Developing Seventh Grade Students' Understanding of Complex Environmental Problems with Systems Tools and Representations: a Quasi-experimental Study

Zerrin Doganca Kucuk[1] · Ali Kerem Saysel[2]

Abstract A systems-based classroom intervention on environmental education was designed for seventh grade students; the results were evaluated to see its impact on the development of systems thinking skills and standard science achievement and whether the systems approach is a more effective way to teach environmental issues that are dynamic and complex. A quasi-experimental methodology was used to compare performances of the participants in various dimensions, including systems thinking skills, competence in dynamic environmental problem solving and success in science achievement tests. The same pre-, post- and delayed tests were used with both the comparison and experimental groups in the same public middle school in Istanbul. Classroom activities designed for the comparison group ($N = 20$) followed the directives of the Science and Technology Curriculum, while the experimental group ($N = 22$) covered the same subject matter through activities benefiting from systems tools and representations such as behaviour over time graphs, causal loop diagrams, stock-flow structures and hands-on dynamic modelling. After a one-month systems-based instruction, the experimental group demonstrated significantly better systems thinking and dynamic environmental problem solving skills. Achievement in dynamic problem solving was found to be relatively stable over time. However, standard science achievement did not improve at all. This paper focuses on the quantitative analysis of the results, the weaknesses of the curriculum and educational implications.

Keywords Systems thinking skills · Environmental education · Dynamic problem solving · Dynamic modelling

✉ Zerrin Doganca Kucuk
zerrin.doganca@rc.bau.edu.tr

[1] BAUSTEM Research Center, Bahcesehir University, Istanbul, Turkey

[2] Institute of Environmental Sciences, Boğaziçi University, Istanbul, Turkey

Introduction

Most of today's socio-environmental problems are dynamic and complex in nature. Systemic complexity consists of accumulations, delays, feedbacks and non-linearities that are ubiquitous in biophysical as well as socio-economic systems. The mismatch between the systems complexity and its widely shared mental models creates failure in solving many of the environmental problems that humanity faces. Contamination of the environment arises from the accumulation of a difference between the level of pollutants discharged and their rate of assimilation in atmospheric, oceanic and terrestrial sinks. There are time lags between action and response, which make it difficult to learn about and control pollution. People have fundamental difficulties in perceiving basic accumulation processes (Sweeney & Sterman, 2000; Ossimitz, 2002; Cronin & Gonzalez, 2007). Moxnes and Saysel (2009) found that people's understanding of atmospheric CO_2 accumulation is flawed and that misconceptions about basic stock-flow dynamics lead them to make systematically biased decisions as to emission control and therefore overshoot suggested climate targets. Examples can be generalised to other cases, such as chlorofluorocarbon (CFC) emissions and the stratospheric ozone problem, where there are considerable time lags between CFC emissions and ozone depletion and between the elimination of CFCs and the recovery of the stratospheric ozone layer. Pest control practises on farmlands cause higher pest populations and even pest outbreaks, as genotypes susceptible to chemical pesticides die and resistant genotypes flourish, urging farmers to apply higher pesticide rates, which result in higher overall pest resistance. Introducing alien species into agro-ecosystems for pest control may result in invasion by species. Ecologists are becoming increasingly concerned about non-linearities with tipping points and irreversible changes that increase the vulnerability of species and ecosystems (Scheffer et al. 2001; Duarte et al. 2012).

Education should aim at preparing individuals to face both current and future environmental challenges. Improving people's understanding of complex system structures and dynamic behaviour patterns of environmental problems requires development of systems thinking skills (STS). There are various approaches to systems thinking. Riess and Mischo (2010) described systems thinking as "*the ability to recognise, describe and model complex aspects of reality as systems*" (p. 707). Sweeney and Sterman (2000) view STS as:

- Identifying

    – stocks (accumulations)
    – flows (their rates of change)
    – delays (distance in time between cause and effect)
    – feedbacks (reinforcing and counteracting loops)
    – non-linearities (unproportional response in effect with respect to its cause),

- defining the boundaries of both mental and formal models
- reasoning on certain behaviour patterns created through the interaction of various system components. This paper approaches STS as it was proposed and classified by Sweeney and Sterman (2000).

There is no consensus in the literature as to the distinction between the terms "environmental education", "education for sustainable development" and "sustainability education" (Malone et al. 2004; Locke et al. 2013). However, Tilbury (2004) identified the main goal of education for

sustainability as motivating people to "*change towards sustainability*" (p. 104). In his analysis of the literature on systems theory, operations research, and organisation theory, Porter (2009) concluded that systems thinking was strongly associated with sustainability education, as it makes students conceptualise and acknowledge complexities of environmental problems and enables them to develop and apply solutions. In line with the attempts to maintain sustainability, Sweeney and Sterman (2007) stressed the role of education in the raising of "*systems citizens*" who are able deal to with today's complex problems. The framework of this study was structured upon the literature on environmental education. Several studies in the field of environmental education argue for the development of STS (Grotzer & Basca 2003; Assaraf & Orion 2005; Evagorou et al. 2009; Riess & Mischo, 2010). On the other hand, there are institutional barriers to the development of STS in environment/science classes in schools. One impediment is the heavy load of factual knowledge, with limited and inadequate connections between fragments of knowledge (Brown, 1992). Hannon and Ruth (2000) criticise it for emphasising individual facts and events, rather than processes and interrelations. Another drawback is the limited ecological content in school curricula (Grotzer & Basca, 2003). Taking these impediments into account, the present study aims to design a systems-based environmental instruction, focusing on the interconnections and dynamics in nature. The designed intervention fits exactly into the schedule for teaching the Turkish Science and Technology Curriculum Human and Environment Unit to seventh grade students. It is one of the attributes of the design that makes it a good substitute for the existing curriculum.

In this study, several instruments were developed and implemented to measure educational gains. The abbreviations for these instruments are presented in Table 1. Throughout the text, rather than the explicit titles, corresponding abbreviations will be used. Moreover, some concepts specific to the systems terminology are explained within the legends next to or below the related figures.

Previous Studies on Teaching Systems Thinking Skills in Environment Classes

In the literature, there are some misconceptions regarding even basic concepts and processes related to ecosystems. There is also a lack of holistic understanding of ecosystems at various grade levels. For instance, first-grade students often do not include plants in their list of living organisms (Stromenn, 1995), and Leach et al. (1996) revealed high school students' reductionist understanding of prey-predator relationship. Assaraf and Orion (2005) identified misconceptions of junior high school students regarding the water cycle.

Table 1 List of abbreviations used in the text

Abbreviation	Explanation
DEST	Dynamic Environmental Scenarios Test
DEST_f	Familiar components in Dynamic Environmental Scenarios Test
DEST_unf	Unfamiliar components in Dynamic Environmental Scenarios Test
SCAT	Science Achievement Test
STS	Systems Thinking Skills
STRST	Systems Thinking Required Skills Test
STST	Systems Thinking Skills Test
_del	Extension for delayed test
_pre	Extension for pre-test
_post	Extension for post-test

Researchers design various systems interventions to support teachers when dealing with their students' misconceptions and deficient understanding of ecological issues. Such interventions typically consist of instruction on basic systems concepts, practice with user-friendly computer simulation models and comparison of the learning outcomes with those obtained through standard instruction. One example is by Riess and Mischo (2010), who designed two interventions, one on basic systems concepts and the other exploring a computer simulation model of a forest ecosystem. In their pre-test-post-test research design with a sample of 424 sixth grade students in Germany, they discovered that a combination of the two methods resulted in significantly higher achievement scores in conceptual understanding of prey-predator relationship. They also found that the combined method resulted in a significantly higher conceptual understanding. Previous environmental education studies on promoting STS focused mostly on the understanding of ecosystem structures. Unlike those attempts, Evagorou et al. (2009) concentrated on the development of a number of STS. They conducted a computer simulation based intervention. Thirteen 11–12 years old students were exposed to five 90-min sessions about changes in a marsh ecosystem over time. The study had a pre-test-post-test design, and post-intervention observations revealed that the participants were able to identify elements, recognise boundaries, identify subsystems of a system and make appropriate decisions to manage the marsh ecosystem. However, they were unable to identify feedback mechanisms.

