Final draft post-refereeing; published in Simulation & Gaming Vol. 46(3-4): 293-321, DOI 10.1177/1046878113517536

Cover sheet

SD_2013_6_Kopainsky_Prior-Exploration

8000 words (main text); 9050 word (with references); 2 pp of figures, 2 pp of tables

Birgit Kopainsky (corresponding author)

System Dynamics Group, Department of Geography, University of Bergen, Post box 7800, 5020 Bergen, Norway

Telephone: +47 555-83-092

Fax: +47 555-83-099

E-Mail: <birgit.kopainsky@geog.uib.no>

Stephen M. Alessi

College of Education, The University of Iowa, 370 Lindquist Center, Iowa City, IA 52242, United States

Telephone: +1 319-335-5568

Fax: +1 319-335-6145

E-Mail: <steve-alessi@uiowa.edu>

Matteo Pedercini

Millennium Institute, 1634 Eye Street, NW Suite 300, Washington DC 20006, United States

Telephone: +1 202-383-6200

Fax: +1 202-383-6209

E-mail: <mp@millennium-institute.org>

Pål I. Davidsen

System Dynamics Group, Department of Geography, University of Bergen, Post box 7800, 5020 Bergen, Norway

Final draft post-refereeing; published in Simulation & Gaming Vol. 46(3-4): 293-321, DOI 10.1177/1046878113517536

Telephone: +47 555-84-134

Fax: +47 555-83-099

E-Mail: <pal.davidsen@geog.uib.no>

Effect of prior exploration as an instructional strategy in system dynamics learning environments

Birgit Kopainsky[1], Stephen M. Alessi[2], Matteo Pedercini[3], Pål I. Davidsen[1]

[1] University of Bergen, Norway

[2] The University of Iowa, USA

[3] Millennium Institute, USA

Abstract

In complex simulation-based learning environments, participants' learning and performance may suffer due to demands on their cognitive processing, their struggle to develop adequate mental models, failure to transfer what is learned to subsequent learning or activities, and fear of failure. This study investigates an instructional strategy addressing those four problems, which we call *prior exploration strategy*. It was implemented in a simulation requiring participants to optimize a developing nation's per capita income. The prior exploration strategy allows participants to manipulate and see the results of a simulation model in practice mode before they manage a similar simulation in a more final mode. The strategy was assessed in an experiment comparing participants using the prior exploration strategy with participants studying equivalent content in a non-exploratory fashion. The dependent variables were performance within the simulation and improvement of participants' understanding. The prior exploration strategy significantly improved participants' performance, as measured by per capita income. It also significantly improved *some* aspects of the participants' understanding (e.g., their understanding of the nation's debt accumulation) but not others (e.g., their understanding of the need to balance the nation's health, education, and infrastructure investments; those that appear to have complex interrelations).

Keywords: cognitive load; exploratory behavior; instructional strategies; mental models; simulation games; system dynamics; transfer of learning, learning environments, subsequent learning, fear of failure, prior exploration strategy, per capita income, performance, improvement of understanding, complex systems, feedback processes, time delays, nonlinearities, accumulations, decision making, cognitive resources, instructional overlays, barriers to decision making,

Complex dynamic systems are characterized by multiple feedback processes, time delays, nonlinearities, and accumulations (Sterman, 2002). Most people, even experts, find them difficult to understand and manage successfully (Brehmer, 1992; Funke, 1991; Jensen, 2005; Moxnes, 2004; Rouwette, Größler, & Vennix, 2004; Sterman, 1989a; Sterman & Booth Sweeney, 2007). Much of the research in system dynamics strives to improve the design of computer simulations and games as tools to improve human decision making and management of complex systems.

Computer simulations certainly have potential to overcome limits on cognitive resources (e.g., Sterman, 2000; Tennyson & Breuer, 2002). Simulation models without instructional overlays (e.g., interfaces that provide guidance, feedback, and tools to support learning), are, however, generally considered insufficient, the more so as systems increase in complexity (Alessi, 2000a; Spector & Davidsen, 1997). Simulation/games or interactive learning environments (ILEs) which *do* include an instructional overlay are one strategy to transfer the insights gained from a formal simulation model to a wider audience (Machuca, 2000; Spector & Davidsen, 1997). In a typical ILE, participants study textual instructions that describe the structure of the system they should manage. Participants then progress to a simulation-based, decision-making interface (the core of the ILE), where they must solve the task (i.e., manage the system) presented in the instructions. A variety of instructional strategies can be applied when designing the user interface of an ILE. These strategies include providing explanations of observed behavior, giving hints before users take action, and providing feedback following those user actions (Alessi, 2000a). These strategies work well in some simulations, especially simulations representing simpler systems. However, evaluations of the effectiveness of these instructional strategies are limited and give mixed results regarding users' understanding and dynamic decision making performance (see Sawicka & Rydzak, 2007 for a review).

Barriers to decision making in interactive learning environments. Learners' success in making decisions in more complex ILEs is likely to be hampered by four barriers:

1. The complex models underlying the ILE impose too much load on learners' cognition. Cognitive Load Theory provides a clear theoretical explanation for this phenomenon (Sweller, 2005). Learners' difficulties in complex environments are exacerbated when overall cognitive load is high and alleviated when good design decreases cognitive load. Designers can do that, most easily, by decreasing extraneous cognitive load, such as that due to the complexity of a user interface, without sacrificing the interesting and important details of the content itself (Sweller, van Merrienboer, & Paas, 1998). In situations where the content is innately and unavoidably complex, research and principles from simulation design and instructional design in general suggest that complex content be carefully sequenced beginning with simpler or more general aspects and gradually increased in complexity (e.g., Alessi, 2000b; Reigeluth, 1999; Bruner, 1960).
2. Learners must, but often cannot, form an adequate mental model to support decision making. Mental model theory asserts that people learning about a phenomenon or situation will form a model in their mind of what is important and how things work (Seel, Al-Diban, & Blumschein, 2000; Doyle & Ford, 1998). Such models can take many forms (e.g., images, sets of rules, procedures) and unfortunately they are not always correct. If learners in a simulation-based ILE (or in a real situation) have errors in their mental model, such that it differs from the *actual* underlying model, their understanding and performance in the ILE, or in the real situation, will suffer.
3. As people aquire knowledge and skills in an ILE, they must apply them (i.e., must transfer what they learn) to subsequent activities, either those later in the ILE, or those in real life (about which the ILE teaches). However, such transfer is often unsuccessful. People frequently acquire knowledge or skills yet fail to transfer them to applicable situations. On a basic level, transfer of learning is simply being able to apply what has just been learned at the *present* time to activity at some *later* time. That later time may be five minutes later. For example, if a learner sets parameters in a simulation and observes a particular outcome, using that just-acquired knowledge to do the *next* simulation run with sensible parameter choices (such as to test a new hypothesis) is an example of transfer, albeit very short term and near transfer. A key principle of near learning transfer is the similarity principle (Gagné, 1954; Osgood, 1949), which says that the more similar the stimuli and responses of two situations, the more likely transfer of learning will occur. In contrast, far transfer, which has been likened to generalization in learning, is more dependent on variation in situations, stimuli and responses (Barnett & Ceci, 2002), and designing such variation into an ILE should improve far transfer to real world activities.

4. When faced with important decisions (even in a simulation) learners often demonstrate the anxiety typical of risk (or loss) avoidance. People vary greatly in their willingness to take risks. However, a general rule is that people decrease their risk taking when they perceive the stakes as being high (e.g., dangerous or potentially very expensive) and increase their risk taking when the stakes are low (nothing consequential will happen). Kermer, Driver-Linn, Wilson, & Gilbert, (2006) suggest that people will strive to avoid losses unless they perceive the amount of potential gain as much greater than the amount of potential loss. They also suggest that people generally expect the emotional effect of losses will be much greater than they in fact turn out to be.

Prior exploration as an instructional strategy in ILEs. Based on these four barriers to complex dynamic decision making (and managing complex systems), we designed an instructional strategy for ILEs. The strategy comprises a simulation-based *prior exploration* phase that precedes the simulation-based *decision-making* or *management* phase. The prior exploration phase incorporated four critical design features. First, it started with reduced complexity, so as to minimize initial cognitive load, and gradually increased complexity so as to manage cognitive load at a reasonable level. Second, it employed user controllable visuals (sliders controlling dynamic graphs) to induce an accurate mental model of how key input variables affected outcome variables. Third, it employed the similarity principle (by maintaining maximum similarity between the actions and visuals of the prior exploration phase and the subsequent management phase) to foster near transfer of learning. Fourth, the prior exploration phase incorporated *reversible* decisions; participants could move sliders up and down, and observe the results across all 50 simulated years. They were allowed to move back in time and try again. This appeared safer to participants than the management phase (in which they could only go forward in time). The possibility to go back in time should encourage exploratory behavior, even potentially risky exploration, because it is necessary for learners to see what *can* lead to catastrophic outcomes in order later to *avoid* such outcomes. The four barriers (cognitive load, mental model errors, poor transfer of learning, and risk avoidance) are the theoretical foundation. These four design features are an example of putting theory into practice via the prior exploration phase. We did not, however, treat the four features as separate independent variables in the study reported here. Rather, the use of the prior exploration strategy, which combined all four features, was the independent variable.

