Complexity-Based Gaming Approach to Improve Learning from Simulation Games [1]

Onur Özgün[2] and Yaman Barlas

Boğaziçi University

Department of Industrial Engineering, 34342 Bebek, Istanbul, Turkey

Tel: +90 212 3597343, Fax: +90 212 2651800

E-mail: onur@onurozgun.com.tr, ybarlas@boun.edu.tr

Abstract

This study investigates whether a procedure in which games are played in an increasing order of complexity can improve game performance, conceptual learning, and transfer of learning. Using controlled experiments, we test whether playing simpler versions of a game in increasing complexity improves performance and learning as compared to playing the simpler versions in random order, or repeatedly playing the same complex game without any change in complexity. The results are not in favor of gradual complexity increase in terms of performance and learning, indicating that it is not straightforward to establish a gradual-increase-in-complexity method for improving performance and learning, due to subtleties related to task structure, game procedure and cognitive effects of the playing sequence. Subjects perform slightly better when they are first introduced with relatively simpler versions of a task, and when the complexities of consecutive games are close. Probable factors behind these results are discussed. In depth analysis of factors causing these results is a potential further research topic.

Keywords: simulation games, systemic complexity, learning

1 Introduction

Simulation games have many advantages that make them popular tools for learning. First, being simulation models, they are simplified representations of the complex systems, only including the essential components. They allow repeated experimentation, with compressed time and space, hence providing more direct feedback to the player. Also, they are interactive, allowing the player to get involved rather than to observe. And after all, games are fun to play, which brings a motivational advantage over other teaching tools.

Although they are simplified representations of real systems, learning from simulation games can be limited, since they usually still contain systemic complexity factors such as delay, nonlinearity and feedback loops. The extent of learning is even more limited when we do not only mean gaining the necessary skills to achieve a good performance in a game (which is called *procedural learning*), but also meaningful, transferrable information acquisition toward managing the real problem the game represents (*conceptual learning*) (Größler *et al.*, 2000).

[1] Research supported by Boğaziçi University Research Grant no 09HA301D

[2] Current Address: Northeastern University, School of Public Policy and Urban Affairs, 360 Huntington Ave. 310 RN, Boston, MA 02115, USA

Based on different scientific theories, different procedures have been proposed to enhance learning from many disciplines such as computer science, media and cultural studies, psychology, education, physics and youth studies (Kirriemuir and McFarlane, 2004). In this paper, we approach the learning problem from a systemic complexity perspective, and analyze whether it is possible to enhance learning by various complexity-based gaming procedures.

The effectiveness of gaming experience depends on many factors about the game itself, as well as the gaming procedure, which encompass a range of issues such as content and method of instructions, number and separation between trials or the reward given to the players. However, systemic complexity is a critical component, given the fact that factors such as delay, nonlinearity and feedback loops are major barriers of learning.

Complexity-based sequencing of tasks for improving learning is a well-known method in experiential teaching literature (Van Merriënboer *et al.*, 2003; Salden *et al.*, 2006). Similar approaches can also be found in simulation-based learning literature. To assist developing the complex mental model required by a complex task, White and Frederiksen (1990) introduced the idea of model progression. Many studies have adopted such gradually increasing complexity methods for simulation-based learning environments, in which gradual model progression is found useful over repetitive exposure to simple or complex versions, in terms of transfer of learning (Alessi, 1995) and conceptual learning (Swaak *et al.*, 1998; Mulder *et al.*, 2011). Also, Yasarcan (2010) demonstrated that by playing simpler versions of a complex simulation game, it is possible to further improve game performance after the performance progress stops after certain number of repetitions. However, in some cases, model progression found ineffective on learning. For example, De Jong *et al.* (1999) found no difference between the subject group playing a game sequence involving five-levels of increasing complexity and the group playing only last two levels of these five levels. In another study, Quinn and Alessi (1994) found that the strategy of breaking the simulation into sections of increasing complexity to be less efficient than presenting the overall task initially.

In many cases, the gradual complexity increase is determined by non-systemic complexity elements such as game speed (Nurmi and Lainema, 2002) or amount of information displayed De Jong *et al.* (1999). Even in the cases where the complexity increase is due to systemic variables such as number of stocks and feedbacks (Swaak *et al.*, 1998; Mulder *et al.*, 2011) or delay duration (Yasarcan, 2010), there is no supporting evidence that the levels are distinct in terms of their systemic complexity.

2 Research Design and Hypotheses

This study compares three different gradual complexity increase procedures against two control groups in terms of subjects' procedural learning, conceptual learning and transfer of learning. A stock management game is used as the task environment. The complexity levels are selected based on a previous study carried out on the same game (Özgün and Barlas, 2012), to obtain a clear complexity progress.

Table 1 presents the experimental design. Simple, moderate, challenging and complex levels represent main complexity levels, which have been demonstrated to be significantly distinct in terms of average scores obtained in these versions. Moderate-to-challenging

Table 1: The experimental design for the learning-oriented experiments.

Trial	Exp. Group 1	Exp. Group 2	Exp. Group 3	Control Group 1	Control Group 2
1	complex	complex	moderate	complex	complex
2	complex	complex	moderate	complex	complex
3	complex	complex	moderate-to-challenging	complex	complex
4	simple	moderate-to-challenging	moderate-to-challenging	complex	challenging
5	simple	moderate-to-challenging	challenging	complex	simple
6	moderate	challenging	challenging	complex	moderate
7	moderate	challenging	challenging-to-complex	complex	simple
8	challenging	challenging-to-complex	challenging-to-complex	complex	challenging
9	challenging	challenging-to-complex	challenging-to-complex	complex	moderate
10	complex	complex	complex	complex	complex
11	complex	complex	complex	complex	complex
12	modified version	modified version	modified version	modified version	modified version
13	modified version	modified version	modified version	modified version	modified version
14	different game	different game	different game	different game	different game
15	different game	different game	different game	different game	different game

and challenging-to-complex are intermediate levels created to provide smooth transition between complexity levels. There are five experimental conditions: three experiment groups and two control groups. Each group consists of ten subjects. The experiment groups are designed to test the effectiveness of three alternative game sequences, all of which involving a gradual increase in complexity.

The players in the first experiment group play the complex game three times at the beginning, to familiarize themselves with the game and to allow stabilization of game scores as a result of procedural learning by repeated trials. Then, they play three relatively simple versions of the game twice, each in an increasing order of complexity. At the end, they again play the complex games, twice. After completion of 11 games, the players play a modified version of the stock management game to test vertical transfer of learning, i.e. transfer of learning from a simpler game to a more complex game. This modified version is created to be more complex than the complex game. In addition, subjects complete a questionnaire to test their conceptual understanding of the underlying system. Finally, they play another simulation game called Scuba Diving Simulator (Barlas and Dalkiran, 2008), which is essentially a complex and non-linear stock management game. The objective of this experiment is to test horizontal transfer of learning, i.e. transfer of learning to a game with similar structure and complexity, but with a different cover story.

Second and third experiment groups are variations of the first experiment group. In the second experiment group, the simpler versions played in fourth-to-ninth trials start from a higher level of complexity, and continue with smaller increments. Possible advantages of this procedure are that smaller increments can yield better transfer between trials, and playing a more similar game to the complex game before trials 10 and 11 can improve performance in the complex game. In the third experiment group, the initial three complex games are removed to eliminate possible adverse effects of starting with complex games, and to open up space for a simpler (moderate) version to start.

Among two control groups, the first one is to make sure that any performance improvement in the last two complex games of the experiment groups is beyond the effect of repeated trials. If the performance of the experiment group subjects turns out to be superior to the performance of first control group subjects, then we can claim the strong conclusion that the increasing-complexity games significantly contribute to performance improvement. Similar comparisons are made for questionnaire results and the two additional games performances.

The second control group is to verify that the order of complexity sequence makes a difference. The subjects play the same games with Experiment Group 1, but in a non-increasing, random order of complexity. If subjects performances turn out to be inferior to the performances of Experiment Group 1 subjects, then this will prove the effectiveness of increasing complexity order as a learning procedure.

There are four performance measures that represent different types of learning: (1) average game scores in 10^{th} and 11^{th} games that measure procedural learning, (2) questionnaire scores that measure conceptual learning, (3) average game scores in 12^{th} and 13^{th} games that measure vertical transfer, and finally (4) average score in 14^{th} and 15^{th} games that measure horizontal transfer of learning.

Our hypothesis is that experiment groups will perform better in terms of conceptual learning, as well as vertical and horizontal transfers of learning. On the other hand, first control group is expected to be better in terms of procedural learning. Second control group is expected to be worst in terms of all performance measures.

3 Game Description

We use a typical stock management game. The subjects play the role of a production manager who is responsible for t-shirts. Their objective is to bring *inventory* level to a target level as soon as possible and keep it there, by entering their *desired production*. In some game versions, due to decisions of higher-level hypothetical production planners, *actual* production may be different than *desired production*, due to capacity constraints. The *inventory* stock grows by *production* and diminishes by *sales*. *Sales* is normally distributed around a constant mean (*base sales*) with a mild standard deviation, unknown to the player. Initially, the game is not in equilibrium; the *inventory* level is higher than *target inventory*, and *production* is higher than *base sales*. The time unit of the model is *days* and $dt = 1$. The time horizon is 40 days. The subjects know the general structure of the model, but they do not know the parameter values (See Appendix A for the user interface and the game description seen by the subjects).