Purpose of the Study

The way STS is explored in this study does not only focus on understanding of systems structures, but also on students' understanding of dynamic system behaviour. For this reason, the present study intends to teach STS in order to interpret systemic complexities of certain environmental problems and study system behaviours dependent on system structures. The instructional design incorporates hands-on dynamic modelling as well as systems tools and representations. Furthermore, unlike other studies, delayed tests were introduced to measure long term educational attainments in STS and competence in dynamic problem solving as an additional measure. Studying the transfer of STS to familiar and unfamiliar contexts is another endeavour of this study.

For this study, the seventh grade science chapter "Human and Environment" in the Science and Technology textbook was redesigned and constructed with a systems approach, and its comprehensive impact on various student skills was compared with the identical educational attainments obtained through standard instruction of the Science and Technology Curriculum of the Ministry of Education in Turkey. Human and Environment is the only chapter covering ecological issues in the Science and Technology Curriculum. It includes the concepts of population, habitat and ecosystems. The ultimate teaching goal of this chapter is for students to be able to discuss various environmental issues and make inferences as to how environmental problems will affect the future of their local and global environment (Ministry of Education Board of Education, 2013). However, the curriculum presents environmental content on a factual basis. One can have a look at the questions at the end of the chapter to understand the factual (knowledge-based) nature of the chapter (Tunç et al. 2011). None of these questions are divergent (open-ended) questions that lead learners to go beyond the facts and make inferences about processes and behaviour over time. Moreover, feedback causalities are not emphasised. For example, the chapter starts with an activity on the relationship between several

environmental concepts. The expected outcome of this activity is development of uni-directional cause-effect relationships, although the presented concepts are highly nested and inter-connected. Another deficiency is the lack of numerical and/or graphical activities in teaching population dynamics. Considering the presented environmental education literature and the drawbacks of the Human and Environment chapter, the main motivation of this paper is to answer the question of whether a systems approach is a more effective way to teach environmental issues that are dynamic and complex to seventh grade students (12–14 year olds). To elaborate on this question, a number of research questions were formed:

- Does systems-based intervention significantly improve systems thinking skills of students?
- Does systems-based intervention significantly improve understanding of dynamic environmental problems?
- Are there any significant differences between science achievement levels on environmental questions of the subjects in the experimental and the comparison group immediately the intervention and after 6 months?

To obtain answers to the research questions, a number of instruments were developed to measure students' systems thinking required skills and STS, competence in dynamic environmental problem solving and science achievement in line with the chapter studied. In order to increase the relevance of the program for middle schools, our intervention was designed in line with the objectives of and the time allocated by the Science and Technology Curriculum.

Context of the Study

School System

The education system in Turkey is centrally controlled by the Ministry of Education. The Turkish education system has experienced a series of radical reforms since the foundation of the Republic of Turkey. After the most recent major reform in 2013 (publicly known as "4 + 4 + 4"), the duration of compulsory education was increased to 12 years, with separate four year compartments of primary, middle and high school education. Education is a free public service; however, some state schools collect tuition fees through school-parent associations. According to Bakioğlu and Kirişçi-Sarıkaya (2015), more than 98% of students attend public schools. Students' placement in public primary and middle schools is based on their residential district; however, before they graduate, students take a highly competitive nationwide test for their placements in different high schools.

According to the OECD report (2013), "Turkey has a higher than average proportion of underperforming students" (p. 4) at international exams such as PISA exam in 2009 regarding the reading, science and mathematics performances of 15-year-olds. The socio-economic status of students is found to be an important factor affecting student achievement (OECD 2013).

Student Context

The public school selected for this study can be seen as representative of most of the middle schools in Istanbul with respect to students' educational and socio-economic background. The school is located at a working class district mostly made up of low- and middle-income households. In the school, there were 52 seventh grade students in two different classes, and all of them were present at the beginning

of the study. The classes were not categorised according to their academic achievement. There was one science teacher at the school and she followed the regular Science and Technology Curriculum and the suggested course books. During the course of the study, some students attended preparatory courses for the high school entrance exam that they were to take prior to graduation. The number of students attending the preparatory courses increased the following year, at the beginning of which they took the delayed tests.

Teaching Sequences and Pedagogy

Both groups were instructed by the researcher her/himself to eliminate any possible effects of teacher attitudes on comprehension. The Human and Environment chapter covers four sections: introduction to concepts in ecology, food chains and webs, biodiversity, and environmental problems. The classroom activities in the textbook mostly include fill-in-the blanks, matching, and short-answer type of items. There is one group activity on recognising different ecosystems. In other words, there is no variety in the instructional tools used. Science and Technology Curriculum (2013) allocated 14 lesson hours to cover this chapter. Both standard and systems-based instruction continued for a month, including the administration of pre- and post-tests. Descriptions of standard instruction for the comparison group and the systems-based instruction for the experimental group are presented below.

Standard Instruction Standard instruction was restricted to the Human and Environment chapter plan suggested by the Science and Technology Curriculum. The activities and content knowledge were taken from the course book and student workbook (Tunç et al. 2011). Standard instruction involves some systems elements, such as the hierarchical organisation, from "atoms" to "the planet Earth", and cause-effect relationships between the elements of an ecosystem. However, looking at it from the systems perspective, standard instruction lacks crucial systems elements and processes. First of all, the standard approach focuses on the facts and events in ecosystems, rather than on the processes. For instance, facts about various environmental problems are introduced without mentioning future effects of the problems on ecosystems and any further estimates as to the fate of affected populations. Secondly, the standard approach claims to generate a holistic understanding of ecosystems, such as food chains or the relationships between biotic and abiotic factors within an ecosystem. However, from the dynamic systems perspective, it lacks the crucial element of feedback and feedback thinking, which is fundamental to the dynamics of environmental systems. Lastly, the standard approach does not enable the study of population dynamics, such as focusing on changes in different populations over time. The content is presented as descriptive, without referring to any mathematical tools such as graphs and charts, or related skills to utilise these tools for creating systems models and predicting the behaviour of environmental systems.

Systems-Based Instruction The systems-based instruction aims at developing STS to improve students' understanding of complex, dynamic environmental issues. The emphasis is on presenting environmental content with systems tools (e.g. Figs. 1 and 3) and representations (see Fig. 2).

The system instructions started with an introduction to system dynamics. These classes aimed at introducing systems in general by means of causal loop diagrams, stock-flow diagrams, and construction of simple dynamic models on STELLA® (Systems Thinking for Education

Research Software). The systems-based activities were adopted from "Road Maps—a Guide to Learning Systems Dynamics" (System Dynamics in Education Group, 2005) and were modified to include terms and concepts more familiar to Turkish students.

After the introductory classes, the systems-based instruction continued with an activity on interrelating ecosystem concepts and variables, where causal loop diagramming was used as the instructional tool. One of the participants in the activity sketched the loop in Fig. 1 and said: "*As human population increases, hunting will increase. As hunting increases, there will be less species, and eventually human population will decrease.*"

The goal and perspective of the systems-based instruction is to introduce population dynamics into the first part of the Human and Environment chapter. This was accomplished through the stock-flow modelling activity, adopted from the book *Shape of Change* (Quaden et al. 2008). In this activity, the participants were expected to construct a stock-flow diagram of a tree population in a forest (Fig. 2) and to hand simulate the behaviour over time on a table. After examining changes in the stocks of seeds, sprouts, saplings and trees, a classroom discussion took place on behavioural responses of various variables in the model. One focus was the time lag between seeding and growing saplings, i.e. seeding unable to replace the harvest immediately.

The food chains and webs in the second section were introduced and explored through causal loop diagramming. The circular causalities in the food chain exercises represent information flows within an ecosystem that generate energy flows. This circular view of ecosystems helps to understand population dynamics in the long term, as opposed to the limited information provided by uni-directional energy flows exercised in the Turkish Science and Technology Curriculum.