Kopainsky and Sawicka, (2011) provided evidence that including prior exploration in an ILE improves both performance in the management phase of the ILE and learners' understanding of a problem and its solution. That evidence was, however, limited to one specific task (reindeer management, a system dynamics simulation model with only one stock variable) and to a rather small number of participants. In this research we investigate whether the prior exploration strategy improves performance and understanding in more complex systems, namely, those with more than one accumulation variable.

Towards this end, we designed a simplified experimental version of an ILE that we have been developing and refining since 2006. That ILE, called BLEND (the Bergen Learning Environment for National Development), its underlying simulation model, and our initial pilot tests are described in detail in (Alessi, Kopainsky, Davidsen, & Pedercini, 2008; Kopainsky, Pedercini, Alessi, & Davidsen, 2010; Kopainsky, Alessi, Pedercini, & Davidsen, 2009). To summarize, participants in BLEND (the original ILE) play the roles of government ministers in a virtual developing nation who must make and revise budget decisions concerning domestic expenditures, taxes, and borrowing. They see the effects of their budget decisions in different national outcomes including economic indicators (e.g., income, national debt, government revenue), social indicators (e.g., population, literacy, life expectancy), and environmental indicators (e.g., water quality and forest preservation). The ministers work simultaneously and interact as they make their decisions. The goal of BLEND is for participants to experience and understand that national development processes are characterized by non-linear relationships, delays, feedback loops, and multiple sources of causation (all key principles of system dynamics) requiring design of long term policies across government ministries.

Our use of BLEND in classes and workshops made it clear, however, that people do not learn those important (and difficult) principles by mere exposure to them. As in many simulation-games, players found the management tasks difficult and they frequently relied on either trial-and-error decision making or on their traditional (often faulty) decision making strategies. It was clear that better instructional strategies were needed.

The prior exploration strategy seems an ideal one, given its design based on the four previously discussed barriers to learning. We implemented the strategy in an experimental version of BLEND. This version contains only five key stocks and allows participants to manage the nation individually (rather than in conjunction with several other ministers) in the role of the nation's prime minister, with authority for all the key national decisions. We tested its effectiveness in an experiment that compared participants using the prior exploration strategy with participants studying equivalent content in a non-exploratory fashion. We investigated the effects of prior exploration as an instructional strategy on both participants' performance and understanding.

Performance, one of our two dependent measures, refers to the success of participants' problem solving *within* an ILE. In addition to performance, our second dependent measure was participants' understanding. While performance is primarily a measure of initial learning, understanding is more indicative of learning transfer, because the tasks and questions that measured understanding were different from the activities within the ILE. It is a measure of near transfer in contrast to far transfer (Laker, 1990) because the concepts are the same and the context is very similar. Measuring both performance and understanding is important when evaluating the effectiveness of an instructional strategy because the relationship between understanding and performance is not straightforward. Good understanding of the complex dynamic problem is necessary for consistently making good decisions, that is, for good performance. However, good performance can also be achieved without understanding, for example, through trial and error. Furthermore, although understanding is necessary, it is not always sufficient for good performance. Practice, for example, is often needed in addition to understanding. In general, we should strive to implement instructional strategies that not only lead to improved performance, but do so *as a result of* better understanding.

The next section describes the laboratory experiment. The results section presents the experiment's outcomes for performance and understanding with 51 introductory level system dynamics students. Finally, we discuss the implications of our findings for ILE design and theory.

Method

A pilot test of the experiment is described in (Kopainsky, et al., 2009). As the results were promising, we left the procedure unchanged and only refined the measures for assessing understanding.

Task and decisions

Participants play the role of the prime minister in Blendia, a virtual sub-Saharan African Nation that, at the outset, is one of the poorest nations in the world (per capita income of $300 per person per year). Their task is to achieve and maintain the highest possible per capita income in the course of 50 simulated years (see Appendix 1 for the complete instructions). The time horizon of 50 years is necessary because it takes more than two decades for investment decisions to significantly impact the development pattern of a nation (Arndt, Jones, & Tarp, 2009). Behavior patterns such as worse-before-better or better-before-worse only become visible with a reasonably long time horizon.

The prime minister in Blendia has far reaching financial responsibilities and full decision-making authority regarding:

- Expenditures for education (an explicit decision)
- Expenditures for health (an explicit decision)
- Expenditures for roads (an explicit decision)
- Borrowing to finance the above expenditures (an implicit decision resulting from the three previous ones).

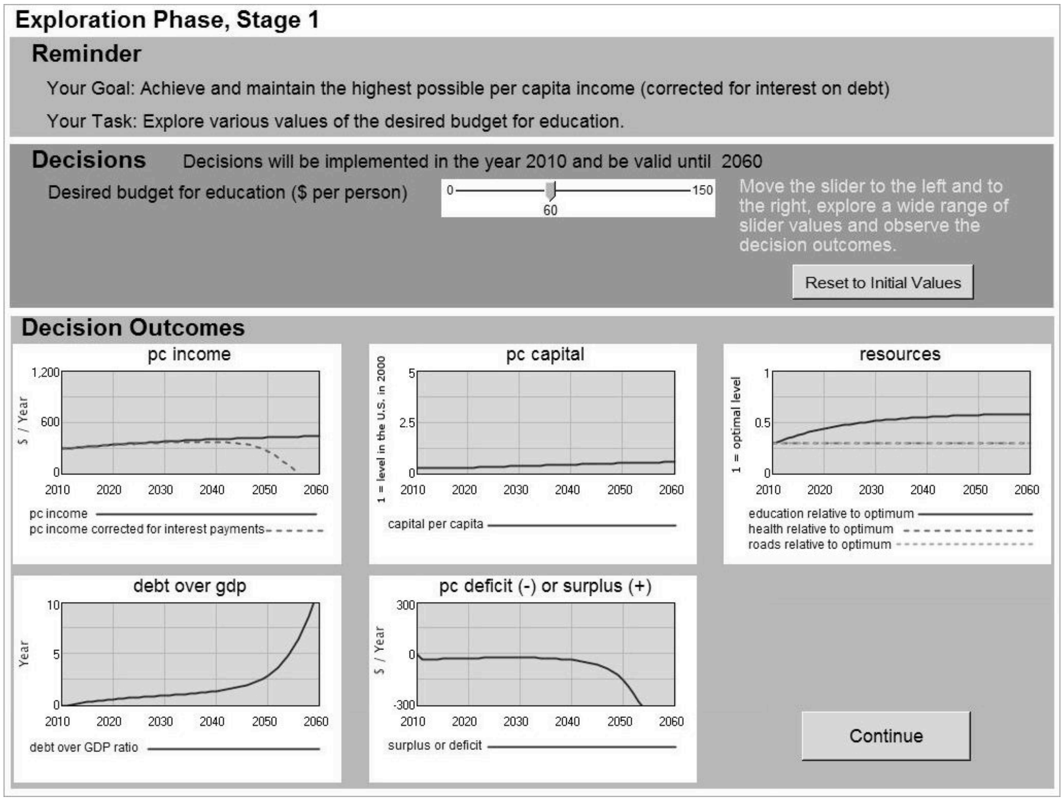
The simulation model used for the task is based on an extensive cross-country analysis that identified the role of a country's resources for its long term economic development (Pedercini, 2009) and it is described in detail in (Kopainsky, et al., 2009). The model depicts the development of per capita income over time as a consequence of reinforcing processes between capital accumulation through private sector development and capital accumulation through infrastructure and human development. Economic development can be severely hampered by the accumulation of debt as a consequence of too aggressive borrowing strategies. All variables are calculated on a per capita basis (e.g., capital per capita, debt per capita, and government development budget per capita) so that demographic development (i.e., the change in population) need not be taken into account. Expenditure and borrowing decisions are made every five years. The simulation starts in equilibrium and the prime minister remains in office throughout all 50 simulated years no matter how poor a participant's performance. The simulation model and the user interface for the experiment were designed and implemented in *LiveCode* (previously called *Runtime Revolution*). The instructions (Appendix 1) describe the structure of the simulation model underlying the experiment and the decisions participants could make.

Experimental design

Treatments. The study compared participants using a simulation-based exploratory activity (the prior exploration strategy) before encountering a more complex management task (the experimental group) with participants who were given a textual introduction (in place of the prior exploration strategy) before doing the same management task (the control group).

