Questionnaire, so only 11 responses were received on that question from this sample. This response represented 92.3 percent of the sample and 8.0 percent of the population.

Demographics

The demographic characteristics of the two samples are presented in Table 2. The pretest sample is included for comparison, although none of these responses were used in hypothesis testing. Only two of the 25 students from the Air Force Institute of Technology (AFIT) had a Masters Degree from another institution, and neither had the degree in a management-related field. Both groups contained

Table 2

Demographics

	87 Students	88 Students	Pretest
2. EDUCATION			
Engineers	7	8	2
Non-Engineers	5	5	3
3. DEGREE			
Bachelors	5	0	0
Bachelors+	5	13	2
Masters	1	0	0
Masters+	1	0	3
4. MILITARY EXP			
< 4 yrs	0	1	3
4+ yrs	12	12	2
5. OPERATIONAL EXP			
Zero	1	5	5
< 1 yr	0	3	0
1-2 yrs	0	3	0
2+ to 4- yrs	3	0	0
4+ yrs	8	2	0
6. ACQUISITION EXP			
Zero	7	3	3
< 1 yr	0	1	0
1-2 yrs	0	1	0
2+ to 4- yrs	2	3	1
4+ yrs	3	5	1
7. WARGAME EXP			
None	6	5	3
Some	5	4	1
A lot	1	4	1
8. TEMPO EXP			
Yes	7	8	0
No	5	5	5

students with about the same military experience.

Since the primary hypothesis involves a comparison of students from the two AFIT classes, it was desirable that the students contained approximately equal representation among the other factors that might potentially influence game performance. A chi-squared test of homogeneity was performed to determine if the populations contained approximately equal representation among these factors. The results of this test are presented in Table 3.

The degree of operational experience was the only significant difference found between the two samples. Since operational experience is expected to improve managerial comprehension, the '88 class might be expected to score higher than if the two samples were equally represented.

Table 3

Test of Homogeneity--AFIT Class

	'88	'87	X^2	$X^2_{.05,1}$	Significant?
Engineers	7	8			
Non-Engineers	5	5	.027	3.843	NO
>2 yrs Op Exp	11	2			
<2 yrs Op Exp	1	11	14.548	3.843	YES
>2 yrs Aq Exp	5	8			
<2 yrs Aq Exp	7	5	1.035	3.843	NO
Some Game Exp	6	8			
No Game Exp	6	5	.337	3.843	NO
TEMPO Exp	7	8			
No TEMPO Exp	5	5	.027	3.843	NO

Performance

TEMPO-AI was designed to teach management principles to students in a correspondence course. One purpose of the AFIT curriculum is also to develop these skills. Therefore students who have completed a 15-month AFIT program would have more exposure and awareness of the management principles taught in the game, and might be expected to perform better than a similar group who had not received this training. The results of the experiment, presented in Table 4, support this theory. At alpha level equal to .05, the students with the advanced management education scored significantly better than their counterparts. The mean score for Sample A was -5688.8 net utils per war--almost a full standard deviation less than the mean score for Sample B. Note that the negative sign in the performance measurement indicates that the computer outscored the players. This was true for 23 out of 25 scores.

The computed Sample mean scores and Sample standard errors yield a t-score of +1.922. With 22 degrees of freedom, this score is sufficient to reject H1 at alpha=.05. This implies that TEMPO-AI does reward some of the management skills learned in an advanced management program. This is a

Table 4

AFIT Education

	Mean	Sigma	n	$t_{.05,22}$	T	Reject H_o?
'87 class	-5688.4	7425.5	13	+1.717	+1.922	YES
'88 class	-11064.4	6225.8	12			

Sample Correlation Coefficient = +.40

necessary prerequisite for the game to be useful in teaching management concepts. Sample correlation coefficient for advanced management education and game performance was +.40. This is only a moderate correlation, but is considered significant in this particular field due to the difficulty in ascribing management success to any one factor. For example, the most widely accepted predictor of management success is a student's Grade Point Average (GPA). Yet GPA has an even more modest correlation with management success (15:54).

Technical expertise did not appear to be much of a factor. Engineers scored slightly better, but the difference was insignificant. H2 was not rejected, as the t-score was only +.498. The engineers did seem to catch on to the rules quicker and expressed somewhat more confidence. However, this did not translate into increased performance. The correlation was a negligible +.11. The results are summarized in Table 5.

Among the non-engineers, both samples contained students with degrees in computer science, accounting, or the

Table 5

Technical Education

	Mean	Sigma	n	$t_{.05,22}$	T	Reject H_o?
Engineers	-7743.6	6134.3	14	+1.717	+.498	NO
Non-Engineers	-9263.0	8843.2	10			

Sample Correlation Coefficient = +.11

physical sciences. Since TEMPO-AI does involve a lot of basic computations, this could have had an impact. However, the current Air Force emphasis on 'hard' engineering degrees prevents many of the others from entering or remaining in the acquisition career field. For example, even engineering physics is considered a non-technical career field by Air Force definition. Since this narrow definition of technical education is used by the Air Force, it was the basis of the secondary hypothesis (H2). Perhaps a more meaningful investigation would be to test the effect of a quantitative educational background on performance. Unfortunately, only three of the 25 AFIT students sampled had a truly non-quantitative background.

Perhaps the biggest surprise in the performance measurements was the negative association between operational experience and TEMPO-AI performance. H3 was not rejected, but this was due to the directionality of the test. If anything, a positive association had been expected due to the Air Force emphasis on operational assignments for acquisition managers. The two mean scores (-10,985.4 and -4725.4 net utils per war) had the largest difference of any two groups. Clearly the -2.269 t-score displayed in Table 6 would be a significant difference if the test had been structured as a Left-Tailed or even Two-Tailed test.

The slightly negative correlation (-.25) between operational experience and game performance is probably due to a lack of other types of experience rather any

Table 6

Operational Experience

	Mean	Sigma	n	$t_{.05,22}$	T	Reject H_o?
>2 yrs Op Exp	-10985.4	6112.2	14	+1.717	-2.269	NO
<2 yrs Op Exp	- 4724.4	7394.8	10			

Sample Correlation Coefficient = -.25

detrimental effects of operational experience itself. For example, the results of the test for homogeneity (Table 2) found that a disproportionate number of students in the junior class had operational experience. In fact, 91.7 percent of the '88 students had more than two years experience. This contrasts sharply with the '87 students with only 15.4 percent indicating that they had more than two years experience in that area. This is an important consideration because the '87 students had received advanced management education which did show a significant positive association with game performance.