For the last section on environmental problems, the experimental group created a dynamic computer simulation model on bioaccumulation as illustrated in Fig. 3. The cover story was real and took place in Clear Lake, California (Giusti, 2009), where mercury contamination from an evacuated mine intruded into the food chain. The modelling activity included the steps of identifying the variables, drawing stock-flow diagrams, specifying units, writing equations and implementing and simulating the models on STELLA®. After simulating their final models, the participants were asked questions about equilibrium mercury concentrations in fish and human populations. They were expected to explore ways of reducing mercury concentration in humans by conducting simulation experiments.
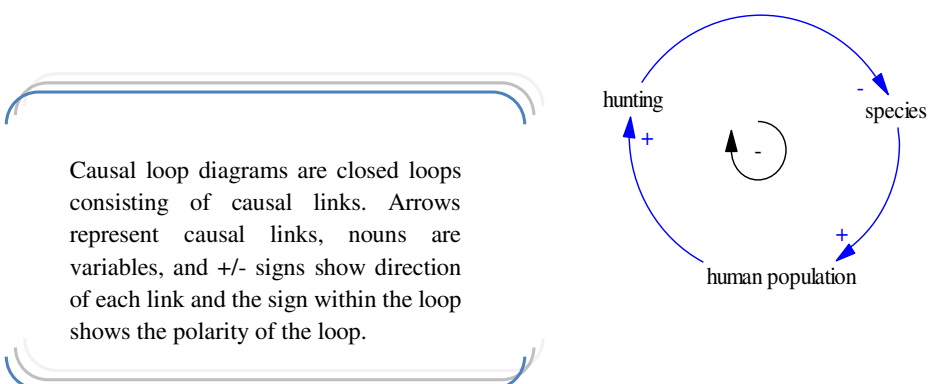
Causal loop diagrams are closed loops consisting of causal links. Arrows represent causal links, nouns are variables, and +/- signs show direction of each link and the sign within the loop shows the polarity of the loop.

Fig. 1 A causal loop diagram drawn by a subject during "Relating Environmental Concepts" activity
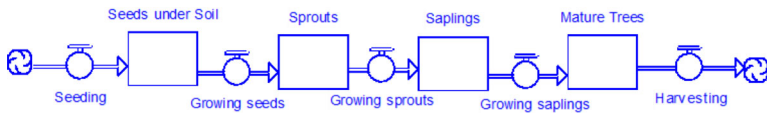
Stock-flow diagrams show the structural elements that produce system behavior by showing stocks and flows. Stocks, represented with rectangles, are simply accumulations in a system. Flows, represented with taps, are rates of change.

Fig. 2 Stock-flow diagram drawn during "Trees in a Forest" activity

Methodology

Design of the Study

The research design is quasi-experimental, due to non-random assignment of the participants. Gribbons and Herman (1997) support quasi-experimental research in studies where effects of certain educational programs are evaluated and in cases where it is not feasible to make random assignments of participants. The study consists of both quantitative and qualitative analyses; however, in this paper, we focus on the quantitative ones. Table 2 summarises the research design and the flow of activities. Uppercase letters, A and B, stand for the alternate forms of the tests on Table 2. Alternate forms refer to two different sets of questions with the same measurement purposes so as to avoid problem solving based on memorised facts and answers.
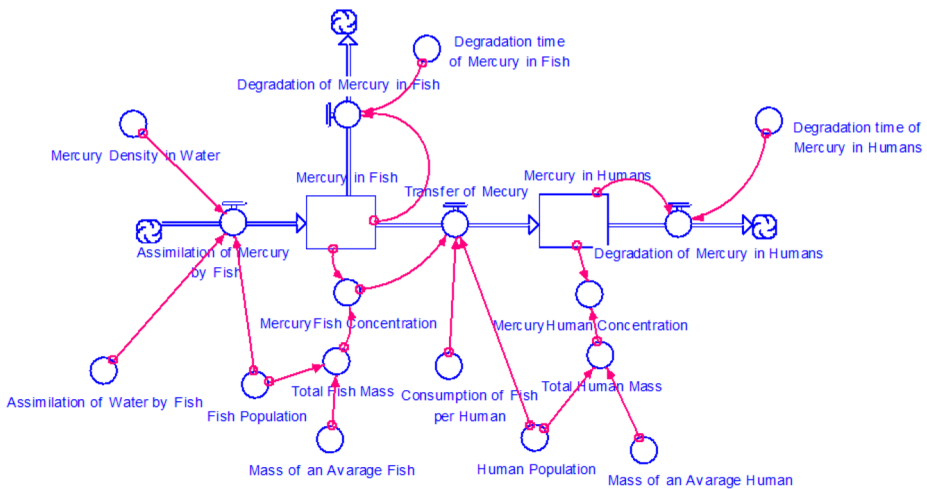
A dynamic model deals with relationships that vary with time. The circles in the model are converters and they can be either constants, or calculated with formulae, or identified with graphs. The arrows represent connectors that carry information from one element to another in the model. They can also either be constants, or algebraic functions, or graphical relationships.

Fig. 3 Bioaccumulation model created by the students on STELLA

Table 2 Design of the study

Experimental group	Comparison group
Pre-tests	Pre-tests
- Demographic Information Sheet	- Demographic Information Sheet
- Systems Thinking Required Skills Test (A)	- Systems Thinking Required Skills Test (A)
- Systems Thinking Skill Test (A)	- Systems Thinking Skill Test (A)
Introduction to system dynamics	Meeting with students
Systems-based instruction	Standard instruction
Immediate post-tests	Immediate post-tests
- Systems Thinking Required Skills Test (B)	- Systems Thinking Required Skills Test (B)
- Systems Thinking Skill Test (B)	- Systems Thinking Skill Test (B)
- Science Achievement Test	- Science Achievement Test
- Dynamic Environmental Scenarios	- Dynamic Environmental Scenarios
Delayed post-tests	Delayed post-tests
- Systems Thinking Required Skills Test (B)	- Systems Thinking Required Skills Test (B)
- Systems Thinking Skill Test (B)	- Systems Thinking Skill Test (B)
- Science Achievement Test	- Science Achievement Test
- Dynamic Environmental Scenarios	- Dynamic Environmental Scenarios

A pilot study was carried out in Spring 2011 to diagnose potential problems with instruction and instruments. Inter-rater reliability of the instruments was evaluated through random assignment of the student tests to three experts; two system dynamics specialists and one science educator. The instruments were tested with a number of students from two private schools to study reliability. Modified versions of the Systems Thinking Skill Test (STST) items adopted from the well-known systems tasks (Sweeney and Sterman, 2000, 2007; Ossimitz, 2002) increased face validity of the test. To ensure construct validity, one Turkish teacher and a graduate student studying Turkish literature examined the wording of the teaching materials and the tests developed. One science educator from a state university examined the tests by considering the cognitive level and science background of seventh grade students. Two system dynamics experts from a state university evaluated all the teaching materials and the tests for content validity. Based on feedback from these experts, all the lesson materials and the instruments were revised.

Sample of the Study

The study took place in a state middle school in Istanbul, Turkey, in Spring 2012. Among 52 seventh grade students at the school, 42 attended the classes regularly and took all the tests. Hence, the results are reported for all 42 participants. One of the two seventh grade classes (ages between 12 and 14 years) was selected as the comparison group ($N = 20$) and the other as the experimental group ($N = 22$). The distribution of the sample according to sex is summarised in Table 3.

Table 3 Distribution of sex between the groups

		Sex		Total
		Female	Male	
Groups	Experimental	12	10	22
	Comparison	9	11	20
Total		21	21	42

Instruments

Sweeney and Sterman (2007), Nuhoğlu (2008) and Boersma et al. (2011) stressed the lack of instruments to evaluate systems thinking skills of children and designed specific instruments for the purpose of their own studies. With a focus on environmental issues and by taking into account the specific age group of the participants, a number of instruments and their alternate forms were developed for this study. The content and the role of the instruments within the study are explained below.

Systems Thinking Required Skills Test

Systems Thinking Required Skills Test (STRST) consists of four questions. The questions aim at measuring the skills that are required for learning and assessing systems thinking, as proposed by Sweeney and Sterman (2000). These skills include the ability to read and interpret graphs, create graphs from data, tell a story from a graph, create a graph of behaviour over time from a story, identify units of measurement and understand the basic concepts of probability, logic and algebra.

Systems Thinking Skills Test

The questions in the Systems Thinking Skills Test (STST) were designed in a way that a "natural system thinker" (Sweeney and Sterman, 2007, p. 285) would respond to without any need for knowledge on specific systems terms like feedback loop, stocks, flows, etc. The systems thinking literature includes some specifically designed tasks, such as the Bathtub Task by Sweeney and Sterman (2000) and the Department Store Task by Sterman (2002). However, the literature lacks in tasks addressing other STS attributes.