The performance is measured by *relative deviation from target*, defined as:

$$\textit{Cumulative deviation from target} - \textit{Benchmark's cumulative deviation from target} \quad (1)$$

Benchmark behavior is defined as the best possible decisions yielding the minimum total inventory deviation from the target level. By subtracting the benchmarks cumulative deviation, we make sure that different game version results are comparable.

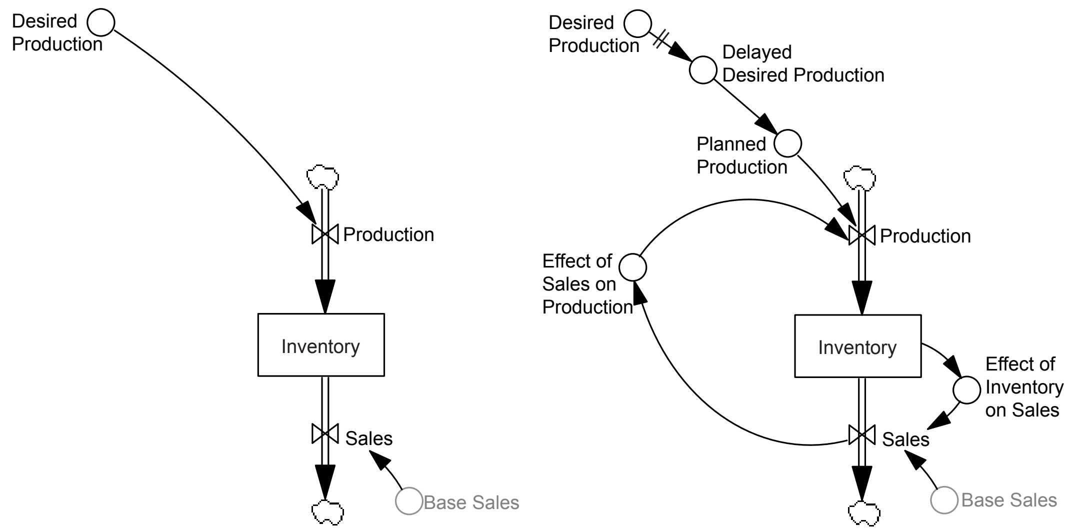
Different complexity levels are formed by varying the number and strength of systemic complexity factors in the game (Table 2). Three factors are varied: delay (order and duration), nonlinearity, and strength of feedback loop. The specific levels of each complexity factor present in each game version are determined as a result of an earlier study (Özgün and Barlas, 2012).

Figure 1 presents the structure of two game versions: simple and complex game. In the simple game there is no systemic complexity factor. The only challenge in the game is the unknown *sales*. Moderate game is obtained by adding *delayed desired production* with a first-order, 2-day delay. When there is delay, *inventory* typically shows oscillations around target, which become larger with increased delay duration.

Table 2: Systemic complexity factors present in the game versions.

Game version	**Systemic complexity factors**
simple	—
moderate	First-order, 2-day delay
moderate-to-challenging	First-order, 2-day delay, and moderate nonlinearity
challenging	Third-order, 4-day delay, and moderate nonlinearity
challenging-to-complex	Fourth-order, 6-day delay, high nonlinearity and moderate feedback
complex	Fifth-order, 7-day delay, high nonlinearity and strong feedback
modified version	Discrete-order, 8-day delay, extreme nonlinearity and extremely strong feedback

Starting with moderate-to-challenging game, we introduce nonlinearity, by adding *planned production*, which is a nonlinear function converting *delayed desired production* to *production*. This nonlinearity by itself does not have a significant effect on game scores. Therefore, to make sure challenging game is significantly complex than the moderate game, we simultaneously increase the order and duration of the delay. Finally, we introduce a positive feedback loop by making *sales* dependent on *inventory*, and *production* dependent on *sales*. The strength of feedback determines the extra *production* automatically introduced as a result of increased sales, in addition to the players decisions. Introducing feedback to the game with delay creates a further deterioration in the game scores. The formulation details and exact definitions of systemic complexity factor levels can be found in the Appendix B. The game files are in the supporting material.



(a) Simple game

(b) Complex game

Figure 1: The structure of the stock management game for two different versions.

4 Procedure

Subjects are volunteer undergraduate and graduate students. 29 of them are female and 21 are male. Half of the subjects are industrial engineering students (25 out of 50). The remaining subjects come from diverse backgrounds. Students are distributed to the experiment and control groups so as not to bring any bias in the results.

Subjects are given a written instruction about the game rules (See Appendix A). Before starting the experiment, they play a trial game to familiarize themselves with the game interface and controls. Then, they sequentially play first 12 games. At the end of 12 games, they fill out a paper-based questionnaire that measures their understanding about the game structure (Appendix C). Next, they play a modified version of the game. Before starting the Scuba Diving Game, they are given a short description of this game and they play a trial game. They complete the experiments by playing the Scuba Diving Simulator twice. Subjects are given a monetary reward based on their relative performances compared to other subjects that play the same game sequence. The reward is largely based on their average performance in games 10 and 11 (70%). Their average score in other games is taken into consideration with a smaller weight (30%).

5 Analysis of Results

First, we analyze the first three games to understand whether there is a significant difference between subject pools of four groups in which the complex game is played at the beginning. As Figure 2 shows, the scores of Control Group 1 are slightly higher (worse) than the scores of other three groups. This is due to three poor scores in Control Group 1, all belonging to the same subject. Further analysis revealed that the subject had difficulty in understanding the game objective and gave almost random decisions Thus, as an outlier, we removed this subject from following analysis. After this removal, there are nine subjects in Control Group 1. The worst score in Experiment Group 2 is also extremely high compared to other scores in that group. We replaced this score by the average of first two trials of the same player. Finally, we also removed one player from Experiment Group 3, who consistently displayed extremely bad performances even in the simplest games. This removal left nine players in this group.

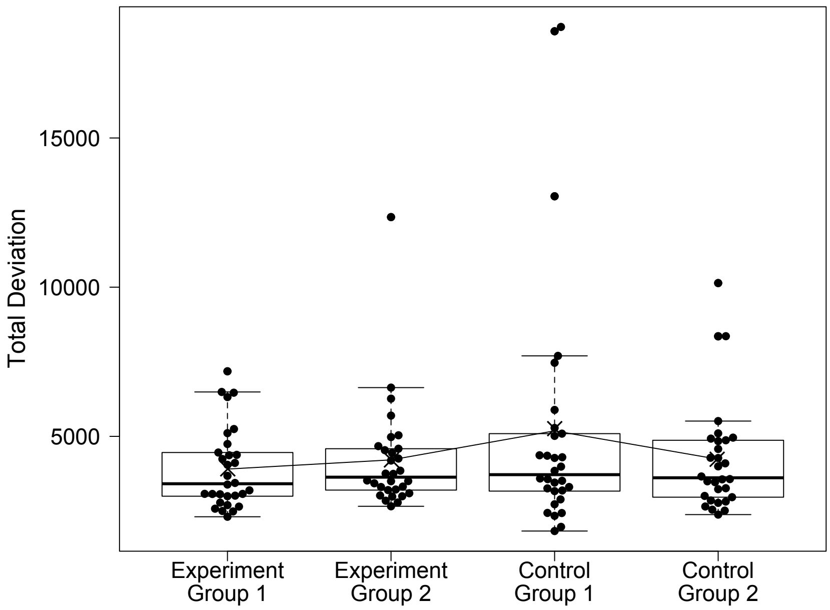


Figure 2: Game scores in the first three trials. Dots represent scores of individual players. Box plots show the distribution of the scores. The line connects the mean values.

Table 3: p-values for pairwise differences between subject groups, of trial 10–11 scores.

	Exp. Gr. 2	Exp. Gr. 3	Control Gr. 1	Control Gr. 2
Experiment Group 1	0.5277	0.2698	0.0110	0.0022
Experiment Group 2	—	0.0764	0.0002	0.0009
Experiment Group 3	—	—	0.0338	0.1679
Control Group 1	—	—	—	0.1651

Next, we analyze the complex game scores from trials 10 and 11, in order to see the effect of playing with simpler games in trials four to nine. Figure 3 shows the scores of five groups. Table 3 presents the p-values for the score differences between these groups. Among three experiment groups, Experiment Group 3 yields best performance, but the differences are not statistically significant. This shows that starting from simpler games and gradually increasing the complexity can be somewhat more helpful than exposing the subjects to the complex game at the beginning. The performances in Experiment Groups 1 and 2 are significantly worse than both control groups, whereas there is no statistical difference between Experiment Group 3 and the control groups. The best scores are obtained in Control Group 1. This may not be a surprising result since subjects in Control Group 1 play the same (complex) game through the experiment, and hence acquire better procedural learning. However, the difference between the experiment groups and Control Group 2 is surprising. Common sense suggests that gradual increase in complexity should lead to better performance compared to a random gaming sequence. However, the results suggest the opposite. We can see the effect of the difference between two game sequences by comparing Experiment Group 1 and Control Group 2. These two groups play exactly the same games, but in different order, yielding statistically significantly different results.