During the prior exploration phase participants in the *experimental group* first explored the effects of education expenditures alone (step 1), health expenditures alone (step 2) and road expenditures alone (step 3), after which they explored the combined effect of all three decisions taken together (step 4). Figure 1 shows the interface for the prior exploration phase for step 1, where participants could move a slider for desired budget for education up or down, and observe the effect of any changing education expenditure in Year 0 (the year 2010) on the behavior of several indicators over the entire 50 years time horizon. The other expenditure categories were held constant at their initial levels.

Figure 1: User interface of step 1 in the prior exploration phase



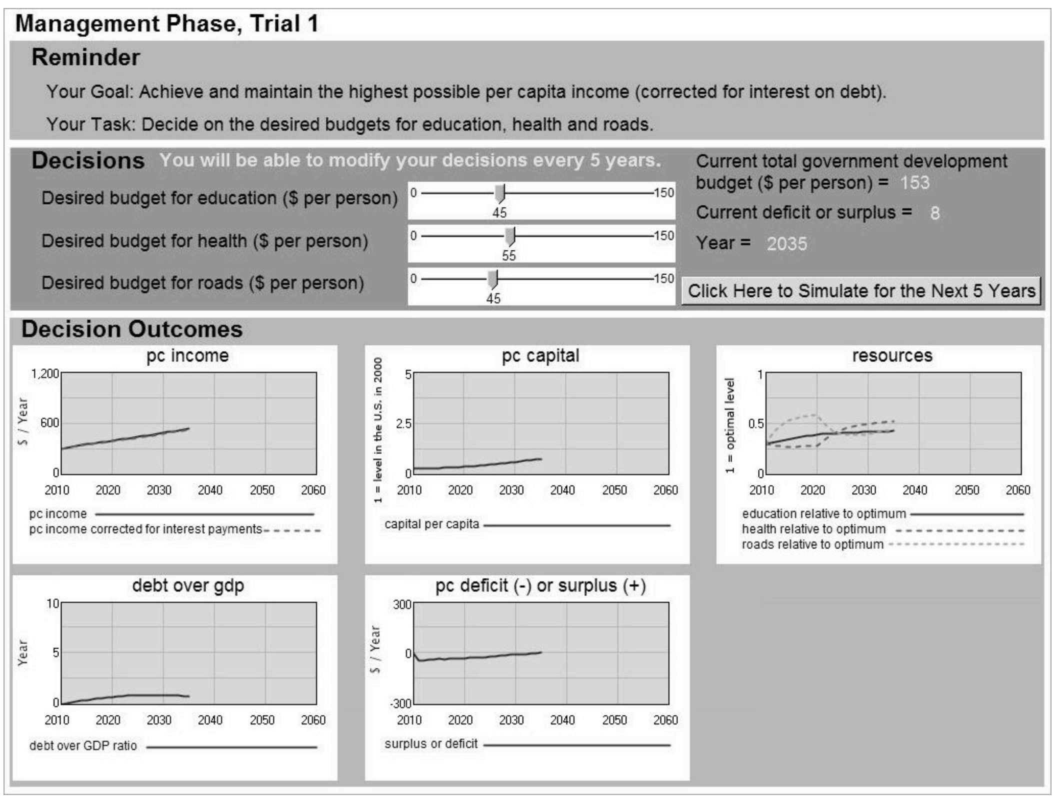
It is important to emphasize that the interface shown in Figure 1 uses dynamic graphs. As participants move the slider (in the case of Figure 1, the desired budget for education) the graphs at the bottom of the page (e.g., per capita income - interest payments) change immediately for the entire 50-year period, in accordance with the slider movement.

After each step, participants were asked to record their observations and to explain the resulting national outcomes indicated by the graphs. This pause after each exploration activity provided participants with the opportunity to reflect and was intended to prevent the impression that they are merely manipulating parameters in a trial-and-error fashion. The expectation is that through reflection, participants gain greater insight into the complexities of the system (Spector, Christensen, Sioutine, & McCormack, 2001). These “reflection opportunities” are considered part of the instructional strategy, and were not outcome measures.

After all four steps of the prior exploration phase, the experimental group proceeded to the management phase. The interface for the management phase is shown in Figure 2. The overall task is the same, to manage the nation well. Now, however, participants must more finely manage all three variables (education, health, and roads) and do so five years at a time. That is, they can modify the three expenditures every five years. To be clear, in the prior exploration activity, participants only set expenditures for the first year and observed outcomes across the subsequent 50 years. In the main management task, participants set expenditures, observed the results across the next five years, changed the expenditures, observed results across another five years, and so on for ten successive five-year periods.

While the interface for the prior exploration phase used dynamic graphs (which changed immediately as the sliders were moved by the participants and which were reversible), the graphs in the management phase changed only when a participant had made all decisions and chose to progress five years. In the main management task, the participants could not reverse their decisions once implemented. As a participant moves the sliders, nothing happens to the graphs. Only when the participant clicks the button to "Simulate for the next 5 years" does the model recalculate and update the graphs for that next five-year period. Thus, although the underlying model is the same for both the prior-exploration activity and the main management activity, participants interact with it differently. The management activity is more like decision making in real life, including (in many cases) being irreversible and risky. The prior exploration strategy is reversible and therefore "feels" less risky to participants. However, it does not reveal the optimal management strategy. It only provides participants with a tool to discern it on their own.

Figure 2: Interface for the management



Participants in the *control group* studied the same instructions as the experimental participants. They then engaged in an alternative task to the simulation-based prior exploration activity. They were provided with a text that verbally described the relationships between key variables and the reaction of outcome variables (e.g., per capita income) to changes in input variables (e.g., expenditure for education). The text was designed to require approximately the same study time as was required for the prior exploration activities (see Appendix 2). Thus we hoped to provide the control participants with a comparable amount (both time and content) of instruction as the experimental participants. The control participants then proceeded to the simulation-based management phase, which was exactly the same as that given to the experimental participants.

Hypotheses. Based on our theoretical framework and on previous studies using prior exploration, we expected participants in the experimental group to perform better and to have better understanding of the system. Concerning performance, our null hypothesis predicted no difference in performance between the experimental and the control group and our alternative hypothesis predicted a difference favoring the experimental group. Concerning understanding, our null hypothesis predicted no difference in understanding between the experimental and the control group and our alternative hypothesis predicted a difference, once again favoring the experimental group.

Participants. Data was collected with 51 introductory level system dynamics students in the fall 2009 and spring 2010. The students were recruited from the University of Bergen in Norway, the University of St. Gallen in Switzerland and the University of Freiburg in Germany. Participants from all three locations were assigned evenly and randomly to the experimental (25 participants) and control group (26 participants). That is, *at each university*, about half of the participants were randomly assigned to the experimental condition and half to the control condition. Participants were between 23 and 30 years old and 60% male and 40% female. The experiment was conducted in English, which was a foreign language for all participants at all three universities. Their random assignment to the experimental and control groups should, however, prevent any language-related bias in the results.

Procedure. Participants were assigned randomly to the experimental or control group. Before starting the task, all participants received the same pre-briefing. It emphasized that they were about to manage a virtual nation over a long time horizon. They were then presented with the general schedule of the experiment.

The participants proceeded at their own pace and required between 45 and 90 minutes to complete all the activities. They worked at separate computers with no communication between them.

In addition to the simulation activities, participants completed several questionnaires designed to explain their performance and assess their understanding of the system. The first occurred immediately after the participants had been introduced to the nation of Blendia in the instructions. They were asked:

1. To describe the problem situation in Blendia at the beginning of their term as Prime Minister. This included identifying the key variables relevant to the problem and explaining the relationships between them.
2. To describe their strategy for increasing per capita income while maintaining low interest payments on debt. This included explaining which policies they would implement and why they thought these would have the desired effect.

After the management phase, the participants were again given a questionnaire. They were asked:

1. To revise their description of the problem situation in Blendia which was copied into the answer field.
2. To revise their description of a strategy to manage the nation which was copied into the answer field.
3. To report their interest in, prior knowledge of, and experience with national development issues and the use of simulations for national planning. This was used to control for the effects of participants' backgrounds on performance and understanding.

Debriefing

The experimental session was followed by a plenary debriefing session which included an exchange of participants' experiences while performing the experiment, collaborative development of the underlying model structure and a discussion of the short and long term effectiveness of different expenditure strategies. The three groups of participants were taking classes in which the experimental activity was pedagogically relevant and the debriefing served to reinforce their understanding of the underlying system dynamics principles and their application to the respective courses, including correction of any misunderstandings students may have had at the end of the simulation. Qudrat-Ullah (2007) has shown that debriefing can reduce misperceptions of feedback. This study, however, focused on the effects of the prior exploration strategy on performance and understanding, so the debriefing followed and did not affect the measurement of those dependent variables. Its purpose was to ensure that the experimental activities contributed to the participants' education in their respective courses.