The STST includes questions adapted from well-known systems tasks (see Appendix 1). The feedback thinking question was modified and translated from the "Systems-based Inquiry Protocol" designed by Sweeney and Sterman (2007). The delay question was developed by the researchers. There are two questions addressing stock-flow thinking. The first question was inspired by the Federal Deficit Task, designed by Ossimitz (2002), and was written in a "true/false" format. The theme and content of the question were simplified for children. The second stock-flow thinking question was also inspired by the "Arrivals and Departures in the Alpenhotel Work" task, designed by Ossimitz (2002). The theme was transferred to focus on the number of passengers in a bus in order to create a more familiar context for the participants.

Dynamic Environmental Scenarios Test

The Dynamic Environmental Scenarios Test (DEST) has two purposes. First of all, it aims to reveal whether the systems-based and standard instructions would result in any differences in understanding of dynamic socio-environmental problems. Secondly, this instrument is designed to assess different levels of transfer of learning regarding the dynamic and complex environmental issues.

The DEST includes authentic questions related to local environmental problems and they are free from any systems-specific terms. DEST contains familiar (DEST_f) and unfamiliar components (DEST_unf) to test near and far transfer, respectively (Perkins and Salomon, 1992). The familiar components were built around population dynamics and bioaccumulation, while the

unfamiliar components were built on the construction of the third bridge in Istanbul (see Appendix 2) and waste collection problem.

Science Achievement Test

The questions in the Science Achievement Test (SCAT) were selected from two randomly selected science test-banks (Güvender, 2009; Oran, 2008) and modified for this study. SCAT was prepared by taking into account the objectives listed in the Science and Technology Curriculum. The test is a standard one assessing students' comprehension of the "Human and the Environment" chapter.

Descriptive and Reliability Statistics of the Tests

Table 4 shows the descriptive statistics for all tests, as well as their respective reliability coefficients. It is observed that the reliability coefficients are within the acceptable ranges for most of the tests. Alternate forms of STRST and STST were also developed and implemented. Spearman-Brown coefficients for the alternate forms are reported on Table 4. DEST includes open-ended questions, so inter-rater reliability was studied for this instrument. As the intra-class reliability coefficient implies (on Table 4), two raters were in almost complete agreement.

Analysis of the Data

The design of the study enables assessment of between group and within group statistics. Since each group contains less than 30 participants, the normality of the scores for each test should also be tested. For sample sizes smaller than 50, using Shapiro-Wilk for testing normality is recommended (Razali and Wah, 2011). After the normality test, it would be better to choose parametric tests for normal distributions and non-parametric tests for the non-normal ones.

For the between group comparisons, one-way ANOVA was used for normal distributions and Kruskal-Wallis one-way analysis of variance test was used for non-normal distributions. For the within group statistics, repeated measures ANOVA and Friedman tests were used for the normal and non-normal distributions, respectively. To study performances of the participants on specified skills, the items were also analysed separately by using a chi-square test.

While analysing DEST and DEST_del, analysis of covariance (ANCOVA) was applied to control "pre-treatment group differences" by taking into account any covariates (Huck, 2012).

Table 4 Basic information about the tests

Test	Number of items	Max. possible score	Length of test (min)	Type of questions	Reliability test	Reliability coefficient
STRST	9	5.5	15	Filling in the blanks	Cronbach's α	0.62
				True-false short-answer	Spearman-Brown	0.64
STST	11	11.5	40	Filling in the blanks	Cronbach's α	0.73
				True-false short-answer	Spearman-Brown	0.69
SCAT	26	23	30	Multiple choice Matching	Cronbach's α	0.76
DEST	10	25	40	Essay	Cronbach's α	0.74
					Intra-class	0.91

To decide on the covariates, a correlation matrix was formed. This matrix included demographic data (parental education, science and math grades from previous semesters), STRST_pre and STST_pre scores. Among these variables, the significantly correlated ones with DEST and DEST_del scores were selected as covariates to run the ANCOVA test.

Results

Table 5 presents the basic descriptive statistics of all tests at each stage. The significance values for Shapiro-Wilk tests with an asterisk ($p < .05$) represent non-normal distributions on the table. For these distributions, non-parametric tests were applied, as presented in the following sections.

Table 6 presents the between group comparisons. STST_post, DEST and DEST_del tests result in statistically significant differences (marked as italics) between the groups. Profile plots in Fig. 4 show between-group and within-group differences. To further explore the research questions, individual tests are analysed in more detail in the following sections.

Systems Thinking Required Skills Test Results

Figure 4a compares STRST scores between and within groups. Trends for both groups appear to be similar. There were no significant differences between the groups at .05 significance level (Table 6). This is an expected result, since instruction did not include any specific content related to the mathematical skills addressed on STRST. To examine within group dynamics on STRST, ANOVA was used. Statistically significant differences are reported on STRST scores, both for the experimental ($F(2, 40) = 10.32$, $p < .001$) and the comparison groups ($F(2,

Table 5 Descriptive statistics for all tests

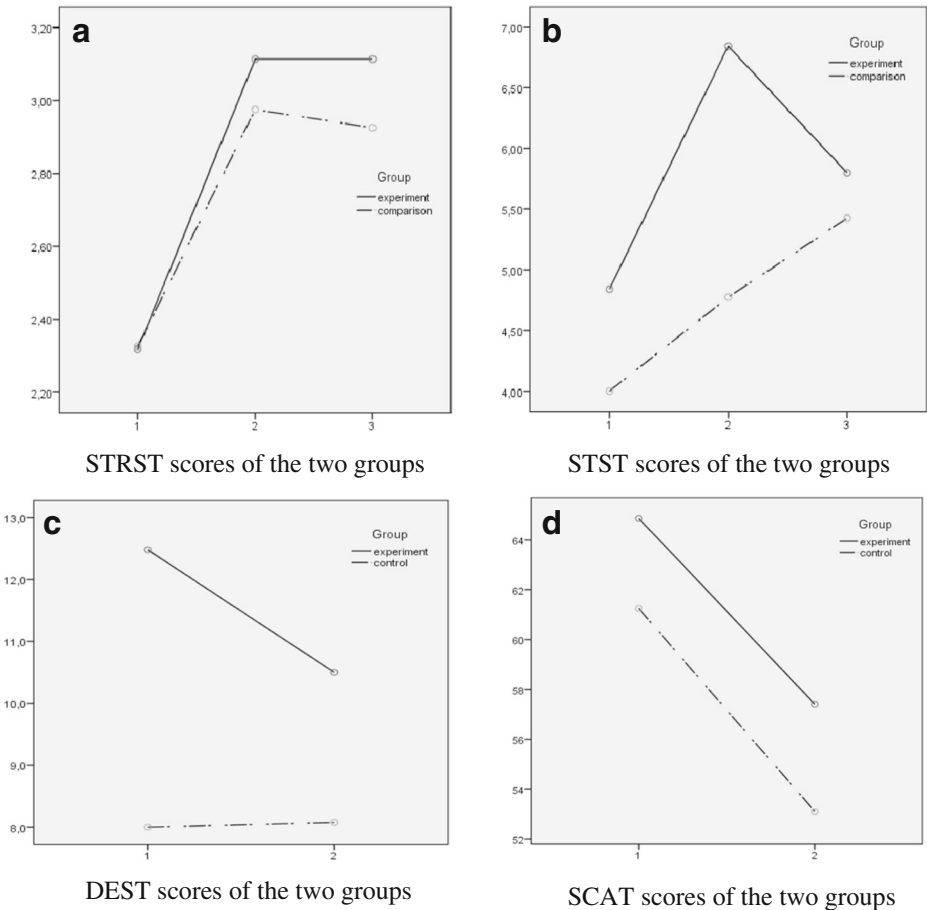|	Group	Number	Mean	Range	Std. dev.	Unanswered items (%)	Shapiro-Wilk statistics	Shapiro-Wilk sig.
STRST_pre	Exp.	22	2.32	0.5–4.5	0.91	13.63	0.86	0.01*
	Com.	20	2.33	0.5–4.5	1.04	10.5	0.90	0.04*
STRST_post	Exp.	22	3.11	1–4.5	0.84	3.03	0.82	0.01*
	Com.	20	2.98	1.5–4.5	1.02	13.80	0.91	0.06
STRST_del	Exp.	22	3.11	2.5–5	0.94	2.68	0.84	0.01*
	Com.	20	2.93	1.5–5	0.98	10.61	0.81	0.01*
STST_pre	Exp.	22	4.84	2.5–10.5	2.37	6.61	0.96	0.63
	Com.	20	4.00	0.5–10.5	2.56	10.45	0.89	0.03*
STST_post	Exp.	22	6.84	2.5–10	1.96	1.23	0.96	0.54
	Com.	20	4.78	1–10	2.88	3.63	0.95	0.34
STST_del	Exp.	22	5.80	1.5–8.5	2.14	1.86	0.93	0.12
	Com.	20	5.43	1–9.5	2.24	2.01	0.97	0.83
DEST	Exp.	22	12.48	5–21.5	4.93	0	0.94	0.18
	Com.	20	8.00	3–20	4.59	4.50	0.94	0.25
DEST_del	Exp.	22	10.50	6–15.5	3.28	1.15	0.95	0.38
	Com.	20	8.08	2–18	4.18	3.56	0.97	0.71
SCAT	Exp.	22	64.86	29–91	17.32	0.82	0.90	0.05
	Com.	20	61.25	28–93	20.97	1.95	0.91	0.09
SCAT_del	Exp.	22	57.41	23–87	20.53	0	0.94	0.22
	Com.	20	53.10	13–88	19.06	0	0.93	0.15