Acquisition management experience appears to have been less of a factor. Those students with two or more years of experience scored only slightly better than those with less. However, when the groups were split between those students with zero experience and those at least some experience the performance gap widened. As Table 7 indicates, the relation became stronger, although still not statistically significant at alpha level .05.

Of the ten students with no previous experience in acquisition management, nine had 4 or more years of

Table 7

Acquisition Management Experience

	Mean	Sigma	n	$t_{.05,22}$	T	Reject H_o?
>2 yrs Aq Exp	-7655.2	7566.3	13	+1.717	+.522	NO
<2 yrs Aq Exp	-9230.1	7103.8	11			
>0 yrs Aq Exp	-7083.7	7208.1	15	+1.717	+1.137	NO
0 yrs Aq Exp	-10632.7	7177.2	9			

Sample Correlation Coefficient (<2 yrs and >2 yrs) = +.11
Sample Correlation Coefficient (0 yrs and >0 yrs) = +.24

operational experience. Since lack of acquisition experience showed a weak positive correlation (+.24) with game performance, this finding lends further support to the idea that lack of experience may be an influential factor, although the degree of experience may not be relevant past an initial familiarity with the subject.

Since proficiency with wargames is not considered a factor associated with managerial success, a significant association between wargame experience and game performance could indicate that the game lacked discriminant validity. In other words success might be linked to making good game decisions which were not necessarily representative of good management decisions.

The results in Table 8 do not provide a decisive answer to this question. Students with at least some wargame experience seemed to be more at ease with the game, and also score higher, although the difference was still not significant to Reject Ho at alpha = .05. A correlation of +.27 indicates that there might be a weak relationship

Table 8

Wargame Experience

	Mean	Sigma	n	$t_{.05,22}$	T	Reject H_o?
No exp	-10895.6	5124.0	10	+1.717	+1.470	NO
Some+ exp	- 6585.3	8144.1	14			
Some exp	- 6653.0	6890.7	9	+1.782	+.040	NO
A lot	- 6463.4	10980.3	5			

Sample Correlation Coefficient (None and Some+) = +.27
Sample Correlation Coefficient (Some and A lot) = +.04

between the format of the instructional tool, and the student's comprehension of the concepts. This is not necessarily undesirable, as long as students are not taught exclusively by one method.

The fact that Table 8 shows no difference between those with a lot of experience and those with only some, again seems to indicate that a lack of experience is the key factor. The implication of this is that a certain minimum proficiency may be important, but that there is little added value beyond this minimum level.

The results of Table 9 indicate that previous experience with TEMPO, was also weakly correlated (+.27) with game performance in TEMPO-AI. Again the t-score of 1.325 was not significant at alpha = .05, but is sufficiently high to warrant further investigation. There may very well be a learning curve associated with this game, as those few students who were able to play twice (second scores not included in results) scored much higher in the

Table 9

TEMPO Experience

	Mean	Sigma	n	$t_{.05,22}$	T	Reject H_o?
Experience	-6884.1	6903.8	15	+1.717	+1.325	NO
No Experience	-10865.2	7499.8	9			

Sample Correlation Coefficient= +.27

second game. This is consistent with the goals of the game, since the students should improve if they learned anything the first time. A possible reason why the previous TEMPO players did not show more significant improvement in TEMPO-AI, may lie in the differences in the rules. This learning curve aspect suggests an interesting line of research not pursued in this thesis.

Student's Perceptions

Overall. After completing a game of TEMPO-AI, students were asked to fill out the Post-Test Questionnaire that is included as Appendix C. In addition to the demographic data, students were asked to respond to some subjective questions concerning what they thought about the game and how they played. A Likert Scale was used, with 1 representing strong disagreement and 7 representing strong agreement. The students responses were then separated into the same groups used for the hypothesis testing to test for any difference in how the different groups responded to any one question. The responses were tested using a chi-squared test for homogeneity with the combined median being

used to separate the responses into a 'success/failure' category. The tests were conducted at alpha level .05, with one degree of freedom.

The 25 AFIT students who were enticed into actually playing TEMPO-AI had a very strong favorable reaction to it. The combined median was '6' which indicates that the students agreed with the statement that they enjoyed the game. Of the 25 students, all but two (8 percent) of the responses were favorable, and 25 percent said that they were in strong agreement with the statement. This is encouraging as it suggests that those students who start a game are more likely to finish it, but says little about the validity of the game itself.

Responses varied considerably with respect to student's perceived performance. The group median was '3' indicating slight disagreement with the statement that they performed well, and 56 percent disagreed with the statement to some extent. Sixteen percent were neutral, with the remaining 38 percent in some degree of agreement. This perception of low performance was probably a result of 'losing' a succession of wars to the computer opponent. All but two of the 25 students lost to the computer in terms of net utils per war. Despite the perceived poor performance, students still seemed to enjoy the game. Most indicated in th· open-ended comments that they thought they could do better if given a second chance.

Most of the test subjects felt that they had made a

best effort, and that they had put a lot of thought into each decision round. A full 80 percent agreed with this sentiment, with only 16 percent disagreeing. This could have been influenced by the use of financial rewards as an incentive, but most of the students continued to try their hardest even after they perceived that they were not going to be in the top of the class.

By the end of the test, most students felt that they had caught on, and that they understood the impact of their decisions. The group median was '5' indicating slight agreement with the statement, and 60 percent agreed with it to some degree. The remaining 40 percent were split evenly between those that were neutral and those that disagreed.

Similarly, most students were confident about the decisions they were making by the end of the game. The group median was again '5' and again 60 percent agreed that they were confident with remaining split between neutral and disagreement.

In contrast, the majority of those tested did not feel that the game was realistic. Those agreeing with the median of '3' (slight disagreement') constituted 64 percent of all responses. The remainder were evenly split between agreement or no opinion.

Despite the perception that TEMPO-AI was unrealistic, the majority felt that it was useful anyway. The students recognized, for the most part, that the game was deliberately abstract and felt that it has a place as an

instructional tool. Almost two-thirds agreed that the game was useful as presented, and almost all of the remainder were neutral (28 percent). Of those not agreeing, most had only one or two minor objections, which might be corrected through simple revisions.