Measurement of learning outcomes

To test the hypotheses we compared measures of performance and understanding to benchmark values. For performance, we calculated the optimal quantitative values for each budget category and decision period using Vensim's® policy optimization algorithm. Those values are the best that participants could have obtained if they made the best decisions possible. We evaluated participants' performance by subtracting interest payments on debt from the per capita income they attained and comparing these values to the optimal ones. The distance to the optimal value has been used in previous studies (e.g., Diehl & Sterman, 1995; Moxnes, 2004; Paich & Sterman, 1993; Sterman, 1989b) for the statistical analysis of experimental and control groups' performance and the quality difference between them.

This assessment of performance was objective, as it was based on straightforward calculations on outcome variables in the simulation and optimization by an unbiased computer program. In contrast, our assessment of understanding (described next) was more subjective, as it was based on participants' open-ended questionnaire responses and raters' evaluation of those responses.

To assess understanding, we compared participants' verbal responses on the questionnaires to responses to the same questions by experts. The experts' responses contained descriptions of the key relationships in the simulation model and of the necessary steps for solving the national development planning task in an optimal manner.

Although rating of verbal descriptions always includes some subjectivity, the following procedures were used to do the ratings in as unbiased a manner as possible. Participants' verbal responses were first printed on one side of an index card and the participant number was printed on the reverse side to enable blind scoring (primarily so raters would not know the condition a participant was in). A scoring protocol was created, comprising a list of phrases showing evidence of understanding either detail or dynamic complexity (Senge, 1990). Detail complexity represents the overall amount of content learned, for example, by the number of variables or concepts and the number of links between them that participants wrote in their verbal responses. Dynamic complexity refers to the presence of specific concepts that reflect important system dynamic concepts including feedback, delays, nonlinearities and multiple causation. The scoring protocol awarded points to each of these elements with the maximum number of points determined by the expert text.

Two of the authors evaluated and discussed the expert texts, identifying 16 relationships between important variables (detail complexity). These relationships are summarized in the upper half of Table 1 and are labeled Relationships (detail complexity). Participants received one point for each relationship identified, the maximum being 16.

Table 1: Coding scheme for measuring understanding

Relationships (detail complexity)	the goal is to maximize pc income minus interest payments
	per capita income depends on capital and total factor productivity
	capital increases with investment
	investment increases with per capita income
	investment increases with education

	investment increases with health
	investment increases with roads
	the prime minister can regulate expenditure on education, health and roads
	available budget equals tax revenue minus interest payments
	tax revenue equals per capita income times tax rate
	a deficit arises when the desired budget exceeds the available budget
	a surplus arises when the desired budget is below the available budget
	deficit leads to borrowing
	borrowing leads to debt
	debt leads to interest payments
	surplus leads to paying down debt
Characteristics of successful strategy (dynamic complexity)	balance resources (education, health, roads)
	invest in education early
	invest in roads early
	invest in health later
	borrow early
	pay down debt later

We similarly identified six characteristics in the expert texts that represented dynamic complexity and the understanding of stock and flow variables and their interactions (the part of Table 1 labeled Characteristics of successful strategy (dynamic complexity).

Participants received one point if their description included the concept of balancing the expenditures on education, health and roads. Neither roads, health nor education alone can stimulate per capita income very much. Per capita income grows fastest when the three resources are balanced.

Participants received one point each if their description included education and roads requiring early increases in expenditures and health requiring a somewhat delayed increase in expenditures. Many years must pass before expenditures on education affect the nation's per capita income. Knowledge, skills, techniques and capabilities embodied in labor can be acquired through education and training but they require time to show their effect. The same holds true for expenditures on health. The time necessary to achieve improvements in average life expectancy (a major indicator of health in a nation) is, however, considerably shorter than the time required to increase the average adult literacy rate (a major indicator of the quality of education in a nation). The results of expenditures on roads becomes visible fairly soon. To attain balanced growth, the optimal strategy is to prioritize education and roads in the early years, education because it has the longest implementation time and roads because those expenditures have a growth-stimulating effect fairly soon.

Finally, participants received one point each if they included the notion that borrowing early (to provide funds for expenditures) was important and that, at a later time, debt should begin to be paid off. The danger in *not* paying off debt is that with increasing debt, interest payments increase and these payments are deducted from tax revenue every year. Too aggressive borrowing or borrowing over too long a time period can easily bankrupt the nation and destroy any improvements in per capita income. With reasonable debt in the early years and adequate allocation to education, health and roads, the economy starts growing so well that debts can be paid back and spending can be increased even more in all three budget categories. This allows per capita income to increase considerably without having the negative effects of debt and interest payments.

Scoring was fairly liberal in all cases. Any phrase suggesting participants understood a particular concept in Table 1 was awarded a point. One of the authors rated all participants' responses and another author rated 20% of them in order to assess inter-rater reliability, which was .76 (Cohen's Kappa). The reliability of the scale used was .79 (Cronbach's Alpha).

Results

The data were not normally distributed. We therefore present results based on non-parametric tests. Participants from the three different universities demonstrated no statistically significant differences in performance, understanding, interest in and knowledge about national development planning, or other background information based upon a Kruskal-Wallis test at $\alpha=0.05$. We therefore analyze and report on the entire participant pool as one group, consisting of 25 experimental and 26 control participants.

On average, participants reported a high interest in development issues with a mean of 3.8 on a five-point scale from 1 (not at all interested) to 5 (extremely interested) and a slightly lower knowledge about national development issues with a mean of 3.3 on a five-point scale from 1 (very poor knowledge) to 5 (very good knowledge). A two-tailed Mann-Whitney test ($\alpha=0.05$) revealed no significant difference between the experimental and control group for either interest or knowledge and no significant effects due to participants' other background variables, such as their practical experience, whether they had ever used simulation to study national development issues or had ever taken classes in national development economics.

The time participants spent on the entire experiment varied considerably from one participant to another. However, participants in the control group spent significantly less time than participants in the experimental group (based on a two-tailed Mann-Whitney test at $\alpha=0.05$) for either time on the entire experiment or for the control group's reading activity versus the experimental group's prior exploration activity. We will elaborate on this difference further in the discussion section.

Performance

Figure 3 presents median performance for the experimental group (the solid black line) and the control group (dotted black line) across the 50 years of the simulation. The grey line represents the optimal solution.

Figure 3: Optimal and median performance of the experimental and control groups

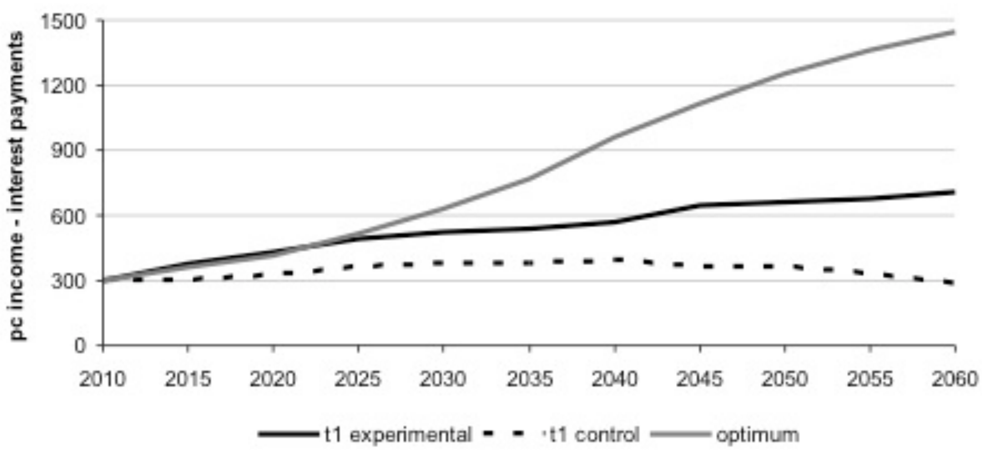
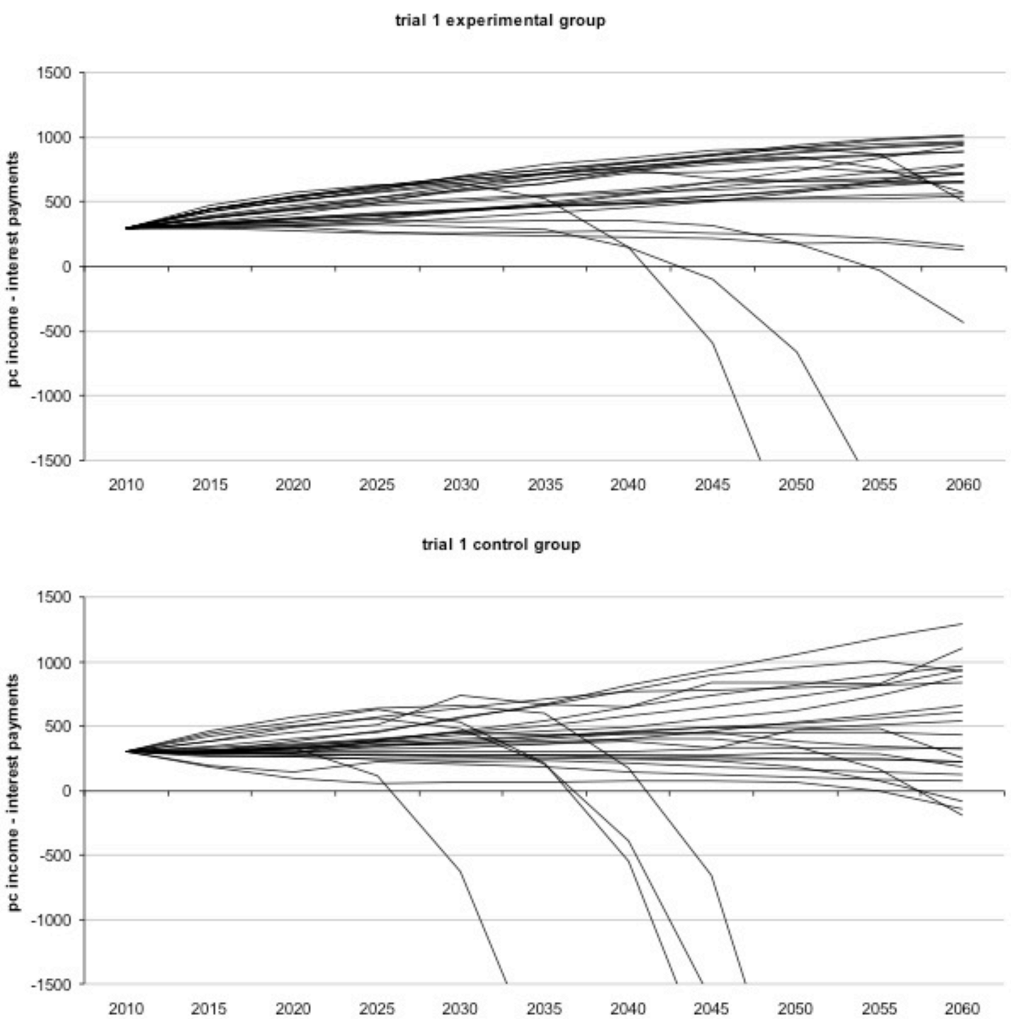