*Significance values for Shapiro-Wilk tests ($p < .05$) represent non-normal distributions

Table 6 Group comparisons for all the tests

	F	Sta. sig.	Effect size	Effect size category
STRST_pre	0.00	.98	−0.01	Very small
STRST_post	1.22	.28	0.14	Very small
STRST_del	0.23	.63	0.19	Very small
STST_pre	0.23	.63	0.34	Small
STST_post	7.52	.01*	0.84	Large
STST_del	0.43	.52	0.17	Very small
DEST	9.22	.01*	0.94	Large
DEST_del	4.42	.04*	0.64	Medium
SCAT	0.03	.87	0.10	Very small
SCAT_del	0.07	.79	0.08	Very small

* significant at $p \le .05$

38) = 9.85, $p < .001$). It is found that the main difference for both groups lies in the difference between STRST_pre and STRST_post scores, and the difference seems to persist even six months after the instruction (STRST_del), based on the Bonferroni post-hoc test results.

Fig. 4 Profile plots for each test

STRST scores of the two groups

STST scores of the two groups

DEST scores of the two groups

SCAT scores of the two groups

Performances of the subjects on individual STRST items are also examined with a chi-square test. No significant differences on STRST items are reported between groups at .05 significance level. Based on the frequencies reported on STRST, it is found that the students were able to draw and interpret graphs. However, they had difficulties with converting units and writing equations, and the difficulties persisted even after the instructions.

The interesting point is that both groups developed the required skills throughout the study. Since the groups exhibited parallel developments in STRST scores, it can be argued that these developments were independent of the different methods implemented throughout this study. One possible explanation is that the students received the same mathematics instruction from the same teacher, parallel to our science classes during the intervention period. Seventh grade students learn about graphs and data analyses during the second half of the academic year. This mathematics chapter is totally related to STRS, and this instruction might have affected their scores on STRST.

Systems Thinking Skills Test Results

Figure 4b compares STST scores between and within groups. The highest peak on the plot was observed as the mean STST score of the experimental group immediately after the training. However, there was also a decline in the STST scores of the experimental group six months after the training. An increase in STST mean scores was also observed for the comparison group. To compare STST scores of each group before, immediately after and 6-month after training, a one-way repeated measure ANOVA test was used. A significant effect of the systems-based instruction was observed ($F(2, 40) = 23.42$, $p = .01$, $\eta^2 = .39$), while the effect of standard instruction was not statistically significant ($F(2, 38) = 3.00$, $p = .06$; $\eta^2 = .21$). This finding positively answers the first research question.

For closer examination of STST performances, the frequencies of the individual STST items were analysed. No significant differences among the item frequencies were found on STST at .05 significance level. It was found that there was a positive effect of systems-based instruction on development of STS; however, the effect did not seem to last long.

Dynamic Environmental Scenarios Test Results

Figure 4c shows the trends in DEST scores for the two groups. As an answer to the second question regarding the comparison of DEST scores, it can be said that there is a significant difference between the groups on DEST ($F(40) = 9.22$, $p = .01$, $d = 0.94$) and DEST_del tests ($F(40) = 4.42$, $p = .04$, $d = 0.64$). The decrease in the experimental group's scores on the DEST_del seems striking, but it is not reported as statistically significant at .05 significance level ($F(21) = 2.81$, $p = .11$, $d = 0.47$). DEST scores of the comparison group do not seem to change after a 6-month period ($F(19) = 0.11$, $p = .92$, $d = 0.01$).

To test whether any pre-determined covariates affect the statistically significant difference obtained for DEST and DEST_del, ANCOVA test was used. Among the possible covariates, science grade (sci_gra), mathematics grade (math_gra) and STRST_pre were found to be significantly correlated to DEST scores. It was observed that DEST scores still showed significant differences between the groups when these significantly correlated covariates were taken into account (refer to group (covariate) values on Table 7).

Sci_gra and STRST_pre were significantly correlated to DEST_del scores. Although the statistical difference between groups on DEST_del diminished when sci_gra was controlled

Table 7 ANCOVA test results for DEST and DEST_del scores with the statistically significant covariates

Test	Source of variance	df	F	Sig.	Partial eta squared (η^2)
	sci_gra	1	15.08	.01*	0.28
DEST	Group (sci_gra)	1	4.55	.04*	0.11
	math_gra	1	6.27	.02*	0.14
	Group (math_gra)	1	6.1	.02*	0.14
	STRST_pre	1	9.98	.01*	0.20
	Group (STRST_pre)	1	11.36	.01*	0.23
	sci_gra	1	5.64	.02*	0.13
DEST_del	Group (sci_gra)	1	1.86	.18	0.05
	STRST_pre	1	11.12	.01*	0.22
	Group (STRST_pre)	1	5.59	.02*	0.13

* significant at $p \leq .05$

with ANCOVA test, DEST_del scores were still significantly different for the groups when STRST_pre were taken into account (Table 7).

The most remarkable differences was observed in DEST scores. DEST results were also disaggregated into its familiar (DEST_f) and unfamiliar (DEST_unf) components. It was found that there are statistically significant differences between groups both in their DEST (Table 6), DEST_f and DEST_unf scores (Table 8). DES_unf and DES_unf_del test scores show non-normal distributions. Hence, non-parametric tests were used, and the corresponding results are reported in Table 8. The most striking result is that the experimental group got significantly higher scores on DEST_del_unf than the comparison group, with a very high effect size ($U = 61.5$, $p < .00$, $d = 1.63$). The main motivation of the study is to examine any possible transfer of STS to dynamic environmental questions that are not introduced in instruction. These results support this motivation.

For individual DEST item analysis, a chi-square test was used. Chi-square test results indicate that there are statistically significant differences between the groups on the third bridge ($\chi^2(1, 42) = 6.64$, $p = .01$) and bioaccumulation ($\chi^2(1, 42) = 4.55$, $p = .03$) questions on DEST at .05 significance level. No statistically significant differences were observed between categories for the groups on the DEST_del test at .05 significance level.

Table 8 Descriptive and inferential statistics for DES_f and DES_unf scores

	Group	Number	Mean	Std. dev.	t	Sig.
DES_f	Exp.	22	8.25	3.280	2.617	.012*
	Comp.	20	5.60	3.275		
DES_f_del	Exp.	22	6.25	2.434	1.741	.089
	Comp.	20	4.88	2.685		
	Group	Number	Mean rank	Sum of ranks	Mann-Whit. U	Asymp. sig.
DES_unf	Exp.	22	26.20	576.50	116.50	.008*
	Comp.	20	16.33	326.50		
DES_unf_del	Exp.	22	28.70	631.50	61.50	.000*
	Comp.	20	13.58	271.50		

Science Achievement Test Results

Figure 4d shows trends of SCAT and SCAT_del scores for the two groups. Regarding the third research question on science achievement levels of the participants, it can be said that the experimental SCAT scores are higher but the differences are not reported as statistically significant at .05 significance level for both SCAT and SCAT_del, respectively ($F(40) = 0.03$, $p = .87$, $d = 0.10$ and $F(38) = 0.07$, $p = .79$, $d = 0.08$). This is an expected result because the questions on the SCAT are factual (knowledge-based) and convergent (with finite sets of answers) type of items. Indeed, science textbooks approved and suggested by the Ministry of Education do not include any divergent (open-ended) questions that lead learners to go beyond factual knowledge about the Human and Environment chapter selected for this study.