Many students took the time to write extensive suggestions for improvement. The single most common comment concerned the difficulty in using the computer program. Several students suggested an 'on-screen' what-if tool so that they could get an overview of their projected expenditures and force structures. The computer calculates these for the players, but only after they have committed themselves.

Most students were frustrated when they overran their budgets, which occurred at least once in nearly every game. The inability to view what a player has purchased, places a premium on attention to detail and keyboard skills. These are not intentionally part of the objective of TEMPO-AI, but many students spent more time of this part of the game than on the cost/benefit tradeoffs that are supposed to test the decision making abilities of the students.

Other comments indicated that several students felt overwhelmed by the flood of information in the first two decision rounds, and didn't figure out what was going on until the middle of the game. Several mentioned that by this point it was too late to recover. This may indicate that at least one 'practice' round might be useful in

introducing this game to new players.

AFIT Education. The two samples indicated that the students from classes '88 and '87 perceived the game in approximately the same manner, as seen in Table 10. The only significant difference concerned the students' perception of realism. The graduating class tended to rate the game lower in realism as compared to the entering class. The senior class had considerable more exposure to a wide variety of management tools and techniques. As a result they may have

Table 10

Student Perceptions--AFIT Education

	'88	'87	X^2	$X^2_{.05,1}$	Significant?
Enjoyment					
<= median	10	10			
> median	2	3	.262	3.843	NO
Performance					
<= median	6	8			
> median	6	5	.337	3.843	NO
Thought					
<= median	7	6			
> median	5	7	.371	3.843	NO
Understanding					
<= median	10	8			
> median	2	5	1.379	3.843	NO
Confidence					
<= median	9	9			
> median	3	4	.103	3.843	NO
Realism					
<= median	5	11			
> median	7	2	4.996	3.843	YES
Utility					
<= median	7	9			
> median	5	3	.787	3.843	NO

identified the game as an abstraction rather than an attempt to simulate defense acquisition or nuclear war. Of particular interest is the fact that the '87 students did not feel they performed well. In fact, their perceived performance was slightly less than their counterparts, even though their actual performance was significantly better.

Technical Education. A comparison of the responses of engineers and non-engineers found no significant differences in their perceptions just as the quantitative measures found no significant differences in game performance. The results of the chi-squared test are summarized in Table 11.

Operational Experience. No significant differences were found between the perceptions of those students with more than two years experience and those with less than two years. It did seem that some of the rated officers placed more emphasis on the labels of the different weapon systems. For example, the new high technology Fighters offered in the game were often preferred over more efficient systems. It may be that the experience these officers had with one or more of the weapon systems, may have prejudiced their decision making. This may have been linked to the negative correlation with game performance with operational experience. Although the chi-squared test used did not find the difference significant, the operationally-oriented officers tended to rate the game higher in realism. For example, 26.7 percent found the game realistic and another 26.7 percent were neutral. None of the students in the

Table 11

Students Perceptions--Tech Edu, Op Exp, Acq Exp

	Tech Edu		Op Exp		Acq Exp	
	Engr	Non-Engr	>2 yr	<2 yr	>2 yr	<2 yr
Enjoyment						
<= median	11	9	13	7	11	9
> median	4	1	2	3	2	3
Performance						
<= median	8	6	9	5	9	5
> median	7	4	6	5	4	7
Thought						
<= median	8	5	9	4	8	5
> median	7	5	6	6	5	7
Understanding						
<= median	10	8	12	6	9	9
> median	5	2	3	4	4	3
Confidence						
<= median	9	8	10	7	10	7
> median	6	2	5	3	3	5
Realism						
<= median	9	7	8	8	7	9
> median	6	3	7	2	6	3
Utility						
<= median	8	7	9	6	6	9
> median	6	3	5	4	7	2

other group rated the game realistic, and only 20 percent were neutral. This may indicate that this group accepted TEMPO-AI as an abstraction.

Acquisition Experience. No statistical evidence was found to indicate that officers with experience in the acquisition management career field perceived the game differently than without this experience. Table 11 does seem to indicate a that those with more than two years

experience with acquisition management tended to rate the game higher in perceived utility. Although not statistically significant, a chi-squared value of 3.234 is noticeable. In fact, a more detailed look at the data suggests a relationship may exist. For example none of the more experienced officers disagreed with a statement that the game was useful, but 27.2 percent of the other group did disagree. In addition, 53.6 percent expressed support among those with two or more years experience, compared to only 22.2 percent for those with less than two years. This tendency is probably related to the nature of the game, which is more directly oriented towards acquisition management, although still very general.

Wargame Experience. Again, the test for homogeneity did not reveal any statistical differences. However, a chi-square score of 3.484 was registered for students perception of their own understanding. This just barely missed being significant at alpha = .05. To a lesser extent, those students with some wargame experience expressed more enjoyment and more confidence. None of these results are particularly surprising; students with more experience with this type of instructional tool felt they understood it better than those who didn't. The results are summarized in Table 12.

TEMPO Experience. Those students who were experienced with TEMPO perceived TEMPO-AI in the same way as those who had played neither game. A chi-square score of 2.679

Table 12

Student Perceptions--Wargame Exp, TEMPO Exp

	Wargame Exp		TEMPO Exp	
	Some+	None	Some+	None
Enjoyment				
<= median	10	10	11	9
> median	4	1	4	1
Performance				
<= median	6	8	8	6
> median	8	3	7	4
Thought				
<= median	6	7	8	5
> median	8	4	7	5
Understanding				
<= median	8	10	9	9
> median	6	1	6	1
Confidence				
<= median	8	9	10	7
> median	6	2	5	3
Realism				
<= median	9	6	9	7
> median	5	5	6	3
Utility				
<= median	8	7	7	8
> median	6	3	7	2

indicates that those officers who played TEMPO before felt that they understood TEMPO-AI slightly better, but this difference was not statistically significant. All off the 15 test subjects who had previous experience with TEMPO, felt that TEMPO-AI played the same way, despite the differences in some of the rules. The results of the chi-square comparison are presented in Table 12.

Other Results

In addition to the test of population means and the tests for population homogeneity, several factors were tested for statistical correlation. The correlations between game performance and the factors that were thought to influence managerial success were computed. The significance of those results were discussed in their respective sections, and are listed again in Table 13. The strongest correlation (.40) indicates a moderate relationship between TEMPO-AI and game performance.