Figure 3 illustrates that the experimental and the control group show considerable differences. They start out the same, but diverge as the simulated years progress. Median performance in the experimental group shows continuous improvement in the stated performance goal, per capita income minus interest payments (also referred to as per capita income corrected for interest). The average behavior pattern for the control group, on the other hand, is an initial slight growth of that goal followed by a decline caused by high debt and the resulting high interest payments.

Figure 4 displays graphs of the performance goal for individual participant, both for the experimental and the control group. A separate line is plotted for each participant. The lines show participants' per capita income corrected for interest across the 50 simulated years. The figure illustrates that the participants in the experimental group manage to increase per capita income (corrected for interest payments on their debt) better than the control group. Although the experimental group had some unsuccessful performers, most experimental participants' performance converges towards fairly high per capita income values (corrected for interest). The performance of the participants in the control group is more varied and ranges from fairly successful participants to unsuccessful ones who buy high values of per capita income with high debt, the result of which is that the per capita income corrected for interest becomes negative half way through the simulation. The range of per capita income values (corrected for interest) at the end of the simulation is much greater for the control group than that of the experimental group. It appears the major impact of the prior exploration strategy is to help avoid the vicious cycle known as the reinforcing debt loop.

Figure 4: Overview of individual participants' performance in the experimental and the control group



To investigate whether those differences are statistically significant, we compared per capita income minus interest for the two groups using a two-tailed Mann-Whitney test. The difference between the experimental and the control group was statistically significant at the 5% level ($p=0.05$). We therefore reject the null hypothesis, (no difference in performance between the experimental and the control group). On the other hand, we retain the alternative hypothesis, that performance in the experimental group is better than in the control group.

Understanding

Table 2 presents participants' descriptions of the problem situation and their proposed solution strategy. The percentages in the cells of the table indicate the percentage of participants in either the experimental or control group who described a specific relationship or characteristic of the successful strategy before (MTP1) and after interacting with the simulation (MTP2).

Table 2: Percentage of participants describing correct relationships and characteristics of the successful solution strategy

	Exp MTP1	**Ctrl MTP1**	**Exp MTP2**	**Ctrl MTP2**	**Exp MTP2-MTP1**	**Ctrl MTP2-MTP1**
Relationships (detail complexity)						
the goal is to maximize pc income minus interest payments	36%	23%	40%	23%	4%	0%
per capita income depends on capital and total factor productivity	32%	19%	36%	19%	4%	0%
capital increases with investment	12%	8%	12%	8%	0%	0%
investment increases with per capita income	4%	0%	4%	0%	0%	0%
investment increases with education	28%	15%	28%	15%	0%	0%
investment increases with health	28%	15%	28%	15%	0%	0%
investment increases with roads	36%	15%	36%	15%	0%	0%
the prime minister can regulate expenditure on education, health and roads	32%	54%	36%	54%	4%	0%
available budget equals tax revenue minus interest payments	24%	8%	24%	12%	0%	4%
tax revenue equals per capita income times tax rate	24%	54%	24%	54%	0%	0%
a deficit arises when the desired budget exceeds the available budget	4%	4%	4%	4%	0%	0%

a surplus arises when the desired budget is below the available budget	0%	4%	0%	4%	0%	0%
deficit leads to borrowing	28%	8%	28%	8%	0%	0%
borrowing leads to debt	32%	31%	40%	31%	8%	0%
debt leads to interest payments	32%	35%	36%	35%	4%	0%
surplus leads to paying down	8%	8%	8%	8%	0%	0%
total relationships					*24%*	*4%*
Characteristics of successful strategy (dynamic complexity)						
balance resources	4%	0%	4%	8%	0%	8%
invest in education early	32%	35%	28%	27%	-4%	-8%
invest in roads early	44%	15%	44%	19%	0%	4%
invest in health later	12%	4%	20%	4%	8%	0%
borrow early	48%	27%	52%	23%	4%	-4%
pay down debt later	16%	0%	24%	0%	8%	0%
total strategy					*16%*	*0%*

Exp: Experimental group, Ctrl: Control group

MTP1: Measurement time point 1 (pre-simulation), MTP2: Measurement time point 2 (post-simulation)

According to Table 2, a fairly high percentage of participants, irrespective of their experimental condition, understood the goal of the task ("the goal is to maximize pc income minus interest payments") and what can be influenced ("the prime minister can regulate expenditure on education, health and roads"). A comparable percentage were able to identify the key stocks in the system (capital, education, health, roads, and debt). While they were able to identify the capital stock ("per capita income depends on capital and total factor productivity"), they failed to mention the relevant inflow ("capital increases with investment"). Similarly, many participants described the debt stock and that it increases with borrowing. However, few were also able to describe the outflow that will decrease the stock (that is, the "surplus leads to paying down" relationship).

A fairly high percentage of participants were able to describe the direct consequences of their decisions ("investment increases with education/health/roads" relationships). They were also quite clear about how they could generate revenue domestically (tax revenue = per capita income * tax rate). However, few participants described the indirect consequences of their decisions, such as the correct mechanisms of the budget (that is, the relationships "a deficit will occur if the desired budget is greater than the available budget" and "a surplus will occur if the desired budget is less than the available budget"). Another instance of an indirect consequence of participants' decisions is the fact that per capita income closes a reinforcing private sector development feedback loop among per capita income, investment, capital and per capita income.

The missing focus on the flows is confirmed by the descriptions of the characteristics of successful strategies. Many participants realized that they needed to finance the important early expenditures for education (because of the long delay) and roads (because of the rather immediate impact on growth) through borrowing. However, few participants mentioned the importance of paying down debt at a later time, in order to avoid exponential growth of interest payments on the debt.

The majority of participants failed to recognize the importance of the non-linearities in the system, with only a minority mentioning that the three resources (education, health and roads) need to be balanced for maximum growth and that as a result, expenditures for health must be increased a bit after expenditure increases in education and roads.

While these general statements are true for both the experimental and the control group, the two groups nevertheless differed from each other. At measurement time point one (after the instructions but before interacting with the simulation) the two groups showed no statistically significant differences with respect to their understanding of relationships (detail complexity; two-tailed Mann-Whitney test at α=0.05). However, the difference in the number of relationships between measurement time points one and two (the increase in the number of described relationships) was significantly higher for the experimental group while it was not for the control group (two-tailed Wilcoxon test α=0.05). The increase in the number of described relationships was due to an increase in descriptions of relationships, such as the goal of the task and the decision mechanism, or the determinants of per capita income and the debt stock. The prior exploration strategy did not, however, appear to increase the understanding of the relationships that were only indirectly linked to their decisions (i.e., the correct budget mechanisms; that investment increases with per capita income).