Discussion

The aim of the study was to investigate whether an appropriate systems approach provides a more effective basis to teach environmental issues, which are dynamic and complex, for seventh grade students (12–14 years old). In accordance with the aim, participants' STS, competence in dynamic environmental problem solving and transfer of STS to familiar and unfamiliar contexts and science achievement were assessed with a number of instruments developed for this study. Based on the results of the STST and DEST, significant results favouring systems-based instruction were reported.

Between group comparisons of STST scores reveal that mean scores of the groups on STST_pre are not significantly different at .05 significance level, but the experimental group scored statistically higher than the comparison group on STST_post. The difference did not last for 6 months and it diminished in STST_del (Table 6). Time was an important barrier to sustainability of the systems thinking skills regarding the experimental group. Resnick (1996) studied high school students and determined that they were able to deal with complex systems using a modelling program called StarLogo. However, she also pointed to the vulnerability of their skills over time. This is the case with the experimental group that did well on STS. However, their performance did not last long in the absence of any systems-based instruction.

Regarding the repeated measures statistics, STST scores exhibit an expected trend for the experimental group, as in many studies in different fields with pre-post-delayed test designs (Quirk et al., 1998; Cho, 2010; Larson and Zahner, 2011). A statistically significant increase on STST_post scores and a statistically significant decrease on STST_del scores were observed. The increase on STST_post scores was already expected after an intensified systems-based instruction. However, the effect of the instruction did not last long enough to be observed on the delayed post-test applied six months after the post-test. An interesting output was observed for the comparison group, as they increased their mean scores on STST throughout the study (Fig. 4b). One possible explanation for this increase on STST scores is their attendance to preparatory courses for the high school entrance exam. Although learning was contaminated (Yıldıran, 2006) by a long summer holiday and attendance of some of the participants to out-of-school preparatory courses, it should be noted that the increases were not reported as statistically significant on repeated measures ANOVA test.

DEST exhibits more enduring effects for the experimental group. The experimental group scores are significantly higher than those of the comparison group both on

DEST and DEST_del at .05 significance level. The statistical significance on DEST still exists when the covariates sci_gra, math_gra and STRST_pre are controlled with ANCOVA tests. The statistical difference also exists for DEST_del when the covariate STRST_pre is controlled with ANCOVA. The covariate sci_gra becomes important when we compare DEST_del scores between the groups six months after the instruction. In other words, the presence of more participants in the experimental group who were better at science courses could account for the difference of performance on DEST_del test. This finding can be explained using information processing theory. According to this theory, new information is processed in the mind and is associated with existing and relevant knowledge. It is assumed that an organised mind creates more associations between pieces of information (Schunk, 2012). Hence, more successful students are expected to have more organised minds, at least regarding school subjects, and thus tend to recall knowledge better in the long run.

Familiar and unfamiliar items of DEST were investigated for between groups and repeated measure bases. This revealed that the experimental group scores are significantly better on DEST_f part, but the difference does not exist for DEST_f_del scores at .05 significance level. When the items on DEST are examined individually, performances on the third bridge question (feedback thinking), bioaccumulation (modelling) and fish population (stock-flow thinking) questions are found to be statistically significant on chi-square tests. However, performances on these questions were not statistically significant on DEST_del. One of the possible reasons for decreasing mean score on DEST_f_del is that the modelling question, which the experimental group did significantly better on DEST, demands high cognitive load. Modelling is a high-level STS (Stave and Hopper, 2007) that needs to be practised for a long period of time for retention. Correspondingly, the item on fish populations has to do with stock-flow thinking, which is a challenge even for adults. Ossimitz (2002) and Sweeney and Sterman (2000) presented specifically designed tasks on stock-flow thinking to university students in Austria and USA, respectively. They reported the incompetency of their subjects in stock-flow thinking. Regarding the unfamiliar DEST questions, the feedback question has an up-to-date and popular theme focusing on the construction of the third bridge in Istanbul. The participants in the experimental group were able to transfer feedback thinking skill to a subject that was not taught in the instruction. However, the number of the participants in the comparison group that were able to elaborate on the third bridge issue increased 6 months after the instruction probably due to effects of mass media. For the other unfamiliar DES question (waste management; see Appendix 2, question 3), the number of the participants who were able to identify the delay in the landfill was the same for both DES and DES_del. Hence, the scores on DES_unf_del are still statistically significant in favour of the experimental group. This is an important finding that implies far transfer of the skills tested with unfamiliar questions on DEST.

To conclude, the subjects in the experimental group were able to perform some STS after an intensified systems instruction and they were able to use some of these skills in some familiar and unfamiliar environmental tasks, even after 6 months. Evagorou et al. (2009) determined that 11–12 years old students were able to perform a number of STS, such as identifying elements, boundaries and subsystems of a system, making some changes in order to observe certain system behaviours after five sessions with ecosystem simulations. However, they had difficulties in performing feedback thinking. In this study, subjects treated with the systems-based approach

were able to achieve feedback thinking by introducing food chains with CLDs rather than emphasising a linear, one-way representation between organisms on a food chain. Presenting food chains in a loop fashion enables 13–14 years old students to engage in deeper discussions on changes in populations and natural resources over time and their adverse effects on one another. Liu and Hmelo-Silver (2009) also supported this finding by suggesting inclusion of conceptual representations of natural systems to science classes to achieve understanding of complex phenomena.

Limitations of the Study

The outcomes of this study depends on the specific teaching environment (the public school), the participant pool (the 42 students with specific demographic backgrounds) and the practitioner researcher's skills in classroom instruction and engagement. To be able to draw stronger conclusions, the instructional design and the instruments should be implemented in different teaching environments to larger groups and by different teachers.

In this study, the researcher held a dual role both as a researcher and a teacher. In the literature, there are a considerable number of studies where researchers acted both as teachers and researchers (Ball, 2000; Lampert, 1986; Simon, 1995; Taylor, 2004; Wong, 1995).With the exception of action research carried out by in-service teachers to identify instructional problems and to improve teaching in their own classes, these researchers carried out their research as teachers in authentic learning environments to understand the dynamics from an insider's perspective (Wong, 1995). However, this duality brings the risk of researcher bias. This is related to unintentional behaviours by the researcher that might affect the results in favour of the experimental group. Gay et al. (2006) suggested "blind" scoring of the tests in which the researcher is unaware of whose tests are assessed. In this study, along with the researcher/teacher, a second researcher blind scored the tests to avoid the researcher bias and to study inter-rater reliability. The decision to place the researcher as teacher was intentional because in-service teachers do not have any background in the teaching of system dynamics and teaching with the systems approach. Hence, rather than training teachers on the systems approach, the intervention was directly tested on the students.

The overall study made use of both quantitative and qualitative methodology. When the researcher's dual role as a teacher and a researcher is examined, there appears the question of credibility. Credibility is the conformity between participants' responses and the researcher's reflection on their viewpoints in qualitative studies (Merriam, 1998). Guba and Lincoln (1989) coined the term "progressive subjectivity", where researchers create their own constructs rather than getting "stuck on some intermediate construction" (p. 238). They suggest documenting and monitoring the study with various techniques, such as reports, field notes and reflective commentaries. Both during the pilot study and the actual study, all sessions were video recorded, and a teaching assistant took field notes to monitor the process. These documents enabled the researchers to revise the content of the intervention program and modify how the researcher/teacher approached the participant group following to the pilot study. Besides, they also served as evidence for the dual role of the researcher.

Various measurement instruments were developed to match the instructional design. A trade-off was observed between the reliability and validity of some of these instruments. For example, the reliability for STRS could be considered weak, with coefficients below

0.70. Omission of selected questions in STRS could have improved this coefficient. However, that raised doubts concerning validity, since such omissions implied elimination of questions related to specific systems thinking skills suggested by Sweeney and Sterman (2000). This problem is clearly observed when the percentages of unanswered items on STRS in Table 5 are examined.