Other pairs of factors were ranked and a sample correlation coefficient obtained per the method described by Devore (2:448). These coefficients are also listed in Table 13. There was a weak to moderate negative correlation (-.36) between game performance and how well the students liked the game, but a weak positive correlation (+.23) and how well they liked it and how they rated their own performance.

The only strong correlation found was between perceived performance and assessment of the TEMPO-AI's utility. This correlation (+.91) is consistent with virtually all literature on the instructional value of games. Given human nature, it is not surprising to find that students who think they did well also think the game has value. However, no correlation was found between actual performance and perceived performance. This indicates that the students self-assessment was flawed despite the strong attempt to

Table 13

Correlation Between Factors

Factor 1	Factor 2	r
game performance	(perceived) performance	+.15
game performance	enjoyment	-.36
game performance	understanding	+.02
(perceived) performance	enjoyment	+.23
(perceived) performance	utility	+.91
understanding	utility	-.38
realism	utility	+.26
game performance	AFIT education	+.40
game performance	Technical education	+.11
game performance	Operational Experience	-.25
game performance	Acquisition Experience	+.11
game performance	Wargame Experience	+.27
game performance	TEMPO Experience	+.27

build feedback into TEMPO-AI.

Summary of Results

Statistical evidence supports the research hypothesis of this thesis. Students who had completed an advanced management degree at AFIT did outperform similar students who had not had that education. The difference in the population mean scores was significant at alpha = .05. A positive correlation of .40 is also significant in light of the historical difficulty in attributing managerial success

to any one factor.

The implication of this finding is that TEMPO-AI is at least partially successful. Students playing the game were rewarded for making correct managerial decisions. Students with more advanced education are assumed to be better able to make 'real-world' decisions correctly. This same body of students scored higher in TEMPO-AI. Indirectly, this implies that the decision-making process in TEMPO-AI is at least partially analagous to real-world decision-making.

The test subjects themselves felt that TEMPO-AI was a useful learning aid. Of the few students who either disagreed or were neutral, most thought the game needed only one or two revisions. Despite this strong support from the intended users of TEMPO-AI, students' perceptions are of questionable use in testing validity of the game itself. For example, no correlation (+.15) was found between perceived performance and actual performance, but a very strong correlation (+.91) was found between perceived performance and perceived utility. Historically, test subjects tend to downgrade the validity of games where they perform poorly and elevate those where they perform well. However, since the majority of the students did not believe they had performed well, more credibility can assigned to their belief that the game was useful. Additionally, students who accept a game as valid are more likely to attempt to apply the lessons learned.

An examination of other factors that could possibly

affect game performance yielded mostly inclusive results. For example, while it is fairly clear that those students with 'hard' engineering degrees did not outperform those without, it is questionable whether this factor truly enhances managerial effectiveness in the first place. It possible that a technical background might influence TEMPO-AI performance, but that the Air Force definition of 'technical' excludes officers with considerable quantitative backgrounds. This issue could not be tested because of an inability to find officers with non-quantitative backgrounds willing to play TEMPO-AI.

Population differences were found between some of the other factors, but these results were either negative or not statistically significant. For example, acquisition experience, wargame experience, and TEMPO experience all seemed to enhance performance, but the differences were not significant at alpha = .05. A larger sample size might show significant differences. It did appear that the extent of experience in these areas was not relevant--only the fact of experience.

A surprising result was the negative correlation (-.25) between operational experience and game performance. On the surface, this would tend to contradict the validity of TEMPO-AI, since the Air Force considers operational experience to be a factor that enhances managerial ability. However, this result is most likely due to a <u>lack of experience</u> in other areas. For example, a chi-squared test

of homogeneity indicated that a greater proportion of the students with operational experience had not had advanced management education. Since advanced education was significantly correlated with game performance (+.40), the negative result for operational experience is suspect. Additionally, those officers with only operational assignments may lack the more general management experiences. For example of the ten officers who indicated they had 4 or more years of operational experience, nine had no previous experience in acquisition management--a factor that showed a positive correlation with game performance (+.24). Figure 3 presents a graphic representation of the difference in the mean scores of the different groups.

Even though TEMPO-AI appeared to have external validity, it still may not be useful as a learning tool. Pretests revealed that the game could not be played by an inexperienced student without considerable aid from a human instructor. The on-line documentation was inadequate and had to be supplemented with additional rules and a human game director just to run the experiment. TEMPO-AI lacks a final debriefing which is a crucial part of the learning experience in TEMPO. Without this debriefing, students may draw incorrect conclusions from their experience.

Perhaps an even greater hindrance to implementing TEMPO-AI as a successful learning aid may be the difficulty in getting students to play in the first place. Although most students enjoyed the game when they sat down to play,