Concerning the number of good strategy descriptions, the two groups already differed significantly at measurement time point one, an unfortunate random variation in the composition of the two groups (based on a two-tailed Mann-Whitney test at α=0.05). The difference between the two groups persisted at measurement time point two. The change between measurement time point one and two was not significant for the comparison between the total number of strategy descriptions (Wilcoxon test at α=0.05). However, at measurement time point two, a significantly higher proportion of participants in the experimental group (Mann-Whitney test at α=0.05) described the need to borrow in the early years, which is crucial for solving the national development planning task.

Given these results (the experimental group showing a significantly greater increase in the number of relationships descriptions between measurement time points, and showing a significantly greater increase in one of the good strategy descriptions), the null hypothesis, (no difference in understanding between the experimental and the control group), is not supported. However, the alternative hypothesis, that understanding in the experimental group is better than in the control group, was only partially supported.

Discussion

These results provide some evidence that engaging participants in the free and safe exploration of important variables is an instructional strategy that improves their learning from a simulation-based interactive learning environment. This is true not only for initial learning (performance during the ILE), but also for a measure of learning transfer (evidenced by their understanding of the underlying model and how to manage it). Both performance and to some degree understanding were significantly better in the experimental group than in the control group. The significant difference in performance agrees with results from similar studies (Langley & Morecroft, 2004; Kopainsky & Sawicka, 2011; Skraba, Kljajic, & Borstnar, 2007) and with results from our own pilot applications of the materials both with introductory level system dynamics students (Kopainsky, et al., 2009) and with members of the actual target audience of our interactive learning environments (an unpublished study from an experimental session with planners from different government ministries in Swaziland). This is promising in so far as the simulation model underlying our ILE is more complex than the models from previous studies documenting the effectiveness of the prior exploration strategy.

More importantly, the experimental participants' improved performance appears to be for the right reasons, given that they exhibited better understanding of *some* principles. In line with recent methodologies (e.g., Doyle, Radzicki, & Trees, 2008; Karakul & Qudrat-Ullah, 2008; Kopainsky & Sawicka, 2011; Rouwette, Vennix, & Mullekom, 2002; Schaffernicht & Groesser, 2011), we also assessed participants' understanding and changes thereof in the national development task.

Participants in the experimental group improved their description of the structure of the problem (measured in the number of described relationships) more than the participants in the control group. However, the prior exploration strategy was clearly not as effective in improving understanding of dynamic complexity (strategy characteristics) as it was in improving understanding of detail complexity (number of described relationships). From a system dynamics point of view, prior exploration, therefore, seems to be more effective in developing an understanding of feedback mechanisms than of the behavior these mechanisms give rise to. This agrees with previous findings that people have great difficulties inferring behavior even from the simplest structures (e.g., Booth Sweeney & Sterman, 2000; Moxnes & Saysel, 2009; Sterman & Booth Sweeney, 2007). Although we were not able to overcome the latter difficulty fully, our results suggest that improved understanding of the problem structure itself, that is, without its behavioral implications, is necessary for improving performance.

Limitations

Several limitations should be considered when interpreting the results or suggesting their implications for theory and practice.

First, although we had designed the control group's reading task to be equivalent to the prior exploration activity in both content and study time, the data does not show that to be entirely the case. The reading time was significantly less than the prior exploration time. Furthermore, the question about the content equivalence of the conditions arises. One might contend that the reading activity identified variables but not their relationships, and that the prior exploration activity allowed experimental participants to actually *test* those relationships. The prior exploration strategy indeed has this advantage but this is at the same time the very reason for designing the strategy. One could also make the counter-argument that the reading activity, which did state relationships, did so directly, while participants in the prior exploration condition must infer relationships. Nevertheless, the difference in time on task definitely favored the experimental condition and it might be an alternative explanation for improved understanding. An improved experiment must control for time on task better, perhaps by limiting the time allowed for prior exploration. Finally, one should consider that the greater time spent in prior exploration might be an indicator of the value of prior exploration, because participants were allowed to use it as much or as little as they wanted, and they obviously chose to use it a lot (compared to simple reading).

Second, while our measure of understanding relationships was based on sixteen possible statements, our measure of understanding good strategies was based on only six possible statements. Since the reliability of a measure increases with the number of data points (such as the number of items in a test), it is likely that our measure of understanding relationships is more reliable than our measure of understanding good strategies. Although this is a fair criticism of the latter measure, it also suggests that it might be too conservative a measure, and that using a model with more identifiably effective strategies would improve the power (and likelihood of significance) of the understanding strategies analysis.

Third, although our design of the prior exploration strategy was based upon four cognitive learning principles (reduction of cognitive load through gradually increasing complexity, improvement of mental models through controllable visualization, improvement of learning transfer through the similarity principle, and decrease of perceived risk through time-reversible simulating), we did not measure those constructs and cannot attribute the improvements we did observe to one or another of them. Separate measurement of the four constructs would have required participants to devote too much time in those measures, rather than concentrating on the learning activity. In addition, measurements of some of the constructs are not yet well established and would have required creating new measures for this study, which in turn would have required substantial validation of the new measures. Our purpose was not a study and article on measurement techniques, but to evaluate the effectiveness of an instructional strategy.

The fourth and last limitation we point out is that the changes in participants' text from before the simulation to after the simulation were rather small. This may account for our mixed findings for understanding, especially the lack of difference in increase for understanding strategies. This limitation is to some extent inherent in the pre-test and post-test methodology (in contrast to a post-test only methodology). On the one hand, an advantage of having a pre-test is that it accounts for different levels of entering knowledge or skills. On the other hand, a disadvantage of having a pre-test is that participants learn from it, or when answering the same open-ended question feel that they do not really need to add anything to what they said earlier. Alternative measurement methodologies (such as objective questions or asking more explicit short-answer questions) might thus have improved our ability to detect improvements in understanding.

Conclusions and implications for future research

As others have pointed out (de Jong & van Joolingen, 1998), and as our own data shows, participants in simulation-based learning environments need instructional support. Without it, many participants engage in unproductive trial and error. Furthermore, the literature describes evidence that simply providing *opportunities* for exploration or other instructional strategies tends to be ineffective because participants fail to take full advantage of them (Größler, Maier, & Milling, 2000). The current *prior exploration* strategy implemented an instructional strategy in which participants were explicitly presented with an exploratory tool, were asked to use it, and afterwards were asked to reflect upon what they learned from it. The exploratory tool was designed to make the relationships between important variables visually clear through dynamic graphs, and to do so in a way that was risk free for the participants. Our measures do not allow us to separate the contribution (to outcome improvement) of the various factors implicit in the design of the prior exploration strategy (i.e., reducing cognitive load, improving mental models, improving learning transfer, or reducing perceived risk). Still, it does appear that this particular technique of requiring use of a risk-free visual exploratory tool is effective in improving performance. Our results, however, only show partial improvement in understanding.

Future research should therefore focus on both improving the prior exploration strategy and on improving our measures of understanding. One way to improve the prior exploration strategy might be to provide a user interface that makes the structure underlying the observed behavior more transparent (see e.g., Davidsen, 1992). Existing empirical data provides mixed evidence about the effectiveness of revealing the structure underlying a simulation-based game (Größler, et al., 2000). Nevertheless, increased transparency holds potential for improving understanding by helping learners to construct better mental models.

Improving the measurement of understanding, in turn, must go beyond indirect methods such as comparing novice and expert descriptions. Changes in understanding *during* learners' interactions with simulation-based games, including how their existing knowledge hinders or contributes to the acquisition of new knowledge (Kopainsky & Saldarriaga, 2012) must be developed. Additionally, it would be useful to measure intermediate variables that influence both performance and improved understanding, such as cognitive load and mental models, the very constructs that underlie the design of our prior exploration strategy and the strategies of many other researchers.

Acknowledgements

We are very grateful to Prof. Dr. Markus Schwaninger (University of St. Gallen) and PD Dr. Pablo Pirnay-Dummer (University of Freiburg) who gave us access to their students. We would also like to thank the reviewers for their useful comments and suggestions on earlier drafts.

Declaration of conflicting interests

The authors declared no conflicts of interest with respect to authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: During the research phase for this manuscript, part of the work of one of the authors (BK) was supported by the Swiss National Science Foundation under a Fellowship for Advanced Researchers. During the preparation of the manuscript, one of the authors (BK) received support from the Norwegian Research Council through the project *Simulation based tools for linking knowledge with action to improve and maintain food security in Africa* (contract number 217931/F10). The views and conclusions expressed in this article are those of the authors alone and do not necessarily reflect the views of the Swiss National Science Foundation or the Norwegian Research Council.