To make sure that the performance difference in trials 10–11 is not due to player characteristics, we check the performance change from the first three trials to trials 10–11 within each group. Figure 4 shows the progress of average scores through trials for five subject groups. Experiment Group 1 and both control groups show improvement from the first to third trial. The performance of Experiment Group 2 subjects deteriorates in third trial, without any explainable cause. In Control Group 1, where the subjects continue to play the same game, the scores continue to improve quickly until trial six. Control Group 2 also shows

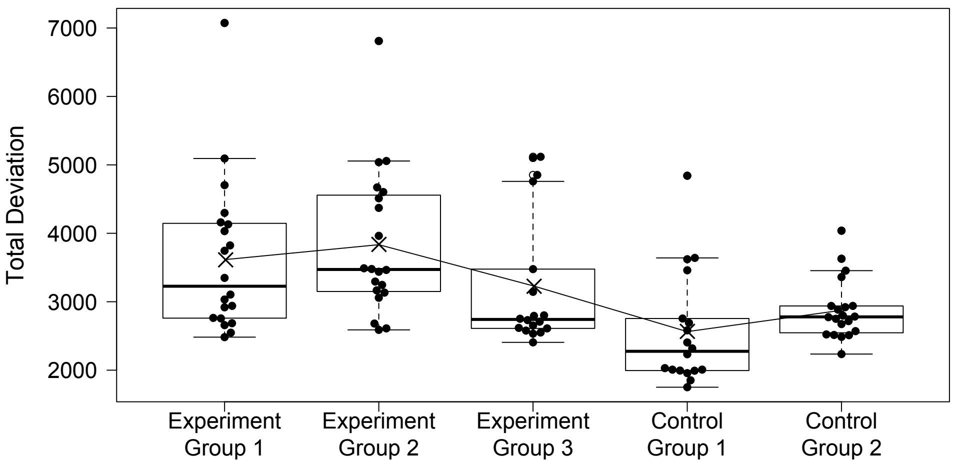


Figure 3: Game scores in trials 10 and 11.

a clear progress from first three trials to trials 10–11. However, in Experiment Groups 1 and 2, there is no or very little performance improvement from third trial to tenth trial. Not only the experiment groups perform worse in trials 10 and 11, they also show worse progress compared to their first three trials. Figure 5 presents the differences of subjects scores in trial 10–11 from their average scores in their first three trials. Experiment groups improvements are significantly worse than control groups improvements (p-values ≤ 0.015).

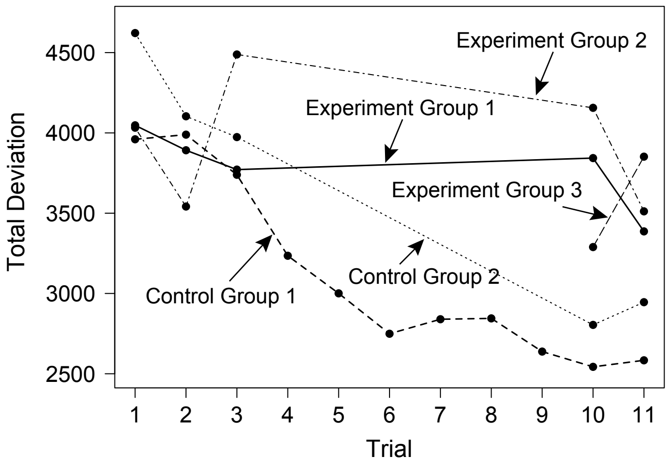


Figure 4: Progress of average scores through trials for three experimental groups.

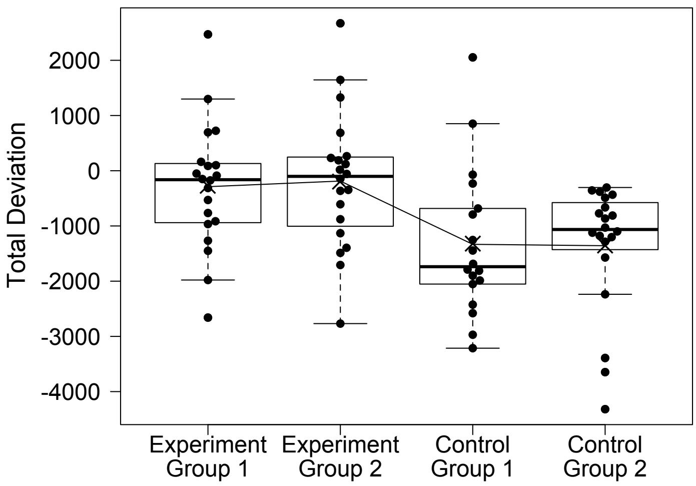


Figure 5: Differences between player scores in trials 10 and 11 and players average scores in their first three trials.

To gain insight about possible causes of the unexpected result that the random playing order yields better performance than the increasing order of complexity, we check the time behaviors of *inventory* for two groups: Experiment Group 1 and Control Group 2. Figure 6 shows the average behavior of *inventory* subjects from these two groups in their first three trials, as well as in trials 10 and 11. In the first three trials, both groups subjects understand that their decisions are delayed. Therefore, after giving low *desired productions* in the beginning to reduce the *inventory*, they increase their *desired productions* well before *inventory* reaches the target. However, the average behaviors show that subjects react early and/or more than necessary. After playing six simplified versions, both subject groups play

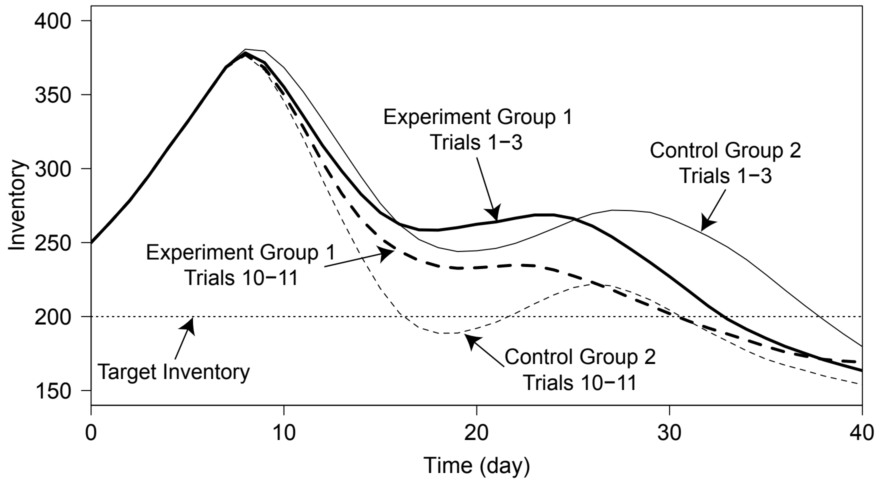
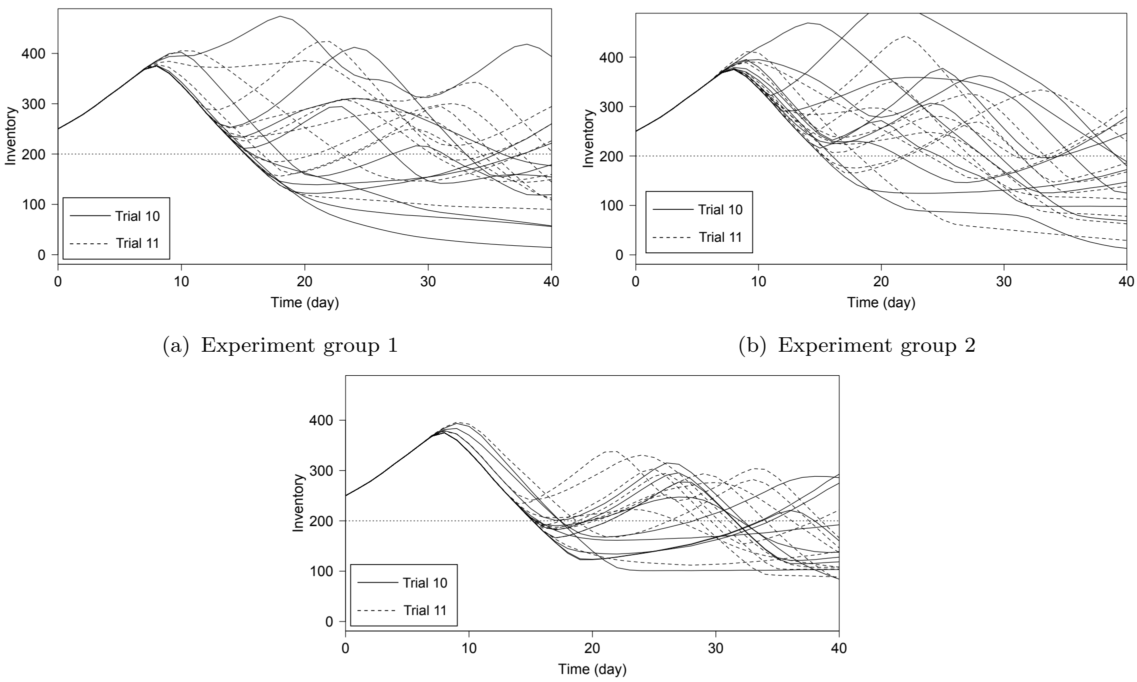


Figure 6: Average time behaviors of *inventory* for Experiment Group 1 and Control Group 2 subjects.

the complex game in trials 10 and 11. Control Group 2 shows more improvement, and on the average, the subjects have better timing of their decisions. Although performance of Experiment Group 1 also improves, the improvement is not as large as the improvement of Control Group 2.