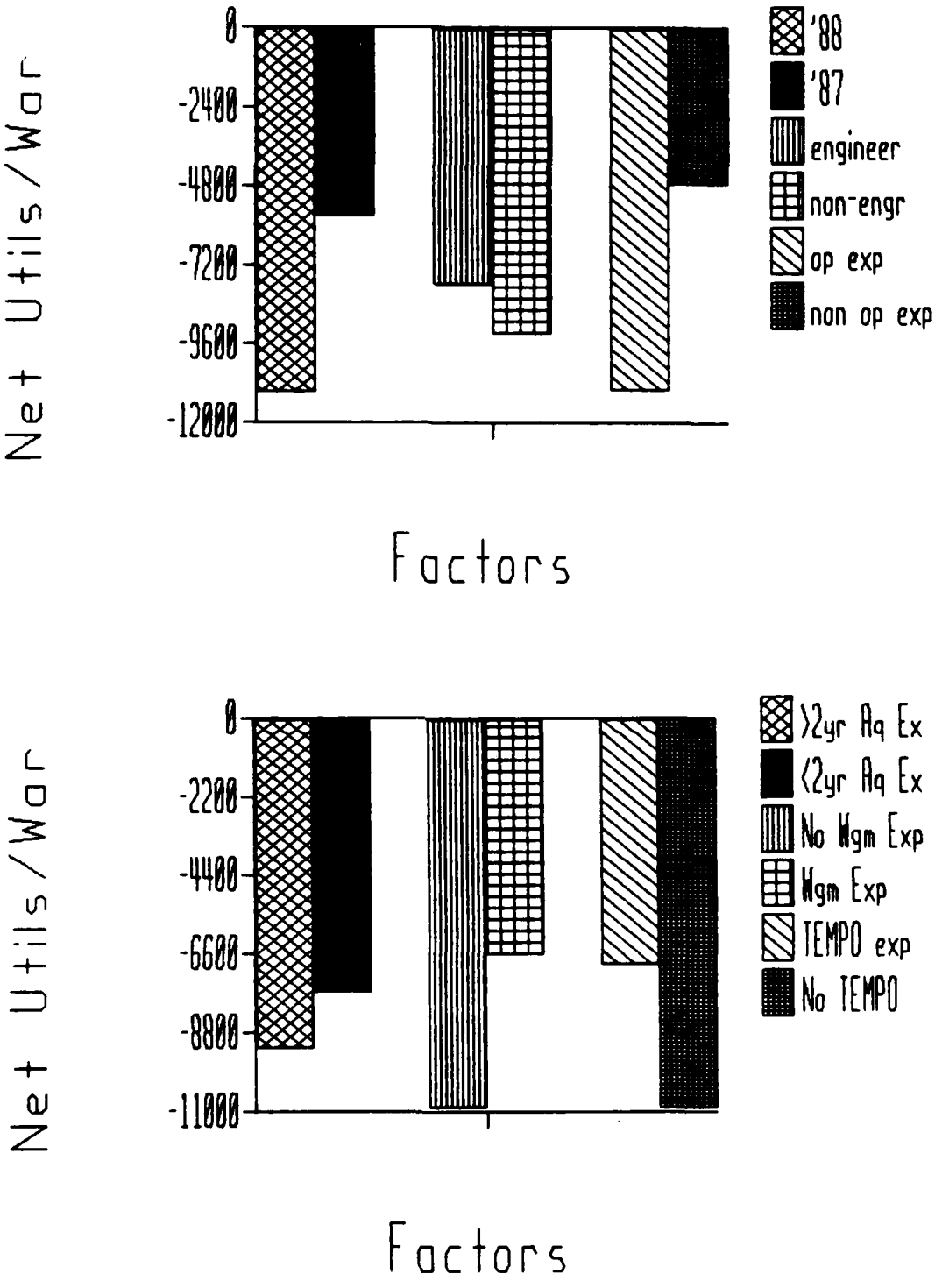


Figure 3. Mean Scores for Factors Affecting Game Performance

there does not appear to be much interest in voluntary participation. Yet this voluntary participation would be required for TEMPO-AI's success in a correspondence program. The initial response rate was 0.0 percent, and even after financial incentives were offered, less than 10 percent of the population volunteered. Student motivation is an obstacle which will have to be overcome before TEMPO-AI can be implemented as intended.

VI. Suggestions and Recommendations

Suggested Revisions to TEMPO-AI

The inadequacy of TEMPO-AI's playtesting was apparent in even the earliest stages of the pretest used to establish the test procedure used in this thesis. This stage of game development is absolutely vital, if the game is intended for serious academic use. Unfortunately, this important step is sadly neglected in nearly all military game development. While TEMPO-AI was extensively debugged as a computer program (i.e. it ran without run-time errors), it proved extremely awkward to use in its unaltered form.

This thesis has demonstrated that provision of a rules supplement can alleviate part of this problem, and is particular useful to former TEMPO players who might otherwise misunderstand the rules differences. In addition, a practice round is recommended for use in any game without a human game director. The practice round should be a short game-turn with no score. The computer's purchases and the player's purchases would then be compared on-screen, with a calculation of the net utils. After this practice game, the game would restart, and the players would have a better understanding of what was expected of them.

During the game, players should be allowed to enter a tentative budget and be given a total by the computer. As currently designed, TEMPO-AI does provide this service--but only after the budget is committed. Not only does this not help, but it causes considerable frustration among the

spends the budget eliminates any chance of that player recovering on the next round. By contrast, even a major loss in a war is not unrecoverable.

An argument could be made that management of a budget is an important part of a manager's job, and that TEMPO also requires that the team submit an accurate budget. However, this function tends to dominate TEMPO-AI and distracts from other, more important, objectives. In TEMPO, there is considerably more time to calculate the budget, and several people can check the bottom-line calculation. The penalty for overspending is minor, and doesn't influence the final result or the lessons learned. In addition, the final team orders are written on paper and can be reviewed any number of times. By contrast, TEMPO-AI offers no opportunity to review the orders, and a single keystroke can cause a major error. An opportunity to review the tentative budget in TEMPO-AI and make any desired adjustments would allow players to concentrate on the decision process and not the tedious task of entering correct numbers.

In a similar vein, TEMPO-AI should be revised to allow easy movement between the different screens of information. Currently, a player must step through eight screens of information to obtain one number. Four of these eight screens are only needed on the first turn, but must be seen again and again throughout the rest of the game. Some kind of expanded menu selection would make the game move more quickly.

A revision of TEMPO-AI should include a return to the generic labels for the weapon systems categories. Air Force labels such as 'Fighter,' 'Bomber,' add nothing to the game, and may confuse some players that have a preconceived notion of what that kind of weapon system should do. These seemed to be a particular problem among those students with actual operational experience, and could even have been a contributing factor to the poor performance of this group.

Finally, a basic change to the TEMPO-AI rules is suggested. There does not appear to be any value to the 'Balance of Forces' rule. It forces players to by a terrible weapon system in order to buy a desired weapon system in another category. This seems counter-intuitive, unless the author's purpose is to simulate Congressional or other political pressures. The boundary defined at a 2:1 ratio is not only arbitrary, but is too sharply defined--no utils are earned for systems bought over this boundary. The 'Diminishing Returns' rule of TEMPO also forces players to diversify, but is more natural. Each new system provides some value, but at a decreasing rate. The optimal buy is not clear, and is very dependent on the alternative choices. The result does force a 'Balance,' but the player has to evaluate all the other systems rather than just choosing a complement. Therefore, TEMPO-AI should be revised to reflect the original TEMPO rules on this subject.