References

Alessi, S. M. (2000a). Designing educational support in system-dynamics-based interactive learning environments. *Simulation & Gaming, 31*(2), 178-196.
Alessi, S. M. (2000b). Simulation design for training and assessment. In H. F. O'Neil Jr. & D. H. Andrews (Eds.), *Aircrew training and assessment* (pp. 497-509). Mahwah, NJ: Lawrence Erlbaum Associates.
Alessi, S. M., Kopainsky, B., Davidsen, P. I., & Pedercini, M. (2008). *A system dynamics-based multi-user domain for improving national development planning*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, NY.
Arndt, C., Jones, S., & Tarp, F. (2009). Aid and growth: Have we come full circle? *Discussion Paper* (pp. 36): United Nations University, World Institute for Development Economics Research.
Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612-637.
Booth Sweeney, L., & Sterman, J. D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review, 16*(4), 249-286.
Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*, 211-241.
Bruner, J. S. (1960). *Process of education*. New York: Vintage Books.
Davidsen, P. I. (1992). *The Structure-Behavior Graph. Understanding the relationship between structure and behavior in complex, dynamic systems.* Paper presented at the 10th International Conference of the System Dynamics Society, Utrecht, The Netherlands.
de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179-201.
Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes, 62*(2), 198-215.
Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review, 13*(3), 253-265.
Doyle, J. K., Radzicki, M. J., & Trees, W. S. (2008). Measuring change in mental models of complex dynamic systems. In H. Qudrat-Ullah, J. M. Spector & P. I. Davidsen (Eds.), *Complex Decision Making* (pp. 269-294). Berlin/Heidelberg: Springer.
Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. Sternberg & P. Frensch (Eds.), *Complex problem solving: principles and mechanisms* (pp. 185-222). Hillsdale, NJ: Lawrence Erlbaum.
Gagné, R. M. (1954). Training devices and simulators: Some research issues. *The American Psychologist, 9*, 95-107.
Größler, A., Maier, F. H., & Milling, P. M. (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.
Karakul, M., & Qudrat-Ullah, H. (2008). How to improve dynamic decision making? Practice and promise. In H. Qudrat-Ullah, J. M. Spector & P. I. Davidsen (Eds.), *Complex Decision Making: Theory and Practice* (pp. 3-24). Berlin & Heidelberg: Springer-Verlag & NECSI.

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological Science, 17*(8), 649-653.
Kopainsky, B., Alessi, S. M., Pedercini, M., & Davidsen, P. I. (2009). *Exploratory strategies for simulation-based learning about national development.* Paper presented at the 27th International Conference of the System Dynamics Society, Albuquerque, NM.
Kopainsky, B., Pedercini, M., Alessi, S. M., & Davidsen, P. I. (2010). A blend of planning and learning: Simplifying a simulation model of national development. *Simulation & Gaming, 41*(5), 641-662.
Kopainsky, B., & Saldarriaga, M. (2012). *Assessing understanding and learning about dynamic systems.* Paper presented at the 30th International Conference of the System Dynamics Society, St. Gallen, Switzerland.
Kopainsky, B., & Sawicka, A. (2011). Simulator-supported descriptions of complex dynamic problems: Experimental results on task performance and system understanding. *System Dynamics Review, 27*(2), 142-172.
Laker, D. R. (1990). Dual dimensionality of training transfer. *Human Resource Development Quarterly, 1*(3), 209-223.
Langley, P. A., & Morecroft, J. D. W. (2004). Performance and learning in a simulation of oil industry dynamics. *European Journal of Operational Research, 155*(3), 715-732.
Machuca, J. A. D. (2000). Transparent-box business simulators: An aid to manage the complexity of organizations. *Simulation & Gaming, 31*(2), 230-239.
Moxnes, E. (2004). Misperceptions of basic dynamics: the case of renewable resource management. *System Dynamics Review, 20*(2), 139-162.
Moxnes, E., & Saysel, A. K. (2009). Misperceptions of global climate change: information policies. *Climatic Change, 93*(1), 15-37.
Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Psychological Review, 56*, 132-143.
Paich, M., & Sterman, J. D. (1993). Boom, Bust, and Failures to Learn in Experimental Markets. *Management Science, 39*(12), 1439-1458.
Pedercini, M. (2009). *Modeling resource-based growth for development policy analysis.* PhD Doctoral thesis, University of Bergen, Norway, Bergen.
Qudrat-Ullah, H. (2007). Debriefing can reduce misperceptions of feedback: The case of renewable resource management. *Simulation & Gaming, 38*(3), 382-397.
Reigeluth, C. M. (1999). The elaboration theory: Guidance for scope and sequence decisions. In C. M. Reigeluth (Ed.), *Instructional Design Theories and Models: A New Paradigm of Instructional Design* (Vol. 2, pp. 425-453). Mahwah, NJ: Lawrence Erlbaum Associates.
Rouwette, E. A. J. A., Größler, A., & Vennix, J. A. M. (2004). Exploring influencing factors on rationality: a literature review of dynamic decision-making studies in system dynamics. *Systems Research and Behavioral Science, 21*(4), 351-370.
Rouwette, E. A. J. A., Vennix, J. A. M., & Mullekom, T. v. (2002). Group model building effectiveness: a review of assessment studies. *System Dynamics Review, 18*(1), 5-45. doi: 10.1002/sdr.229
Sawicka, A., & Rydzak, F. (2007). *Incorporating delays in the decision-making interface: An experimental study.* Paper presented at the 25th International Conference of the System Dynamics Society, Boston, MA.
Schaffernicht, M., & Groesser, S. N. (2011). A comprehensive method for comparing mental models of dynamic systems. *European Journal of Operational Research, 210*(1), 57-67.

Seel, N. M., Al-Diban, S., & Blumschein, P. (2000). Mental models and instructional planning. In J. M. Spector & T. M. Anderson (Eds.), *Integrated and holistic perspectives on learning, instruction and technology: Understanding complexity*. Dordrecht, The Netherlands: Kluwer.
Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.
Skraba, A., Kljajic, M., & Borstnar, M. (2007). The role of information feedback in the management group decision-making process applying system dynamics models. *Group Decision and Negotiation, 16*(1), 77-95.
Spector, J. M., Christensen, D. L., Sioutine, A. V., & McCormack, D. (2001). Models and simulations for learning in complex domains: using causal loop diagrams for assessment and evaluation. *Computers in Human Behavior, 17*(5-6), 517-545.
Spector, J. M., & Davidsen, P. I. (1997). Creating engaging courseware using system dynamics. *Computers in Human Behavior, 13*(2), 127-155.
Sterman, J. D. (1989a). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes, 43*(3), 301-335.
Sterman, J. D. (1989b). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science, 35*(3), 321-339.
Sterman, J. D. (2000). *Business dynamics. Systems thinking and modeling for a complex world*. Boston et. al.: Irwin McGraw-Hill.
Sterman, J. D. (2002). All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review, 18*(4), 501-531.
Sterman, J. D., & Booth Sweeney, L. (2007). Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter. *Climatic Change, 80*(3-4), 213-238.
Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-31). Cambridge, UK: Cambridge University Press.
Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review* (Vol. 10, pp. 251-296): Kluwer Academic Publishing.
Tennyson, R. D., & Breuer, K. (2002). Improving problem solving and creativity through use of complex-dynamic simulations. *Computers in Human Behavior, 18*(6), 650-668.

Bios

Birgit Kopainsky is a senior research fellow at the University of Bergen, Norway. She holds a PhD in agricultural economics from ETH Zurich and a master's degree in Geography from the University of Zurich in Switzerland. Birgit Kopainsky is passionate about simulation models, dynamic decision making and food systems research. She has worked in several sub Saharan African countries and teaches at ETH Zurich as well as at the University of Bergen. Contact: System Dynamics Group, Department of Geography, University of Bergen, Post box 7800, 5020 Bergen, Norway; telephone: +47 555-83-092; fax: +47 555-83-099; e-mail: birgit.kopainsky@geog.uib.no.

Stephen M. Alessi is an associate professor and teaches instructional design and technology at the University of Iowa in the United States. His research includes the application of cognitive learning theory to the design of simulations, multimedia and on-line instruction. He is co-author (with Stanley Trollip) of Multimedia for Learning: Methods and Development. Contact: 370 Lindquist Center, University of Iowa, Iowa City Iowa 52242, United States; telephone: +1 319-335-5568; fax: +1 319-335-6145; e-mail: steve-alessi@uiowa.edu.

Matteo Pedercini is Director of Modeling and Capacity Development at the Millennium Institute, in Washington, DC. He holds a PhD and a Master in System Dynamics from the University of Bergen. Matteo Pedercini teaches System Dynamics applied to development planning at the University of Bergen, Norway. His research interests include development planning, System Dynamics, and scenario analysis. Matteo Pedercini has been consulting for several governments and international agencies in the field of development planning and has created numerous simulation models to support medium and long-term strategic planning. Contact: Millennium Institute, 1634 Eye Street, NW Suite 300, Washington DC 20006, United States; telephone: +1 202-383-6200; fax: +1 202-383-6209; e-mail: mp@millennium-institute.org.