Figure 7 shows individual subjects *inventory* behaviors in trials 10 and 11 for three groups. Control Group 2 not only exhibits a better average *inventory* behavior, but also individual subject performances are better. The variances in the experiment groups are larger.



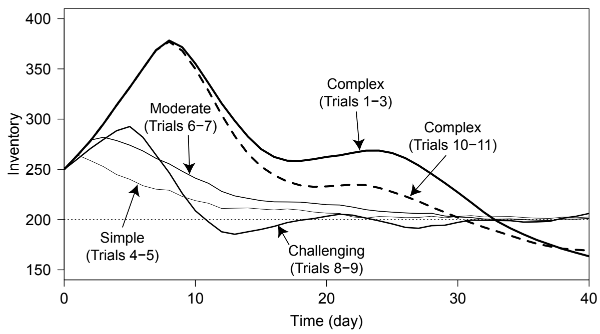
(a) Experiment group 1

(b) Experiment group 2

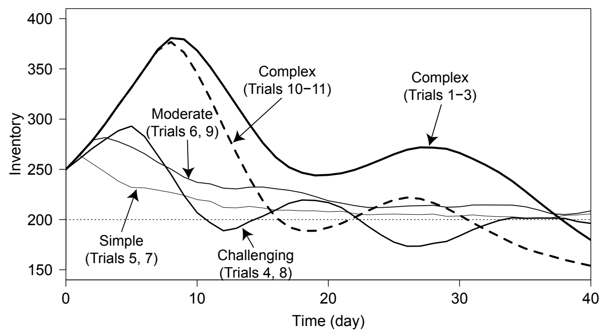
(c) Control group 2

Figure 7: Subjects' *inventory* behaviors in trials 10 and 11.

Figure 8 shows the average *inventory* behaviors for simpler game versions for the above-mentioned two subject groups. Recall that Experiment Group 1 subjects play three simplified versions in an increasing order of complexity, while Control Group 2 subjects play the same games in a shuffled order. Experiment Group 1 subjects know that games will be increasing complexity order, whereas Control Group 2 does not have any idea of the specific ordering of games. Subjects performances in the simpler games are comparable (Figure 9). Although the performances in simpler versions are not very different, the intervening games may have different cognitive effects in the two subject groups.



(a) Experiment group 1



(b) Control group 2

Figure 8: Subjects' average *inventory* behaviors in different game versions.

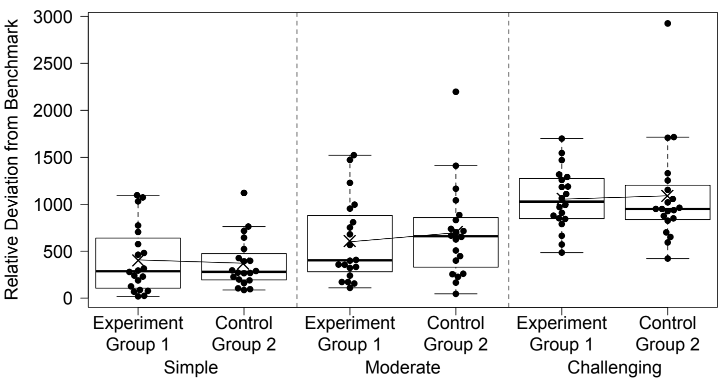


Figure 9: Relative deviation from benchmark in different game versions (trials 4–9).

Figures 10, 11 and 12 show the progress of scores in three experiment groups. It is evident that from the figures that complexity of the consecutive games has strong influence on the performance. For example, moderate-to-challenging game scores in Experiment Group 3 are much better than the same versions scores in Experiment Group 2, probably because they are played after easier games. On the other hand, Experiment Group 2 shows an improvement through trials four-to-nine although the complexity increases, but the trend is not followed when complex game is introduced. Experiment Group 3 scores follow a deteriorating trend that is parallel with the complexity of the games, and reaches (naturally) worst level when the complex game is reached. However, the scores in the complex games of Experiment Group 3 are the best in three experiment groups.

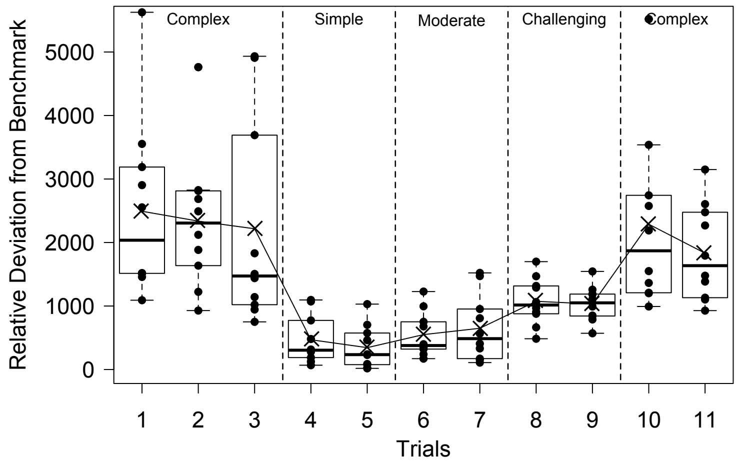


Figure 10: Progress of scores in Experiment Group 1.

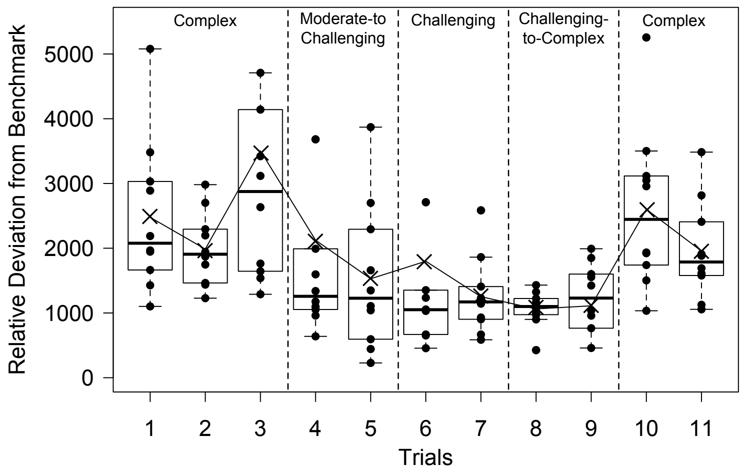


Figure 11: Progress of scores in Experiment Group 2.

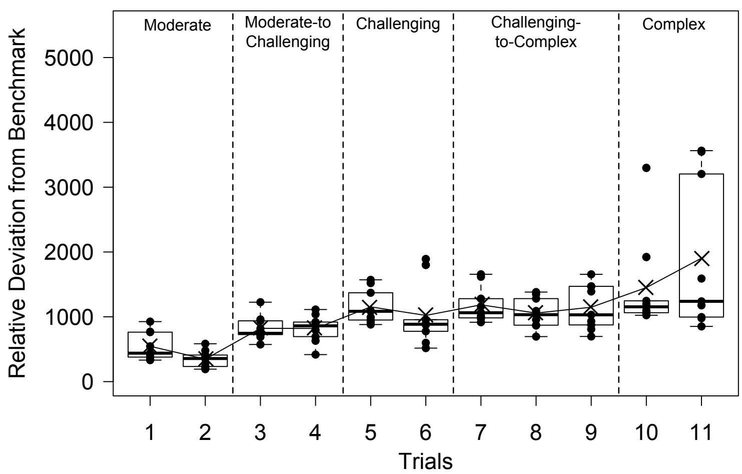


Figure 12: Progress of scores in Experiment Group 3.

Until now, we analyzed the effect of gradual-increase-in-complexity approach on procedural learning. Now, we analyze whether the approach helps transfer of learning. We test vertical transfer by comparing subjects performances in a modified version of the stock management game, which is more difficult than the complex game. Figure 13 shows that there is no significant difference between five subject groups (Table 4). Horizontal transfer is tested by analyzing subject groups performances in the Scuba Diving Game (Barlas and Dalkiran, 2008). As Figure 14 demonstrates, the performances in the Scuba Diving Game are not much different. Only, the scores of Experiment Group 2 and 3 are significantly worse than Control Group 2 (Table 5).

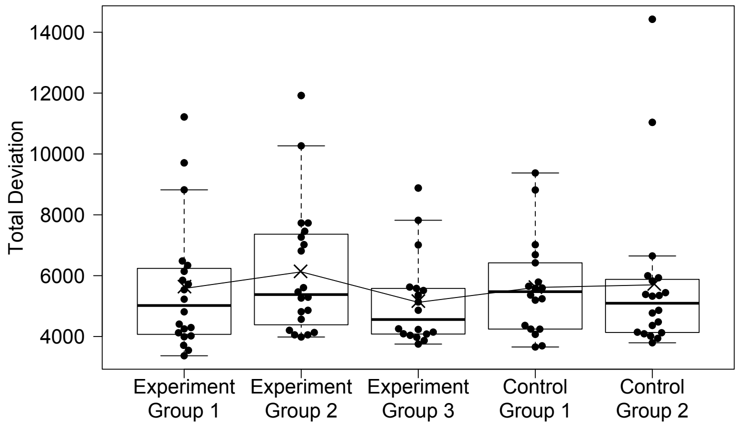


Figure 13: Game scores in modified game version (trials 12 and 13).