Another issue is the nature of written tests, as the participants tended to show limited effort to complete their answers (Patton, 1983). The participants' willingness to answer was generally higher during the interviews supplemented with appropriate probes. Hence, interviews and qualitative analysis enlarge the information pool on which the researchers can build their observations and conclusions. This study also benefited from interviews and qualitative analysis. Since the insights from the qualitative analysis did not significantly change our findings but did strengthen them, this paper solely focused on the quantitative analysis.

Finally, it should be noted that, when the study covers a longer time period, it is more likely that the participants will show improvement regardless of the training. Gay et al. (2006) identified this threat to validity and called it "history" (p. 238), referring to any unplanned events occurring during the course of the study. In this study, students' participation in the high school entrance exam preparatory courses is a threat that probably resulted in increases in STS scores for the participants in the comparison group.

Conclusion

To understand natural phenomena, science education has shifted its focus to teaching complex systems (Assaraf et al. 2013). Yoon (2008) advocates the systems approach as the best teaching method to overcome difficulties in understanding the underlying mechanisms of natural phenomena. In their recent study, Vachliotis et al. (2014) have concluded that, as students' STS develop, their meaningful understanding of related scientific concepts improves. In this context, the study aims to develop the systems thinking skills of seventh grade students to help them better address dynamically complex environmental problems. The quasi-experimental study shows that the instruction plan and the instruments developed are promising tools that enable seventh grade students to develop appropriate systems concepts and deal with dynamically complex environmental problems. In this study, participants' performances on STST and DEST synchronised with the schedule of the training in both areas, and hence strengthened our conclusion. Moreover, unlike standard instruction, the systems-based instruction is designed to include a combination of teaching methods, such as using dynamic simulations, classroom discussions and instructional activities with a student-centred perspective, as demonstrated by Grotzer and Basca (2003) and Riess and Mischo (2010). Turkish Science and Technology Curriculum (STC) adopts educational goals emphasising improvement in students' capacity to infer and create knowledge through interaction and participation. System dynamics in education, in general, and the activities and instruments developed for this study, in particular, can reconcile such educational goals with appropriate classroom practises.

Efforts in teaching about the field system dynamics at various grade levels are not new and have diverse applications (Lyneis & Fox-Melanson, 2001; Nuhoğlu, 2008; Stave & Hopper,

2007; Sweeney & Sterman, 2007; Zaraza & Fisher, 1999). Our results also show that there are immediate gains in basic STS and competence in dynamic problem solving; however, these learning achievements faded over time. For long-term gains, similar instructional tools and activities on various subjects should be presented in different courses at the same time.

To encourage systems-based education and to study the effects of different systems-based education designs, the first step should be teacher education. Educating teachers is a more sustainable way than conducting single studies dependent on researchers also acting as practitioners. A possible further study could focus on teacher education and on examining the outcomes, both on teachers and their students, over a longer period of time. Educating science and biology teachers on both the systems approach and its applications in human-ecosystem interactions should be a priority in the field of environmental education and in education for sustainability.

Appendix 1. Examples of STST Questions

1)   (*Feedback question*)

    A.   Explain the relationship between "hunger" and "eating". Please complete the sentences below and specify the relationship.

> *As our level of hunger increases, we eat _____.*
> *As we eat more, our level of hunger_____.*

    B.   Explain the relationship between "number of chickens" and "number of chicks". Please complete the sentences below and specify the relationship.

> *As number of chickens increase, number of chicks_____.*
> *As number of chicks increase, number of chicken _____.*

    C.   Examine the relationships on the questions a and b. Based on the phrases "increases" and "decreases" compare and contrast these relationships.

2)   (*Stock-flow thinking question*) Erdal cannot handle the pocket money that his parents give and he sometimes borrows money from friends. The amount of money he has borrowed and he has paid back within the first 3 months of the year is listed below:

- He borrowed 20 TL from Murat and paid back 30 TL to Sezin in January.
- He borrowed 10 TL from Sezin and paid back 10 TL to Murat in February.
- He borrowed 10 TL from Murat and made no payment to Sezin in March.

To help Erdal make calculations about his debt, please put "X" to the box you prefer for each statement.

	True	False	I do not know
a. In January, there has been a decrease in Erdal's debt.	X		
b. His total debt in January and February are equal.	X		
c. His total debt has not changed in March.		X	
d. At the end of the three months, there has been a decline in Erdal's debt.		X	

Appendix 2. Examples of DEST Questions

1) Greenpeace Mediterranean made a call for a ban on fishing and trading of juvenile fish. Greenpeace volunteers distribute rulers with corresponding fish lengths to fishermen and consumers.

The Ministry of Agriculture declared the legally fishable, minimum bluefish size as 14 cm. However, bluefish are able to reproduce when they reach at 20–25 cm length (hint: assume that mature bluefish do not eat juvenile bluefish).

    a. If fishermen fish juvenile bluefish, what will happen in time to the

       i. juvenile bluefish population,
      ii. mature bluefish population,
     iii. overall bluefish population?

    b. Propose a policy for saving the bluefish population and explain its reasons.

2) Constructing the third bridge in Istanbul is a hot topic in Turkey.

    a. Some professionals argue that "each bridge creates its own traffic." Can you explain this sentence? (*Hint:* When you are explaining the statement, you should to take into account people's tendency to drive and the new settlements that may be constructed around the bridge.)

    b. If you are the mayor of Istanbul, what will be your alternative solutions to solve the traffic problem in Istanbul? Please give two suggestions for the traffic problem and explain your reasons.

3) Wastes are collected from a district once a week and they are transferred to a landfill. These wastes include plastics, paper, glass and metal substances and they are not sorted (hint: Glass wastes are decomposed approximately in 1000 years. Plastics are decomposed 10–1000 years according to their types. Paper is decomposed in 3 months, and newspapers are decomposed in 3–12 months).

    a. If these wastes are allowed to decompose by themselves (i.e. in case of no intervention), which changes will you expect within the landfill? What is the fate of this landfill?

    b. What do you suggest for more effective usage of this landfill? Please, explain your suggestions.

References

Assaraf, B. O., Dodick, J., & Tripto, J. (2013). High school students' understanding of human body system. *Research in Science Education, 43*, 33–56.

Assaraf, B. O., & Orion, N. (2005). Development of systems thinking skills in the context of Earth system education. *Journal of Research in Science Teaching, 42*(5), 518–560.

Bakioğlu, A., Kirisci-Sarıkaya, A. (2015). Eğitimde Özelleştirme (Eng. Privitazation in Education). Nobel Press, Istanbul, Turkey.

Ball, D. L. (2000). Working on the inside: using one's own practice as a site for studying teaching and learning. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 365–402). Mahwah, NJ: Lawrence Erlbaum Associates.

Boersma, K., Waarlo, A. J., & Klaassen, K. (2011). The feasibility of systems thinking in biology education. *Journal of Biological Education, 45*(4), 190–197.

Brown, G. S. (1992). Improving education in public schools: innovative teachers to the rescue. *System Dynamics Review, 8*(1), 83–89.

Cho, K. (2010). Fostering the acquisition of English prepositions by Japanese learners with networks and prototypes. In S. Knob, F. Boers, & A. Rycker (Eds.), *Fostering language teaching efficiency through cognitive linguistics* (pp. 259–276). Germany: Hubert & Co..

Cronin, M., & Gonzalez, C. (2007). Understanding the building blocks of system dynamics. *System Dynamics Review, 23*(1), 1–17.

Duarte, C. M., Agusti, S., Wassmann, P., Arrieta, J. M., Alcaraz, M., Coello, A., Marba, N., Hendriks, I. E., Holding, J., Garcia-Zarandona, I., Kritzberg, E., & Vaque, D. (2012). Tipping elements in the artic marine ecosystems. *AMBIO—a Journal of the Human Environment, 41*(1), 44–55.

Evagorou, M., Korfiatis, K., Nicolaou, C., & Constantinou, C. (2009). An investigation of potential of interactive simulations for developing systems thinking skills in elementary school: a case study with fifth-graders and sixth-graders. *International Journal of Science Education, 31*(5), 655–674.

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research- competencies for analysis and applications* (8th ed.). Ohio: Pearson.