Pål I. Davidsen (professor, Mag. Art) is educated in mathematics and information science with a speciality in system dynamics (SD). His research interests are in SD methodology, in the application of SD modelling, analysis and policy design across a wide variety of fields in social sciences and in the development of SD-based multi-user interactive learning environments. He served as Associate Chair of the Pre-College Education Project, Sloan School of Management, M.I.T. (1990 – 91) and as Visiting Scholar in the System Dynamics Group, M.I.T. (1986 and 1990 – 91). He is a former President of the System Dynamics Society (2003) and was a Managing Editor of the System Dynamics Review (2007). Contact: System Dynamics Group, Department of Geography, University of Bergen, Post box 7800, 5020 Bergen, Norway; telephone: +47 555-84-134; fax: +47 555-83-099; e-mail: pal.davidsen@geog.uib.no.

Appendix

Appendix 1: Instructions

You have just been elected the Prime Minister of Blendia. You will stay in office as prime minister for a period of 50 years. You are thus in charge of the long term development of Blendia.

Blendia is an island located off the western cost of Africa. It is currently one of the poorest countries in the world with an income per capita of $300 per year. Your task is to bring your country onto a sustainable economic growth path and achieve and maintain the highest possible income per capita.

Income per capita results directly from production and production is driven by the available capital (machinery and its technology level) as well as by total factor productivity. As a government you cannot invest in capital directly. However, you can improve the general investment environment. Investors in capital will invest the potentially available money (a share of per capita income) more when the labor force is more productive and roads provide access to input and output markets for the goods produced. Specifically, you can invest in the following three resources:

- Education
  Education is the stock of knowledge, skills, techniques, and capabilities embodied in labor acquired through education and training. These qualities are important for the labor force to understand and perform tasks, to properly use the available physical capital, and to efficiently organize the production process. Maximum or optimal education would mean an average adult literacy rate of 100% (which is the maximum or optimal value for Human Development Index calculations).
- Health
  Health defines the strengths of the labor force and thus its capability to properly use the available physical capital and to efficiently organize the production process. Maximum or optimal health would mean an average life expectancy of 85 years (which is the maximum or optimal value for Human Development Index calculations).
- Roads
  Efficient and extended infrastructure allows faster and cheaper access to the market, broader access to information, and reliable access to the inputs required for production. Maximum or optimal roads would mean a value of kilometers of roads per person equal to those in the year 2005 in the United States.

Budget issues. For making your investment decisions you will have to take a number of budget mechanisms into account.

Your expenditures for education, health and roads are driven by two sources:

- Revenue: Through taxation (30% tax rate) the government generates revenue from per capita income.

- Borrowing: You can borrow money from foreign sources. If you borrow money you start accumulating debt. Each year you will have to pay interest on your debt. The interest rate depends on your debt. A common measure for the amount of debt is the debt over GDP ratio. The interest rate is 1% for a very low debt over GDP ratio and can rise up to 15% for a very high debt over GDP ratio. For simplicity, assume that per capita GDP is equal to per capita income.

In Blendia, government development expenditure is the total revenue (the revenue generated through taxation from per capita income) minus interest payments on debt.

Decisions. Every five years, as part of a national development planning effort, you will decide on the expenditures for education, health and roads. You can do three things, and as the prime minister you have the absolute power to decide (see also Figure A):

1. Distribute more than the total available development expenditure. In this case you borrow money and create a deficit.
2. Distribute less than the total available development expenditure. In this case you will have a surplus and be able to service (pay down) debt or lend money.
3. Distribute the total available development expenditure without creating either a deficit or a surplus.

Figure A: Budget decisions mechanism with initial values

Government development expenditure	$90 per person
– Education expenditure	$30 per person
– Health expenditure	$30 per person
– Transportation expenditure	$30 per person
Surplus (+)/deficit (-)	0

Evaluation. Your performance will be evaluated based on the following criteria:

- Income per capita: You should try to achieve and maintain the highest possible income per capita. The country's official goal is to reach a value of $600 per capita in 50 years.
- Interest payments on debt: Per capita income can only be maintained if you have not accumulated excessive debt. At the end of the 50 year period the interest payments on debt will be deducted from your income per capita in that year.

Appendix 2: Reading task for the control group

Now that you have been introduced to the imaginary country known as Blendia, let's describe the dynamics underlying its economic development as influenced by investments in the health, education, and roads sectors.

The dynamics of Blendia is described by what we call a System Dynamics model. A System Dynamics model is a set of numeric equations representing the variables in a system, such as a nation's economy, and their cause-effect relationships. The variables in those equations stand for the many quantities relevant to a system such as a nation, for example, per capital income, government development budget, capital investment, debt, and so on. The cause-effect relationships within the equations show how each variable may cause other variables to increase or decrease, and how much. For example, higher per capita income tends to increase the government development budget which in turn tends to decrease debt. For capturing national economic development processes several such variables and cause-effect relationships are necessary.

The variables of a System Dynamics model are of two main types, known as stocks and flows. Stocks represent the amount of something, such as the amount of money, the kilometers of roads in a country, etc. Stocks can increase, decrease, or stay the same. If no investment and depreciation take place, the amount of capital stays the same. With investment, the amount of capital increases. If capital depreciates, the amount of capital decreases. If capital is both being invested and depreciated, the amount of capital changes by the difference between investment and depreciation, so if 4 units of capital are invested per year and 7 units depreciate, the net amount of capital goes down by 3.

Flows represent rates of change to stocks. Investment is a flow of new capital into (or increasing) the stock of existing capital. Depreciation is a flow of capital out of (or decreasing) the capital.

Flows have a direct effect on stocks in that a flow into a stock increases its value and a flow out of a stock decreases its value. The amount of capital invested each year (a flow) directly increases the overall capital (a stock). The amount of capital that depreciates each year (also a flow) directly decreases the overall capital.

Stocks may have a causal connection to flows, including their own flows (whether flowing in or out). For example, the greater the capital (a stock), the more capital will depreciate (a flow) each year.

Other variables (sometimes called auxiliaries or converters) may represent intermediary values. These may affect or be affected by stocks, flows, or other intermediate variables. An example might be the per capita income, that is, the average income for each person in the nation. That intermediate variable is affected by productivity (how much work the labor force accomplishes) and in turn affects other intermediate variables such as the government development budget.

Given these three types of variables, let's describe the structure of Blendia's economy as it operates in this simulation.

The model contains five main stocks, the capital per person, the actual level of education relative to the optimum level of education, the actual level of health relative to the optimal level of health, the actual kilometers of roads relative to the optimum kilometers of roads, and the debt per person in the nation. The capital per person is directly changed by two flows, capital investment which increases it, and capital depreciation which decreases it. The actual level of education relative to the optimum level of education is affected by one flow, the change in amount of education, which may be positive or negative. The same is true for health and for roads, which are affected respectively by the change in health and the change in roads, each of which may be positive or negative. Finally, the dept per person has two flows, borrowing, which increases debt per person, and paying back what was borrowed, which decreases debt per person.

These flows are directly affected by the three variables you will be controlling in the simulation activity, the desired government budgets for education, health, and roads.

For example, the desired budget for education affects, after a delay, the change in education flow. The desired budget for education also affects the governments total desired budget, which affects government borrowing or debt payments, which then affect the debt per person, which affects the actual government budget. In turn, the relative level of education affects national productivity and national investment, thus creating a feedback loop. So while the desired government budget for education eventually affects education, that in turn eventually affects the government budget, including that for education. Such a feedback loop takes a long time to complete (several years), so a change in the education budget affects some things quickly (such as national debt) but other things after a long time (the nations relative level of education and subsequently the government budget).

Similar loops operate for health expenditures and road expenditures. While they affect the current national debt very quickly, they take some time to improve the quality of health and roads, and thus national productivity and per capita income.

This simulation deals with three "input" variables that you (the government) have total control over, the budgets for education, health, and roads. It simulates the effects of those input variables on five important "output" variables, ones that you (again, representing the government) do not directly control, but only indirectly control by setting the input variables. The five output variables are the per capita income corrected for interest payments, the actual resources relative to maximum resources, the ratio of debt over the GDP, the level of deficit (or surplus) per person, and the actual level of capital per person.

Although the relationships are complex and change over time, they tend to operate like the following. Increasing any of the input variables a small amount increases the government development budget per person slowly over time, while increasing the input variables a lot will cause a rapid decrease in the government development budget. Increasing any of the input variables to any extent will increase the actual resources relative to maximum resources somewhat quickly at first, but more slowly later on. Increasing any of the input variables a small amount will increase the debt over GDP ratio slowly, but if input variables are increased too much they will increase that ratio much more quickly. Finally, increasing the input variables to any extent will slowly increase the actual capital per person. Of course, increasing or decreasing the input variables at the same time may have greater or more complicated effects.

Your goal in this activity is to use the simulation to learn as much as you can about the relationships between variables, so that you can manage the country successfully, which we define as maintaining the highest possible per capita income corrected for interest payments on debt.