Table 4: p-values for pairwise differences between subject groups, of modified game scores.

	Exp. Gr. 2	Exp. Gr. 3	Control Gr. 1	Control Gr. 2
Experiment Group 1	0.4254	0.4446	0.9564	0.8748
Experiment Group 2	—	0.1010	0.4087	0.5747
Experiment Group 3	—	—	0.3468	0.4019
Control Group 1	—	—	—	0.9024

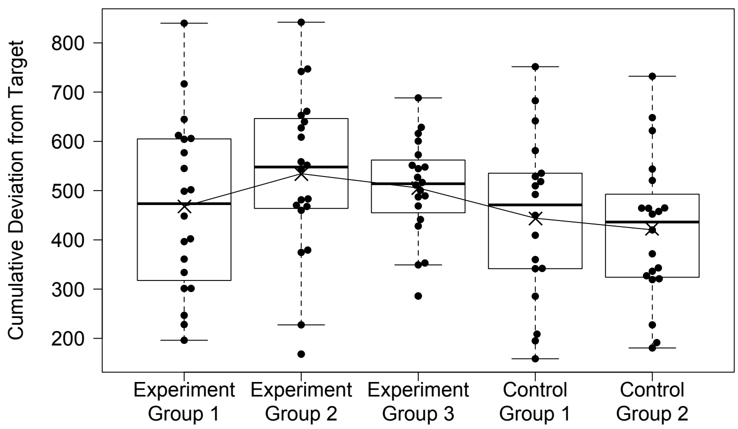


Figure 14: Game scores in Scuba Diving Game (trials 14 and 15).

Table 5: p-values for pairwise differences between subject groups, of Scuba Diving Game scores.

	Exp. Gr. 2	**Exp. Gr. 3**	**Control Gr. 1**	**Control Gr. 2**
Experiment Group 1	0.2301	0.4142	0.6701	0.3587
Experiment Group 2	—	0.5130	0.1087	0.0286
Experiment Group 3	—	—	0.1911	0.0402
Control Group 1	—	—	—	0.6522

Finally, to assess subjects conceptual learning, we analyze the correct answers in the questionnaire given to the subjects after 11 games (Figure 15). The number of correct answers given by the experiment group subjects is significantly lower than that of Control Group 1 (p-value ≤ 0.07). The other differences are not significant. These results show that gradual-increase-in-complexity provides no advantage for conceptual learning of the stock management task, compared to playing the same complex game in all trials.

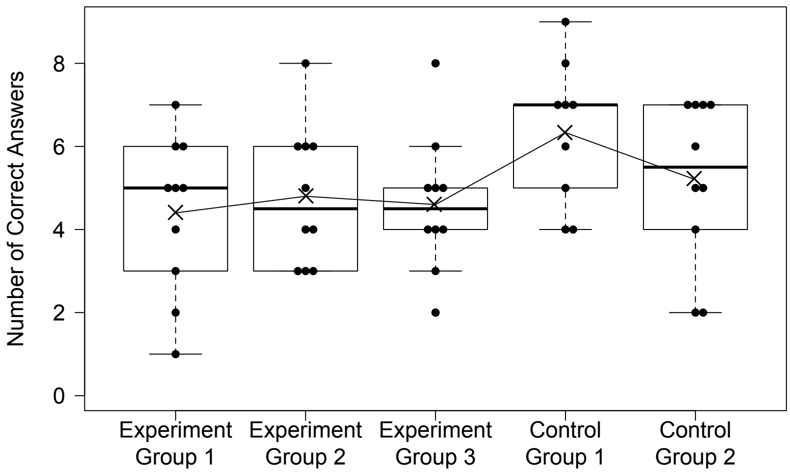


Figure 15: Number of correct answers in the questionnaire.

6 Discussion

In this paper, we tested the effectiveness of the gradual-increase-in-complexity approach on game performance, conceptual learning, and transfer of learning.

While subjects performances in the experiment group improve by playing six games with gradual complexity increase, the performances in the control group improve *more* by playing the same six games in a shuffled order. There may be different explanations for this unexpected result.

First possible explanation is that although the game versions are selected after careful analysis of previous results, the games are not "really" in increasing order of complexity. However, subjects performed significantly better in "simpler" games. Therefore, there is evidence supporting the increasing complexity order of the games.

Second possible explanation is the effect of subjects. Perhaps the subjects assigned to the experiment groups have relatively poor systemic-dynamic skills? We analyzed subjects backgrounds and found that there are no noticeable differences between Experiment Group 1 and in Control Group 2. The performances in the first three identical games also indicate no significant difference between subjects. Thus, we conclude that it is very unlikely that subject difference can be responsible for the unexpected results.

Third possible explanation is related to cognitive effects of the overall sequence of simpler games in two subject groups. Both Experiment Group 1 and Control Group 2 subjects are told that they will play six simplified versions of the first three games in trials four to nine. Experiment Group 1 subjects are further told that they will play these six games in an increasing order of complexity, while Control Group 2 subjects are told that they will play six games in a random order of complexity. One hypothesis is that since Control Group 2 subjects do not know the order of games, they expect a higher complexity in all six games, which allow them to easily adapt to the complex game in trials 10 and 11. In the meantime, gradual complexity increase in the experiment group may have possible cognitive side effects. They may learn better to control simpler games, and have hard time in adapting back to the complex game. By failing to adapt to the complex game, subjects may undershoot the target.

Fourth possible explanation is related to the nature of the game just before the complex game in trial 10, i.e. trial nine. Experiment Group 1 subjects play a challenging game in trial nine, while Control Group 2 subjects play a moderate game. Challenging game has a third order four-day delay coupled with nonlinearity, whereas the moderate game just involves first order two-day delay. As the challenging game involves longer delay, it yields some oscillations around the target inventory. Thus, Experiment Group 1 subjects may have been over cautious in playing the complex game (trial 10). At the same time, since Control Group 2 subjects play a moderate game in trial nine, they may have been more relaxed about the delay. Third and fourth explanations are both plausible, and require further experimentation for verification.

In order to find a gaming procedure involving gradual-increase-in-complexity that enhances learning, we tried three different designs. The results of these three experiment groups yielded interesting results regarding the sequence effect of the games. Compared to Experiment Group 1, the simpler versions used in fourth to ninth trials of Experiment Group 2 are closer to each other, with smaller increments from trial to trial. It is observed that while scores in Experiment Group 1 deteriorate through these trials, Experiment Group 2 scores improve. These observations suggest that smaller increments can better help learning by experimentation. Yet, continued improvement in the simpler versions of Experiment Group 2 is not transferred to the complex game in trials 10–11. This may be due to complexity gap between trial nine and 10.

In Experiment Group 3, we eliminated initial three complex games. It turned out that subjects simpler game performances in Experiment Group 3 are better compared to other experiment group. Also, their complex game performances are superior to other experiment groups. However, it is just barely as good as Control Group 2 subjects performances. The gap between ninth to tenth games may be influential on this result.

Based on these observations, there are possible designs that can potentially yield better results with a gradual-increase-in-complexity approach. Increasing repetitions at each step might help better learning. This may be achieved by either increasing overall number of trials, or decreasing number of simpler games used. At the extreme, only one simple/moderate game with many repetitions can be used. Another design that can be useful is to choose even smaller complexity increments between game versions. Again, this might be done by increasing the total number of trials, or at the expense of decreasing the number of repetitions at each step.

One important objective of this experiment was to see if gradual-increase-in-complexity can yield improved conceptual learning and transfer of learning. However, subjects playing the same complex game in all trials showed the best performances and gave more correct answers to the questionnaire. This result may be trivial and not useful in real life, since it suggests repeatedly performing the actual task desired to be learned, hence begging the question.

We tried to measure the *conceptual learning* of subjects by a questionnaire that measure subjects understandings about the stock-flow structure, as well as delay, nonlinearity, and feedback involved in the task. Control Group 1 subjects correctly answered more questions than all other groups. One possible explanation is that Control Group 1 subjects had more chance to experience with the complex game about which the questions are. The results can be interpreted in two ways: either playing the same game leads to better conceptual learning, or the questionnaire is not an adequate measure of conceptual learning. The only way to resolve this issue is to test with alternative measures of conceptual learning. These alternative measures may include drawing causal relations between variables in the system (Grösser and Schaffernicht, 2012), think aloud protocols (Jensen, 2005) or verbal protocols (Kopainsky *et al.*, 2012).

In terms of *transfer of learning*, the subject groups did not exhibit any significant difference. Although the modified game played in trials 12 and 13 are essentially the same stock management task played in first 11 trials, all subject groups performed equally bad in the modified game, indicating that gradual-increase-in-complexity does not help *vertical transfer* of learning. This may be due to large complexity gap between the complex game and the modified version. The results of *horizontal learning* transfer experiments are also insignificant. These results are in agreement with our findings; transfer of learning is very limited, and only possible if games involved are very similar.