Extensions of TEMPO-AI Research

The experimental procedure used in this thesis should be repeated after the recommended revisions are made to the TEMPO-AI computer program. If possible, a larger sample size should be obtained, but this will prove to be difficult without some kind of coercive authority. Ideally, TEMPO-AI should be tested in the classroom where the sample can be readily controlled, and the lessons learned augmented by lecture and/or discussion.

TEMPO-AI testing could be expanded to include tests of different potential causal factors. For example, other games have shown a correlation between Grade Point Average, comprehensive exam scores, and Aptitude tests. TEMPO-AI could be tested against these other measures of student performance. Of particular interest, would be the confirmation or explanation of the suspect association, found in this thesis, between operational experience and poor performance.

A particular promising line of research would involve multiple tests of TEMPO-AI with the same subjects. Improved scores would demonstrate that the concepts were comprehended and that the learning could be applied. The main obstacle to this line of research would be in obtaining a willing and representative sample. Preferably, this would be achieved in a classroom setting, and the progress of the test subjects tracked in a longitudinal study such as those done by Wolfe (11).

Final Recommendation

The success of military gaming depends on testing the external validity of the games that are developed. A disproportionate amount of resources are applied to game development, without the independent research necessary to ensure that these games fulfill their function.

This thesis has demonstrated a methodology that can be used to evaluate wargames, and raised the issues that need to be addressed in the instructional use of wargames. As a minimum, wargames require extensive playtesting before being implemented. This playtesting should be done not only by the original designers, but by independent researchers, knowledgeable experts in the field, and the target population. *Wargame* design is an iterative process, and there is no substitute for this step.

The need for instructional wargames has been recognized at all levels within the Department of Defense. The need for devoting serious academic attention to this subject has not. The civilian sector has made great progress in substantiating whole classes of business games. The military can, and should, follow suit. Only then can we ensure that the experiental learning that occurs in wargames will correspond to the problems that future managers will encounter in a time of crisis.

Appendix A: TEMPO Game Description (6)

The following material is quoted directly from the rules and suggestions for playing TEMPO:

General Description

Two teams are required for a play of the game. Teams start with identical forces and budget. The budget can be spent on (1) operation of existing forces; (2) procurement of additional forces; (3) research and development; (4) intelligence and counter-intelligence.

All weapons systems are divided into four classes: Offensive A, Offensive B, Defensive A, Defensive B.

Each weapon system is worth a certain number of 'utils' per unit. The 'util' is a measure of effectiveness which has been assigned to each system in order to simplify game play.[1]

In simple terms, the aim of the game is to maximize your team's net offensive utils, which are calculated as follows:

total Offensive A utils minus opposing team's total Defensive A utils

PLUS

total Offensive B utils minus opposing team's total Defensive B utils.[2]

EXAMPLE

OA_x = 2000 utils	OA_y = 1500 utils
DA_y^x = 2500	DA_x^y = 1000
----	-----
-500	500
OB_x = 1000 utils	OB_y = 1500
DB_y^x = 500 (= +100 from DA overdefense)	DB_x^y = 1200
-----	-----
400	300

Total net Offensive Team X = 400
Total net Offensive Team Y = 800
Winning util margin Y = 400

1. Utils are received only for forces currently being operated.

2. No credit is given for over-defending, i.e., defensive utils in excess of the offensive utils of the enemy.

3. Defensive A can only be defend against Offensive A and similarly for 'B' weapons.[3]

4. However, any Defensive A system counts against any Offensive A weapon, also B systems likewise. Thus, defensive utils in DB2 are counted as defensive against OB1, OB2, etc.

5. After a total of 2000 utils in any force unit type and number is obtained further utils are discounted in a sliding scale. Thus, if team X has 3500 utils in OA2, their actual util credit for that weapon is 3300.

6. New weapons systems do not displace or devaluate old systems. All units have the same util value throughout the play.

---

[1] Although determining the 'effectiveness' of weapons is often the most difficult part of military planning, this gross simplification permits the player to concentrate on budget allocation problems.

[2] In reality, the objective is more complex than this statement suggests. The game is played for an undetermined number of periods and maximizing utils for any one year will conflict with maximizing utils in other years. In addition, the game is an educational device and, therefore, the real aim is to learn something about military planning and limited budgets. Other complications will become apparent during the play of the game.

[3] However, 20% of defensive utils in one system over the number necessary to 'neutralize' the opposing team's offensive utils will be credited to the defensive posture of the other system when deciding the result of war.

Detailed Rules

1. Starting the game. At the start of play each team will be given:

   a. A number of Force Information Sheets (FIS). The first set of FIS provides a current inventory of four systems plus estimates of R&D costs and utils of new systems. ...

   b. Two copies of a Budget Allocation Form ... . One copy will be returned to the umpire at the end of the first period and one copy will be retained by the team.

2. Research and Development

   a. Each team will receive a first FIS on all new systems or modifications of old systems during various years of play. This first sheet will provide expected R&D, procurement and operating costs, and expected utils per unit. Note that all values are estimated and may change as R&D progresses.

   b. Additional R&D sheets will be provided only when a team completes the previous R&D.

   c. R&D may be discontinued at any time and resumed at a later date with a penalty payment of $300 or one-half of the last current R&D cost, whichever is the smaller.

   d. Information on costs and utils pertaining to the last year of R&D sheets or changes in information are to be expected.

3. Modifications. Some FIS will provide information about the possibility of modifying existing systems. A modification involves the following special rules:

   a. During the year of modification R&D, existing force units may be modified at the cost given in the FIS.

   b. The old units may be operated at their old costs and values during the modification year or moth-balled during modification.

   c. After modification R&D is completed, additional units of the modified system can be procured at a total cost equal to the procurement cost of the old system plus the modification costs for the modified version.

4. Procurement. A team may procure units of any system which is in inventory and any new system during the last year of R&D and thereafter at any rate not exceeding the maximum acquisition rate stated on the FIS.

5. Operation

a. A team may operate any or all forces in inventory at the start of a year. Units procured during one year are available the next year.

b. A team may operate units undergoing modification during the year of modification at the old costs and utils.

c. Force units not operated in any one year will be assumed to have been scrapped. (You cannot 'moth-ball' old units.) However, those units being modified in any year may be withdrawn from operation for that one year if desired.