Giusti G. A. (2009). Human influences to Clear Lake, California. Retrieved November, 2013, from: http://www.lakecountywinegrape.org/growers/growersfiles/Human%20Influences%20on%20Clear%20Lake%202009.pdf.

Gribbons B., Herman J. (1997). True and quasi-experimental designs. Practical assessment, *Research & Evaluation, 5*(14). Retrieved November, 2013, from: http://pareonline.net/getvn.asp?v=5&n=14.

Grotzer, T. A., & Basca, B. B. (2003). How does grasping the underlying causal structures of ecosystems impact students' understanding? *Journal of Biological Education, 38*(1), 16–29.

Güvender Press (2009). *7th grade 100% SBS science and technology test book [Turkish: 7. Sınıf %100 SBS Fen ve Teknoloji Soru Bankası].* Istanbul, Turkey.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Lampert, M. (1986). Knowing, doing, and teaching multiplication. *Cognition and Instruction, 3*(4), 305–342.

Larson, R., & Zahner, S. (2011). The impact of web-delivered education on preceptor role self-efficacy and knowledge in public health nurses. *Public Health Nursing, 28*(4), 349–356.

Leach, J., Driver, R., Scott, P., & Wood-Robinson, C. (1996). Children's ideas about ecology 3: ideas found in children aged 5-16 about the interdependency of organisms. *International Journal of Science Education, 18*, 19–34.

Lyneis, D. A., Fox-Melanson, D. (2001). The challenges of infusing system dynamics into a K-8 curriculum, *19th International Conference of the Systems Dynamics Society*, Atlanta, Georgia, 23–27 July, 2001.

Hannon, B., & Ruth, M. (2000). *Dynamic modeling* (2nd ed.). New York: Springer.

Huck, S. W. (2012). *Reading statistics and research* (6th ed.). Boston: Allyn & Bacon.

Liu, O. L., & Hmelo-Silver, C. E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching, 46*, 1023–1040.

Locke, S., Russo, R. O., Montoya, C. (2013). Environmental education and eco-literacy as tools of education for sustainable development. *Journal of Sustainability Education, 4* Retrieved February 2016, from: http://www.jsedimensions.org/wordpress/content/environmental-education-and-eco-literacy-as-tools-of-education-for-sustainable-development_2013_02/.

Malone, K., Fien, J., Guevara, J., and Lang, J. 2004. *Education for sustainable development: Australia. Presented at UNESCO-NIER Regional Seminar on Policy Research and Capacity Building for Education Innovation for Sustainable Development*, 27th July–3rd August 2004, Tokyo Japan.

Merriam, S. B. (1998). Qualitative research and case study applications in education. San Francisco: Jossey-Bass.

Ministry of Education Board of Education (2013). "*Talim Terbiye Kurulu Başkanlığı*" (*in Turkish*). Retrieved November, 2013, from: http://ttkb.meb.gov.tr/www/guncellenen-ogretim-programlari-ve-kurul-kararlari/icerik/150

Moxnes, E., Saysel, A. K. (2009). Misperceptions of global climate change: information policies. Climatic Change, 93(1-2), 15–37.

Nuhoğlu, H. (2008). *Studying effects of systems approach on attitude, achievement, and different skills in science and technology lesson [Turkish: İlköğretim Fen ve Teknoloji Dersinde Sistem Dinamiği Yaklaşımının Tutum, Başarıya ve Farklı Becerilere Etkisinin Araştrılması]*, Ph.D. Dissertation, Gazi University.

Oran Yayıncılık (2008). *7th grade SBS science and technology test book [Turkish: 7. Sınıf SBS Fen ve Teknoloji Soru Bankası].* İzmir, Turkey.

Organization for Economic Cooperation and Development (OECD) (2013). Educational outlook: Turkey. Retrieved from: http://www.oecd.org/edu/EDUCATION%20POLICY%20OUTLOOK%20TURKEY_EN.pdf on February 15, 2016.

Ossimitz, G. (2002). Stock-flow thinking and reading stock-flow related graphs: an empirical investigation in dynamic thinking abilities. *20th International Conference of the Systems Dynamics Society*, Palermo, Italy, July 28–August 1, 2002. Retrieved November, 2013, from: http://wwwu.uni-klu.ac.at/gossimit/pap/sfthink.pdf.

Patton, M. Q. (1983). Qualitative Evaluation Methods. (4th Edition). Sage Publications, California, USA.

Perkins, D. N., & Salomon, G. (1992). *Transfer of learning. International encyclopedia of education* (2nd ed.). Oxford: Pergamon Press.

Porter, T. (2009). Three views of systems theories and their implications for sustainability education. *Journal of Management Education, 33*(3), 323–347.

Quaden, R., Ticotsky, A., & Lyneis, D. (2008). *The shape of change: stocks and flows* (Revised ed.). Massachusetts: Creative Learning Exchange.

Quirk, M. E., DeWitt, T., Lasser, D., Huppert, M., & Hunniwell, E. (1998). Evaluation of primary care futures: a faculty development program for community health center preceptors. *Academic Medicine, 73*, 705–707.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21–33.

Resnick, M. (1996). Beyond the centralized mindset. *Journal of the Learning Sciences, 5*, 1–22.

Riess, W., & Mischo, C. (2010). Promoting systems thinking through biology lessons. *International Journal of Science Education, 32*(6), 705–725.

Science and Teachnology Curriculum (2013). "Fen ve Teknoloji Müfredatı" (in Turkish). Retrieved June, 2012, from: http://ttkb.meb.gov.tr

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature, 413*, 591–596.

Schunk, D. H. (2012). *Learning theories: an educational perspective*. Boston: Pearson Inc..

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal of Research in Mathematics Education, 26*(2), 114–145.

Stave, K. A., Hopper, M. (2007). What constitutes systems thinking? A proposed taxonomy. *25th International Conference of the System Dynamics Society*, July 29–August 2, 2007, Boston. Retrieved November, 2013, from: http://www.systemdynamics.org/conferences/2007/proceed/papers/STAVE210.pdf.

Sterman, J. D. (2002). All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review, 18*, 501–531.

Stromenn, E. (1995). Lions, tigers, and bears, oh my!: children's conceptions of forests and their inhabitants. *Journal of Research in Science Teaching, 32*(7), 683–698.

Sweeney, L. B., & Sterman, J. D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review, 16*(4), 249–286.

Sweeney, L. B., & Sterman, J. D. (2007). Thinking about systems: student and teacher conceptions of natural and social systems. *System Dynamics Review, 23*(2–3), 285–312.

System Dynamics in Education Project Group (2005). Roadmaps: a guide to learn system dynamics. Retrieved November, 2013, from Creative Learning Exchange—System Dynamics and Systems Thinking in K-12 Education Website: http://www.clexchange.org/curriculum/roadmaps.asp.

Taylor, S. S. (2004). Presentational form in first person research. *Action Research, 2*(1), 71–88.

Tilbury, D. (2004). Rising to the challenge: education for sustainability in Australia. *Australian Journal of Environmental Education, 20*(2), 103–114.

Tunç, T., Bağcı, N., Yörük, N., Gürsoy Köroğlu, N., Çeltikli Altunoğlu, Ü., Başdağ, G., Keleş, Ö., İpek, İ., Bakar, E. (2011). *Science and technology 7th grade course book [Turkish: Fen ve Teknoloji 7.Sınıf Ders Kitabı]*. MEB Devlet Kitapları, Ankara, Turkey.

Vachliotis, T., Salta, K., & Tzougraki, C. (2014). Meaningful understanding and systems thinking in organic chemistry: validating measurement and exploring relationships. *Research in Science Education, 44*, 239–266.

Wong, E. D. (1995). Challenges confronting the researcher/teacher: conflicts of purpose and conduct. *Educational Researcher, 24*(3), 22–28.

Yıldıran, G. (2006). *Multicultural applications of mastery learning: our thoughts our deeds and our hopes for education*. Istanbul: Bogazici University Press.

Yoon, S. (2008). Using memes and memetic processes to explain social and conceptual influences on student understanding about complex socio-scientific issues. *Journal of Research in Science Teaching, 45*(8), 900–921.

Zaraza, R., & Fisher, D. M. (1999). Training system modelers: the NSF CC-STADUS and CC-SUSTAIN projects. In W. Feurzeig & N. Roberts (Eds.), *Modeling and simulation in science and mathematics education*. NY.: Springer.