References

Alessi SM. 1995. Dynamic vs. static fidelity in a procedural simulation. In *Annual Meeting of the American Educational Research Association, San Francisco, CA*.

Barlas Y, Dalkiran E. 2008. Testing dynamic decision making under real-time pressure: A scuba diving simulator. In *Proceedings of the 26th International Conference of the System Dynamics Society*. Athens, Greece.

De Jong T, Martin E, Zamarro JM, Esquembre F, Swaak J, van Joolingen WR. 1999. The integration of computer simulation and learning support: An example from the physics domain of collisions. *Journal of research in science teaching* **36**(5): 597–615.

Grösser SN, Schaffernicht M. 2012. Mental models of dynamic systems: Taking stock and looking ahead. *System Dynamics Review* **28**(1): 46–68.

Größler A, Maier FH, Milling PM. 2000. Enhancing learning capabilities by providing transparency in business simulators. *Simulation & Gaming* **31**(2): 257–278.

Jensen E. 2005. Learning and transfer from a simple dynamic system. *Scandinavian journal of psychology* **46**(2): 119–131.

Kirriemuir J, McFarlane A. 2004. Literature review in games and learning. Technical Report Number 8, FutureLab, Berkshire, UK.

Kopainsky B, Pirnay-Dummer P, Alessi SM. 2012. Automated assessment of learners' understanding in complex dynamic systems. *System Dynamics Review* **28**(2): 131–156.

Mulder YG, Lazonder AW, de Jong T. 2011. Comparing two types of model progression in an inquiry learning environment with modelling facilities. *Learning and Instruction* **21**(5): 614–624.

Nurmi S, Lainema T. 2002. Collaborative learning with dynamic business game simulation-innovative learning environment for business education. In *Proceedings of ED-MEDIA 2002, World conference on educational multimedia, hypermedia & telecommunications.* Association for Advancement of Computing in Education, 1473–1478.

Özgün O, Barlas Y. 2012. Effects of delay, nonlinearity and feedback on the overall complexity of a stock management game. In *Proceedings of the 30th International Conference of the System Dynamics Society.* St. Gallen, Switzerland.

Quinn J, Alessi S. 1994. The effects of simulation complexity and hypothesis-generation strategy on learning. *Journal of Research on Computing in Education* **27**: 75–91.

Salden RJ, Paas F, van Merriënboer JJ. 2006. A comparison of approaches to learning task selection in the training of complex cognitive skills. *Computers in Human Behavior* **22**(3): 321–333.

Swaak J, van Joolingen WR, de Jong T. 1998. Supporting simulation-based learning; the effects of model progression and assignments on definitional and intuitive knowledge. *Learning and Instruction* **8**(3): 235–252.

Van Merriënboer JJ, Kirschner PA, Kester L. 2003. Taking the load off a learner's mind: Instructional design for complex learning. *Educational psychologist* **38**(1): 5–13.

White BY, Frederiksen JR. 1990. Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence* **42**(1): 99–157.

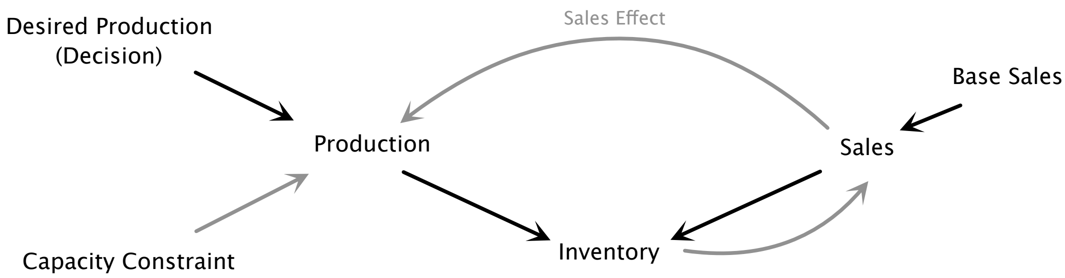
Yasarcan H. 2010. Improving Understanding, Learning and Performances of Novices in Dynamic Managerial Simulation Games: A Gradual-Increase-in-Complexity Approach. *Complexity* **15**(4): 31–42.

Appendices

A Instructions

A.1 Experiment Groups 1 and 2

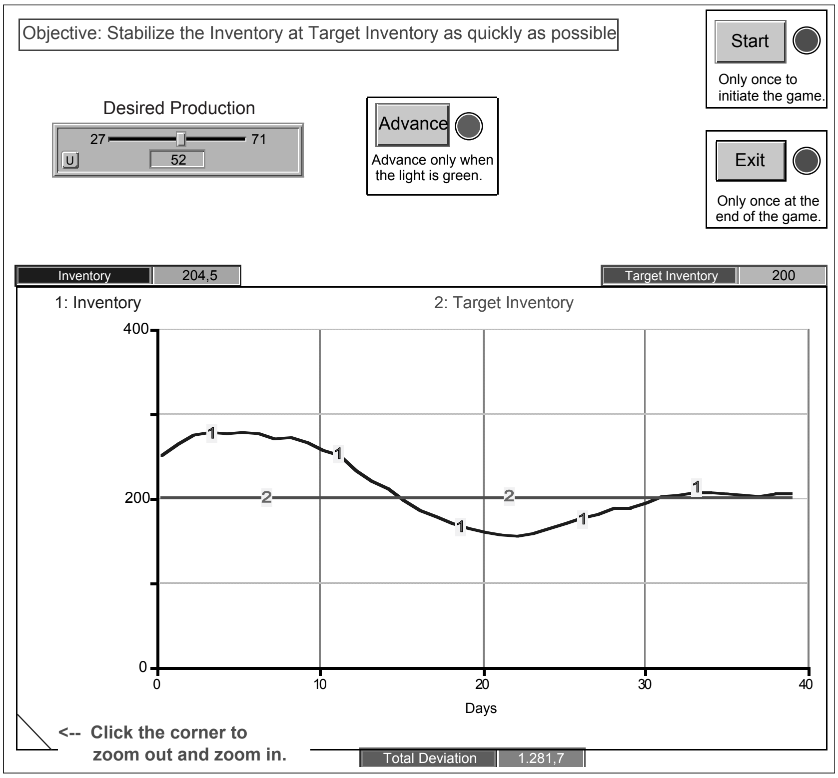
This interactive simulator is about a company that produces textile products. The company operates in a hypothetical world and all the rules of economy may not work as they do in the real world. You are the production manager who is only responsible from the production of t-shirts and your aim is to bring the t-shirt inventory level to a predefined target and keep it there. Your inventory level increases with production and decreases with sales. The figure below gives a broad representation of the causal relationships between key variables. The paragraph below explains the variables in the figure.



For every day, you will be setting a *Desired Production* (unit: boxes/day) for t-shirts, based on the current *Inventory* level (boxes). You can observe the current *Inventory* level immediately and accurately. Your *Desired Production* decisions will be transferred to the production engineers and processed after a **delay**. The duration of this delay will be a few days. The actual *Production* rate (boxes/day) is not determined exactly and entirely by your decision. The production engineers will take your decision and may increase or decrease it to meet capacity and planning constraints. Therefore, the actual *Production* will be a **modified** version of your *Desired Production.* Secondly, there is an adjustment mechanism in the production planning system that increases the *Production* to keep up with the *Sales*, when *Sales* is above a certain level, i.e. amplify the engineers *Production* decisions. This **sales effect** can also work in the opposite direction: when *Sales* is below some predetermined level, *Production* is automatically slowed down. The *Sales* rate is the number of units sold per day (boxes/day). *Sales* is out of your control and handled by the sales department. You do not have access to sales figures, but you can have a sense about it by observing *Inventory*, since *Sales* immediately decreases the *Inventory* level. *Sales* is positively related to *Inventory* level. In other words, as *Inventory* rises, the sales department immediately responds by increasing its sales efforts.

As the production manager, you have one decision to control the *Inventory*: **desired production**. You will decide on desired production for 40 days and **your objective is to stabilize the inventory around the target level of 200 boxes as quickly as possible**. Your performance will be assessed by the total deviation from the Target Inventory. Positive and negative deviations are equally undesired. You will start from an off-equilibrium condition and seek the target level. Pay attention to **delay**, **modification**, and **sales effect** described above, as they will complicate the task.

You will play 15 different games. First three games will have the underlying structure explained above. Next six games will be simplified versions of this game. To be specific, some of the delay, modification, and sales effect will be removed from the game to help learning. You will play these six games in an increasing order of complexity. Then, you will play the first complex game two more times. After playing 11 games, you will be asked to fill in a questionnaire. Finally, you will play a modified version of this game and a stock management game in a different context, twice. There is no time limit in any step of the experiment.



The game screen is as shown above. When you open the game file click the Start button once to start the game. This will initialize the game and advance you to the first day. You cannot change the first days *Inventory*, so do not move the sliders before clicking the Start button. Each day, you must set a *Desired Production* value using the slider and click the Advance button once. You will observe the *Inventory* behavior on the graph in blue and see its numerical value in a blue box above the graph. You will also see the constant *Target Inventory* on the same graph in red. When you complete 40 days, a warning box will appear. When you finish the game you should (1) write down your *Total Deviation* on the sheet provided, (2) click the Exit button and (3) **save the game** when you are asked. Do not play any game more than once, pass to the next game. If you did something by error that you did not intend to do, please stop immediately and inform the facilitator. You will have a trial game at the beginning to familiarize with the game interface.