6. Intelligence. Each team may procure intelligence about the posture of the opposing team in four catefories at a cost of $100 per category. The categories are:

a. Current changes in force structure of offensive forces.

b. Current changes in force structure of defensive forces.

c. Current changes in R&D programs of offensive forces.

d. Current changes in R&D programs of defensive forces.

In addition counterintelligence may be purchased at a cost of $200. When purchased this results in less accurate intelligence being given to the opposing team on force structure, offensive and defensive, if they purchased intelligence that year.

7. War. During each and every period of play there will be a probability of war. If war occurs, the results in terms of net offensive utils will be announced by the umpire.

8. Penalties

a. If war occurs, each team will have $400 cut

from its next year's budget.

b. The loser's budget will be cut by an additional amount equal to the difference between the two teams' net offensive utils. (See General Description.)

c. If a budget allocation sheet is not submitted exactly on time, the late team will be penalized at the rate of $50 per minute for the first five minutes, $100 per minute over five minutes. The penalty will be subtracted from the next budget.

d. A team that overexpends its budget will have its following budget cut by twice the amount of the overexpenditure.

e. Funds not expended in any one year are lost.

Appendix B: TEMPO-AI Game Description

GREETINGS!

Thank-you for signing up to play TEMPO-AI. TEMPO-AI is a military planning game. Other versions of TEMPO are multi-player, but in this version, it is you against the computer. You and the computer start with identical forces and budget. This budget can be spent on 1)operation of exisiting forces; (2) procurement of additional forces; (3) research and development (R&D); (4)intelligence.

All weapons systems are divided into four classes: Bombers, Missiles, Fighters, and ABMs. Bombers and Missiles are offensive. Fighters are a defensive system which counteracts Bombers. ABMs are a defensive system which counteracts Missiles.

Each weapon system is worth a certain number of 'utils' per unit. The 'util' is a measure of effectiveness which has been assigned to each system to simplify game play.

Your objective is to maximize your net offensive utils, while minimizing the computer's net offensive utils. Utils are received only for forces currently being operated during a war period. During 'peace' periods, you will be told whether or not you would have outscored the computer had a war occurred that year.

No credit is given for overdefending, i.e., defensive utils in excess of the offensive utils of the enemy. Unlike other versions of TEMPO, there is no carryover to other defensive categories.

In TEMPO-AI, you are penalized if you do not maintain a 'balance of force.' Balance of forces is achieved in offensive systems if the ratio of bombers to missiles is less than 2:1 and greater than 1:2. Similarly, Defensive systems are considered 'balanced' if the ratio of fighters to ABMs is less than 2:1 and greater than 1:2.

EXAMPLE: You have 17 Bombers and 30 Missiles. This means that offensive systems are 'balanced.'

EXAMPLE: You also have 40 Fighters and 100 ABMs. Defensive systems are therefore not 'balanced.' Only 80 percent of the ABMs will be counted. (ie twice the ratio of the smaller number to the larger.)

Unlike other forms of TEMPO, there is no 'diminishing returns' on any one weapon system. The 'balance of forces' rule replaces that restriction.

1. Starting the Game. The game director will get you to the first screen. The first screen will tell you your starting inventory, your current budget, projected budget, and provide information on existing weapons systems. You do not need to memorize this screen, as you can return to it as many times as you want. Scrap paper is provided for making notes, if you wish.

2. Research and Development.

a. After reading the information on current systems, you may proceed to read information on new R&D programs. All values are 'estimates' until the final year of R&D.

b. R&D may be discontinued at any time. If it is was discontinued after some R&D money was already spent, it can be picked up later, but an additional penalty may have to be paid.

c. Information on costs and performance in the last year of R&D can be treated as certain. Note that for one-year projects, the first year is the last year.

3. Modifications. A player has two choices to modify an existing system.

a. Pay the operation cost for th old system and the modification cost listed for that system.

b. Scrap the old unit and build a new modified unit from scratch. This is often the less expensive option in TEMPO-AI.

c. There is no 'moth-balling' of systems in while in mod.

4. Procurement. A player may procure units of any system which is in inventory and any new system during the last year of R&D provided that the maximum acquisition rate is not exceeded.

5. Operation.

a. A player may operate any or all of his/her forces in inventory at the start of a year and may operate those systems acquired during that year. The acquisition cost includes the operation cost for the first year.

b. Modified units immediately provide the utils at the modified rate.

c. Units not operated are scrapped.

6. Intelligence. Each player may procure intelligence about the posture of the opposing player in four categories at a cost of $100 each. The categories are:

a. Current # and types of offensive units.
b. Current # and types of defensive units.
c. Current offensive R&D programs.
d. Current defensive R&D programs.

e. There is no counterintelligence in this version.

7. War. During each and every period of play there will be a probability of war. This probability can be increased if there is a disparity in the numbers of offensive forces.

8. Penalties.

a. If war occurs, the loser's budget will be cut $800 and the winner's budget will be cut $200.

b. A 50 util penalty will be assessed for not finishing a turn on time.

c. Overspending the budget will result in the next year's budget being reduced by twice the difference. In addition, some systems will be randomly scrapped by the computer to bring the budget in line.

d. Money not spent is lost.

If the above seems a little complicated, don't be too discouraged. The computer will prompt you through most of what you need to do. It was meant to be run without any additional documentation, so this description is a 'supplement.' I will be present to answer rules questions throughout the game.

I have set a two-hour maximum time limit, although I suspect most players will be done in less than that. It helps to have a pocket calculator, and I will provide scrap paper. If you finish early, I have a very short survey about the game for you to answer, and then you are free to leave. Winners will be announced, and prizes awarded when I 'ring the bell.'

Please bring these rules with you to room 312. Good luck!