Make sure that you understand the instructions completely before you start the experiments. If there is anything you do not understand, please ask your questions before you start playing. Work on your own and do not talk to the other subjects.

You must save the game files and fill out the necessary documents for the proper completion of the experiment. If you complete the experiment properly, you will earn a 10 TL base payment[3], plus an additional reward from 0 to 10 TL, depending on your performance in the games that you play. Your reward will be largely based on your performance in games 10 and 11 (70%). However, other games scores and questionnaire results will be also taken into consideration with smaller weights (30% in total). Thank you for your participation.

A.2 Experiment Group 3

Second, third and fourth paragraphs are modified as follows:

For every day, you will be setting a *Desired Production* (unit: boxes/day) for t-shirts, based on the current *Inventory* level (boxes). You can observe the current *Inventory* level immediately and accurately. Your *Desired Production* decisions will be transferred to the production engineers and processed after a **delay**. The duration of this delay will be a few days. The *Sales* rate is the number of units sold per day (boxes/day). Sales is out of your control and handled by the sales department. You do not have access to sales figures, but you can have a sense about it by observing *Inventory*, since *Sales* immediately decreases the *Inventory* level. Sales is positively related to *Inventory* level. In other words, as *Inventory* rises, the sales department immediately responds by increasing its sales efforts.

As the production manager, you have one decision to control the Inventory: **desired production**. You will decide on desired production for 40 days and **your objective is to stabilize the inventory around the target level of 200 boxes as quickly as possible**. Your performance will be assessed by the total deviation from the Target Inventory. Positive and negative deviations are equally undesired. You will start from an off-equilibrium condition and seek the target level.

In the full (complex) game, the actual *Production* rate (boxes/day) is not determined exactly and entirely by your decision. The production engineers will take your decision and may increase or decrease it to meet capacity and planning constraints. Therefore, the actual *Production* will be a **modified** version of your *Desired Production*, as illustrated above. Secondly, there is an adjustment mechanism in the production planning system that increases the *Production* to keep up with the *Sales*, when *Sales* is above a certain level, i.e. amplify the engineers *Production* decisions. This **sales effect** can also work in the opposite direction: when *Sales* is below some predetermined level, *Production* is automatically slowed down. Complex game based on this full model will be played at the end, as games 10 and 11.

The first 9 games to be played for learning purposes will be simpler versions of the full model. To be specific, some of the delay, production modification, and sales effect will be removed or reduced to help learning. You will play these 9 games in an increasing order of complexity. After playing 11 games, you will be asked to fill in a questionnaire. Finally, you will play a modified version of this game and a stock management game in a different context, twice. There is no time limit in any step of the experiment.

[3]equivalent of 5.5 USD

A.3 Control Group 1

Only fourth paragraph is modified as follows:

You will play 15 different games. First 11 games will have the underlying structure explained above. After playing 11 games, you will be asked to fill in a questionnaire. Then, you will play a modified version of this game and a stock management game in a different context, twice. There is no time limit in any step of the experiment.

A.4 Control Group 2

Only fourth paragraph is modified as follows:

You will play 15 different games. First three games will have the underlying structure explained above. Next six games will be simplified versions of this game. To be specific, some of the delay, modification, and sales effect will be removed from the game to help learning. You will play these six games in an random order of complexity. Then, you will play the first complex game two more times. After playing 11 games, you will be asked to fill in a questionnaire. Finally, you will play a modified version of this game and a stock management game in a different context, twice. There is no time limit in any step of the experiment.

B Game Equations

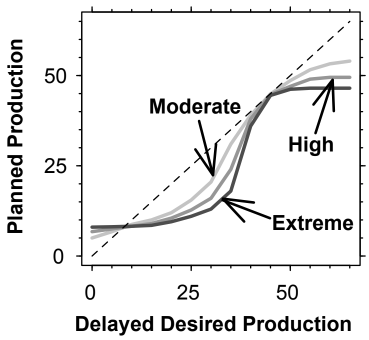
Simple version:

$$\begin{aligned}
\mathit{Inventory}(t+1) &= \mathit{Inventory}(t) + \mathit{Production}(t) - \mathit{Sales}(t) && (2)\\
\mathit{Inventory}(0) &= 250 && (3)\\
\mathit{Target\ Inventory} &= 200 && (4)\\
\mathit{Sales}(t) &= \mathit{Base\ Sales} + \mathrm{NORM}(0,1) && (5)\\
\mathit{Base\ Sales} &= 38 && (6)\\
\mathit{Production}(t) &= \mathit{Desired\ Production}(t) && (7)\\
\mathit{Desired\ Production}(0) &= 52 && (8)
\end{aligned}$$

Starting from moderate version: *Delayed Desired Production* is introduced. *Production* is now equal to *Delayed Desired Production*.

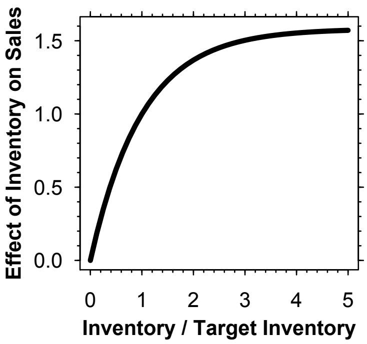
Starting from moderate-to-challenging version:

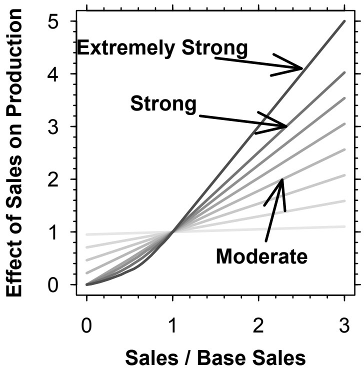
$$\mathit{Production}(t) = \mathit{Planned\ Production}(t) \tag{9}$$



Starting from challenging-to-complex version:

$$\begin{aligned} Sales(t) &= (Base\ Sales + \mathrm{NORM}(0,1)) \times Effect\ of\ Inventory\ on\ Sales & (10) \\ Production(t) &= Planned\ Production(t) \times Effect\ of\ Sales\ on\ Production & (11) \end{aligned}$$



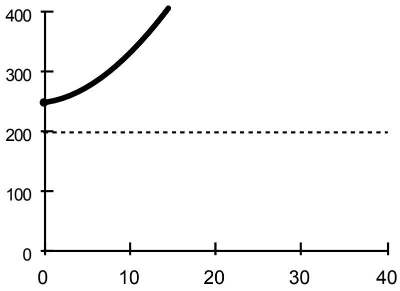


C Questionnaire

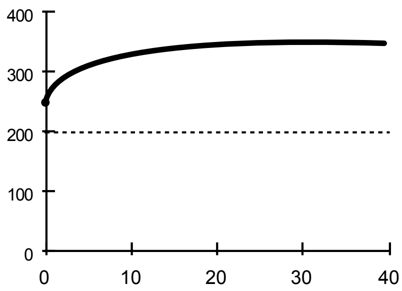
The subjects are specifically instructed that this questionnaire is about the complex game.

1. Which will be the correct inventory behavior if you <u>never change</u> the initial Desired Production throughout the game?

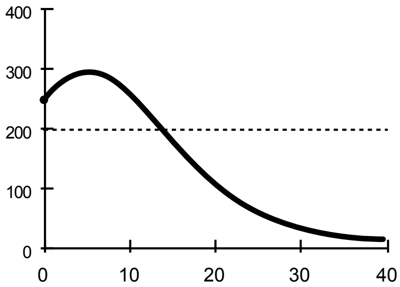
A)



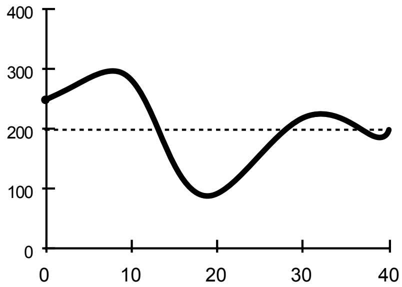
B)



C)

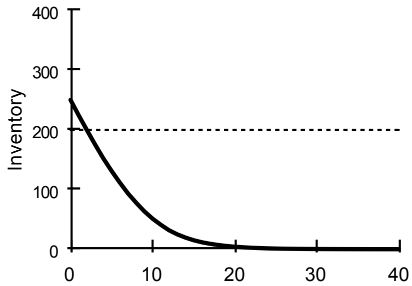


D)

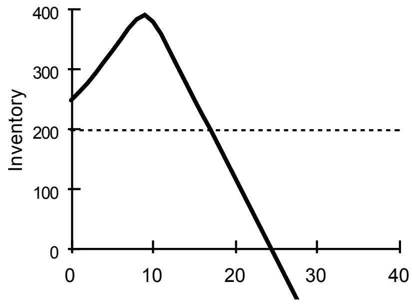


2. Which will be the inventory behavior if you set your Desired Production decisions always at the <u>minimum allowed</u> (27 boxes/day) throughout the game?

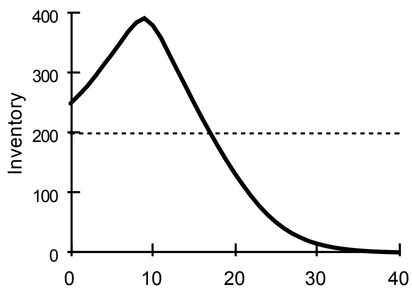
A)



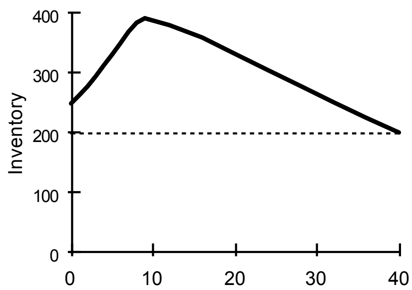
B)



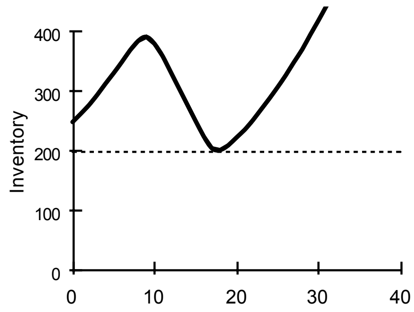
C)



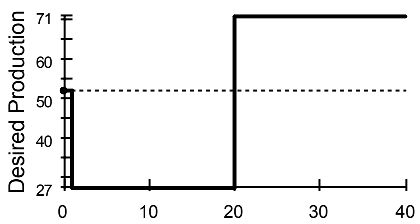
D)



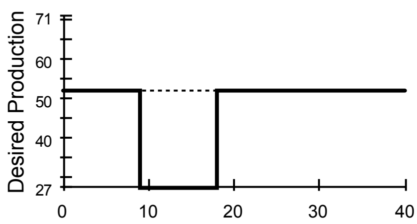
3. Which sequence of Desired Production decisions can give the following inventory behavior?



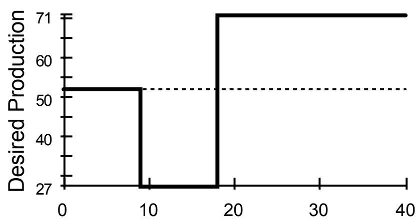
A)



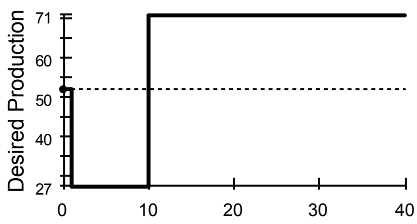
B)



C)



D)



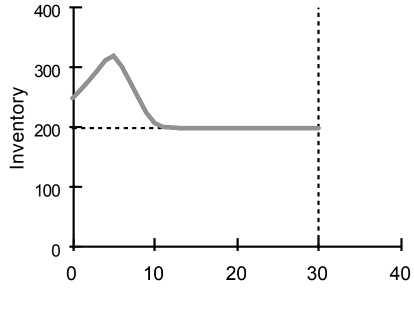
4. Other things being equal, what happens to Sales if Inventory is suddenly increased from 200 to 400?

A) Increases B) Decreases C) Does Not Change D) Cannot be determined

5. Assume that there is no delay, modification, or sales effect in the game, and Sales is constant at 40 boxes/day. What should be the constant Production to bring the Inventory from 180 boxes to 200 boxes?

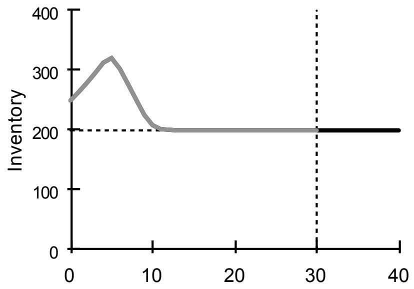
A) 20 B) 40 C) 60 D) 240

6. Assume that you managed to bring the Inventory to the target and you can keep it there with a constant Desired Production (see figure below).

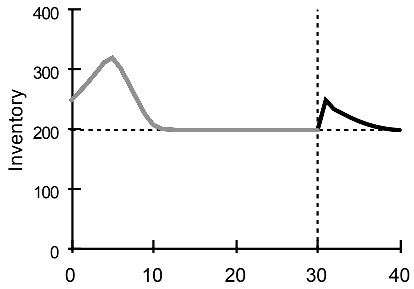


Now, if you increase Desired Production on day 30, what happens to Inventory?

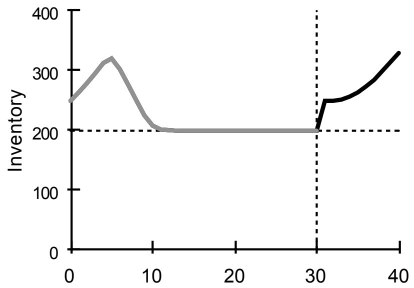
A)



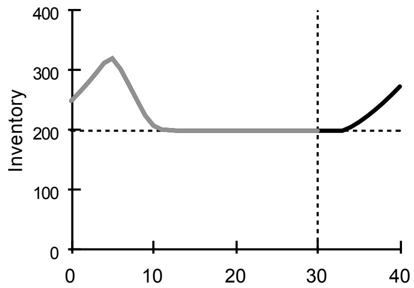
B)



C)

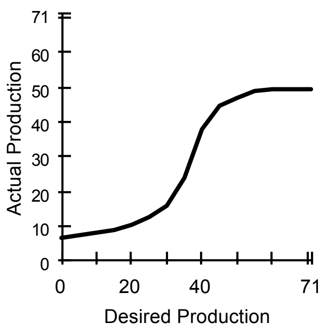


D)

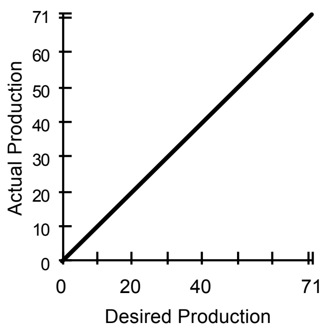


7. Ignoring the delay and sales effect, what do you think is the relationship between your desired production and actual production?

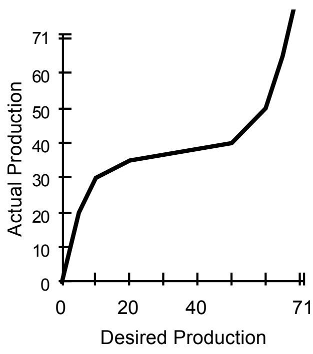
A)



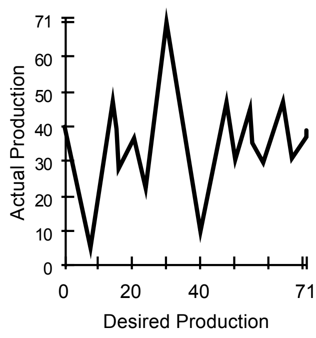
B)



C)



D)



8. Assume that current Inventory level is 300 boxes and Sales is 47 boxes /day. If there is no delay and modification due to capacity constraint, i.e. there is only Sales effect, which of the following can be the Actual Production, if your Desired Production is 70 boxes/day?

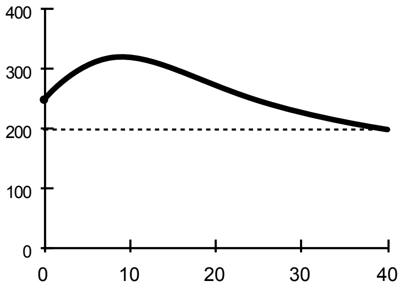
A) 35 B) 59 C) 70 D) 123

9. In your opinion, which factor creates the highest difficulty to the game? [4]

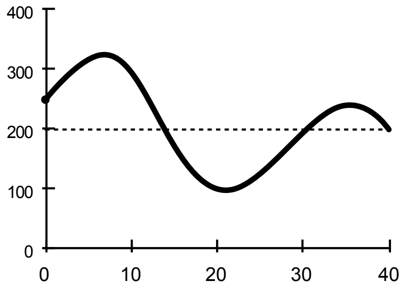
   A) Delay between your decisions and actual Production
   B) Variations in Base Sales
   C) Modification of Desired Production decisions by engineers
   D) Sales effect that automatically adjusts Production according to Sales

10. Which one yields the minimum total deviation (the best score)?

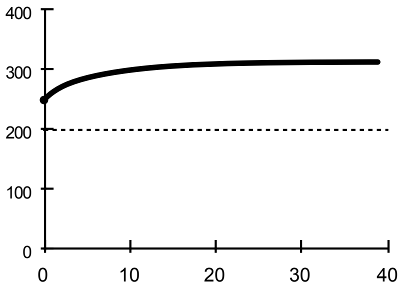
A)



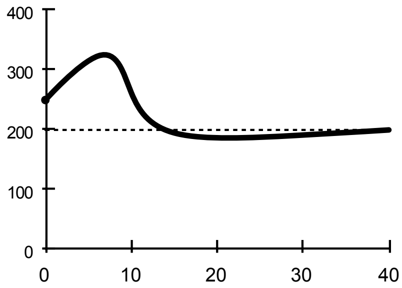
B)



C)



D)



Correct answers: 1-A, 2-C, 3-D, 4-A, 5-C, 6-D, 7-A, 8-D, 10-D

[4]This question is not counted toward questionnaire score