Captain Tony Russo

Appendix C: Post-Test Questionnaire

Player #_____ TEMPO-AI

1. Please write down your Class and Section (eg 87S-GSM, 88D-GIR, etc...)

2. Indicate which areas you are educated in. (Circle all that apply.)

a. Engineering
b. Sciences
c. Social Sciences
d. Education
e. Management
f. Other (Specify)

3. What is the highest degree you have obtained?

a. Bachelor
b. Bachelor with some graduate work (over 12 credits)
c. Masters
d. Masters with some post-graduate work (over 12 credits)
e. Doctorate

4. How many years of U.S. military experience do you have? (any service, include civilian time if working for the military)

a. 0
b. less than 1
c. 1-2
d. over 2 less than 4
e. over 4

5. How many years operational experience do you have? (any service or nationality)

a. 0
b. less than 1
c. 1-2
d. over 2 less than 4
e. over 4

6. How many years of defense acquisition experience do you have?

a. 0
b. less than 1
c. 1-2
d. over 2 less than 4
e. over 4

7. Do you have experience with wargames?

a) none b) some c) a lot

8. Have you played TEMPO before? If so, where?

Please answer the following questions by circling the number that matches how strongly you feel about the answer. Use the following scale:

1	2	3	4	5	6	7
strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree

9. I enjoyed making decisions in this game.

1 2 3 4 5 6 7

10. I think I performed well.

1 2 3 4 5 6 7

11. I put a lot of thought into each game-turn.

1 2 3 4 5 6 7

12. I understood the impact of my decisions.

1 2 3 4 5 6 7

13. I was confident of my decisions.

1 2 3 4 5 6 7

14. I considered this game to be realistic.

1 2 3 4 5 6 7

15. Even if I didn't think this game was realistic, I did feel that it is useful as a learning aid.

1 2 3 4 5 6 7

16. Any other comments?

Bibliography

1. Cohen, Kalman J., and Rhenman, Eric. "The Role of Management Games in Education and Research," Management Science, 7: 131-166 (January 1961).

2. Devore, Jay. Probability and Statistics for Engineers and the Sciences. Monterey: Brooks/Cole Publishing Company, 1982.

3. Dumond, Lt Col John., Assistant Professor of System Management. Personal Interview. Air Force Institute of Technology, Wright-Patterson Air Force Base, August 1986.

4. Greenlaw, Paul S., and Wyman F. Paul, "The Teaching Effectiveness of Games in Collegiate Business Courses," Simulation & Games, 4: 259-292 (September 1973).

5. Harris, J. F., and Nickelson, R. L., A Comprehensive Revision of the Logistics Planning Exercise. MS Thesis, LSSR 30-81. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1981.

6. Hatry, H., and others. TEMPO. Military planning game. General Electric Company, Santa Barbara CA, May 1962.

7. Lawrence, Lt Gen Richard D., "Playing the Game," Defense, pp 22-29 (Jan/Feb 1986).

8. Mauer, William A., Associate Dean, School of Systems and Logistics. Personal Interviews. Air Force Institute of Technology, Wright-Patterson Air Force Base, August 1987.

9. Pierfy, David A., "Computer Simulated Game Design: Stumbling Blocks and Stepping Stones," Simulation & Games, 8: 255-266 (June 1977).

10. Shubik, Martin. Games for Soceity, Business and War: Towards a Theory of Gaming. Amsterdam: Elseveir Scientific Publishing Company, 1975.

11. Vance, S. C., and Gray, C. F., "Use of Performance Evaluation Model for Research in Business Gaming," Academy of Management Journal, 10: 27-37 (January 1967).

12. White, Gregory B., Artificial Intelligence Concepts and the War Gaming Environment: A Case Study Using the TEMPO War Game. MS Thesis, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1986.

13. Wolfe, Joseph, "The Teaching Effectiveness of Games in Collegiate Business Courses," Simulation & Games, 16: 251-282, (September 1985).

14. Wolfe, Joseph, and Guth, Gary R., "The Case Approach versus Gaming in the Teaching of Business Policy: An Experimental Evaluation," Journal of Business, 48: 349-365, (July 1975).

15. Wolfe, Joseph, and Roberts, Richard C., "The External Validity of a Business Management Game," Simulation & Games, 17: 45-56 (March 1986).

VITA

Captain Anthony J. Russo was born on 4 October 1960 in Passaic, New Jersey. He graduated from high school in Bridgewater, New Jersey in 1978 and attended Lehigh University. In 1982, he graduated with a Bachelor of Science degree in Engineering Physics. Upon graduation, he received a commission in the USAF through the ROTC program. He served in several different positions at the 4950th Test Wing, Wright-Patterson AFB OH until entering the School of Systems and Logistics, Air Force Institute of Technology, in June 1986.

Captain Russo is married to the former Judith Ann Buckwalter of Chapel Hill, North Carolina. They have two daughters, Jaclyn and April, and are expecting a third daughter in September/October 1987. Captain Russo is a member of Sigma Iota Epsilon.

Permanent Address: 901 Park View Drive/Apt A-407
King of Prussia PA 19406

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

ADA186962

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; dist unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GSM/LS/87S-28	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Systems and Logistics 6b. OFFICE SYMBOL (If applicable) AFIT/LSY	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Institute of Technology (AU) Wright-Patterson AFB, Ohio 45433-6583	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

AN EVALUATION OF A MANAGEMENT WARGAME AND THE FACTORS AFFECTING GAME PERFORMANCE

12. PERSONAL AUTHOR(S)

Anthony John Russo, Captain, USAF

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
MS Thesis	FROM _____ TO _____	1987 September	94

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	War Games, Teaching Methods
12	04		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Thesis Chairmen: Dr William A Mauer
Associate Dean, School of Systems and Logistics

Approved for [illegible]
24 Sept 87
[illegible]
Air Force Institute of Technology (AU)
Wright-Patterson AFB, Ohio 45433-6583

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr William A. Mauer	22b. TELEPHONE (Include Area Code) (513) 255-6857
	22c. OFFICE SYMBOL AFIT/LS

DD Form 1473, JUN 86 — Previous editions are obsolete — SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

Block 19

ABSTRACT

Despite the long association between military organizations and the instructional use of wargames, virtually no research has been done to validate the concept of a wargame. The objectives of this thesis were to: 1)identify a need for serious academic research on the validity of management wargames, 2)establish a methodology for the objective and quantitative analysis of management wargames, and 3)apply that methodology to a specific management wargame.

TEMPO-AI, a computer version of the TEMPO wargame, was the specific wargame evaluated. The game, which is intended for use in a Squadron Officer School cooerespondence program, was played over 60 times and tested in a controlled experiment. The results of the experiment indicated that students that had nearly completed a Masters Degree program, scored significantly higher than those who were just entering the program.

Although some evidence for the external validity of the game model were obtained, the tests revealed numerous structural problems which might impede successful implementation of TEMPO-AI in a correspondence program. Some suggested revisions may alleviate some of these problems, but the biggest obstacle may be the unwillingness of students to play the game without special incentives.