ANYWHERE AND ANYTIME: EVALUATING STUDENTS' BEHAVIORS IN SCIENCE WEB BASED LEARNING ENVIRONMENT USING LOG FILE ANALYSIS

Galit Ben-Zadok
Knowledge Technology Lab
School of Education
Tel Aviv University and
The Center for Educational Technology
Israel

Moshe Leiba
Knowledge Technology Lab
School of Education
Tel Aviv University
Israel

Rafi Nachmias
Knowledge Technology Lab
School of Education
Tel Aviv University
Israel

Rachel Mintz
The Center for Educational Technology
Israel

Abstract

The main purpose of this research is to examine the learning processes of elementary school students in science web-based learning environment — at school and at home. To this end, the log files of the learning environment and data mining tools and techniques were used. Results suggest that the school-home gap is starting to fade away (learning wise) and a school-home learning continuum can be established. Part of the learning can be transferred to after school hours and can be evaluated similarly by means of data mining tools. Results also suggest that there are differences between learning at home versus learning at school across ages (such as duration of the learning and pace). Our future work will focus on gathering the learning variables and employing data mining techniques in order to find learning patterns at home and at school across ages.

Introduction

ICT have changed both the scope and the nature of learning, setting up new opportunities for learning, as well as offering different ways of learning. Web-

based learning environments offer students the potential to be autonomous in their learning and study anywhere and anytime, especially outside of the school walls. When using these environments in the classroom, the teacher has significant control over the students' learning processes, whereas outside of the classroom the students have the responsibility for their own learning.

While engaging with web-based learning environments, students leave traces of their activity in the form of log files. These files document each action taken by three basic parameters: what was the action taken, who took it and when. Discovering and extracting educational information from these log files using data mining techniques is an emerging field called Educational Data Mining (EDM).

The basic assumption of our research is that learning processes can be reflected in the student's behavior while interacting with the environment; This behavior might be extracted from the log files and may shed light on learning processes for large populations in ways that were not previously possible (the assessment of learning processes is traditionally examined using qualitative tools and methods with small-scale populations).

The goal of this research is to explore the learning processes of elementary students in a science web-based learning environment, at school and at home, and to compare the both age wise.

Background

Information and communication technologies (ICT) have changed the way people learn in recent years. Web-based learning environments have expanded the concepts of time and space of learning. Learning is no longer confined to the four walls of a classroom. It can take place in a specific classroom, but also anywhere at school or at home, and it does not necessarily require the presence of a teacher and students physically together (Barker, 2000).

Web-based learning environments encourage students to exhibit autonomy and control over their learning process. They are encouraged to be more responsible for their learning, and in many cases they have to plan, carry out and evaluate their own learning processes (Besser & Bonn, 1997; Oliver, 2002). But whereas in the classroom the presence of the teacher can influence the learning process, after school hours the students are being confronted with the tasks alone. Research shows, that a primary characteristic that sets successful online learners apart from their classroom-based counterparts is their autonomy in learning (Keegan, 1996). It also indicates that autonomy and responsibility on the learning process are acquired skills, and by the time the students reach higher education, most adults have acquired a degree of autonomy in learning. Younger students need to have a

scaffold during their learning (Cavanaugh et al., 2004). Research also found that older children have more internal locus of control than younger children (Gershaw, 1989).

Contemporary web-based science learning environments, such as WISE[1] and OFEK[2], offer a rich digital science curriculum, focusing on constructive approach to teaching and learning. These environments contain cognitive tools in which the students can learn, such as virtual models, experiments and simulations. These tools enable inquiry-based learning and visualization of scientific phenomena and processes which could not be demonstrated to students in any other way (Linn, Clark & Slotta, 2003; Osborne & Hennessy, 2003; Voogt, 2008). In addition, such environments may present the students a variety of instructional tools in which they can practice their knowledge, such as games, drill and practice exercises and self-tests. Being offered a variety of tools for learning and practice, students can freely navigate along their own chosen path and control their own learning process according to their preferences and needs. The aspects of control are expressed, for example, as control over content, control over time and pace, and control over the learning sequence (Sims & Hedberg, 1995).

The autonomy that the online learners need and implement in such environments, raises the need to enhance our understanding of their learning behaviors. However, this is not an easy task to achieve with traditional research methodologies, which can hardly cope with gathering of information about the online learners (Nachmias & Hershkovitz, 2007). Data Mining is an emerging methodology in the educational research field, which can advance us towards that goal. While learning with web-based environment students leave continuous hidden traces of their activity in the form of log file records, which document every action taken by three parameters: what was the action taken (e.g., the page URL, the file downloaded), who took it (if the system requires login, this field will usually include the student identification), and when (exact date, time). Researchers use data mining techniques to analyze this data and to locate different aspects of learning behaviors, such as patterns of navigation, time spans and sequences of learning (Romero, Ventura, & Garcia, 2007). Web-based learning environments might also hold information in the log files about the student's profile (e.g., age, gender, grades).

In this research we have analyzed the data derived from the students' log files, in order to enhance our understanding of the online learners' behavior in a web-based

[1] http://wise-dev.berkeley.edu
[2] http://ofek.cet.ac.il

science learning environment, and also to learn about the differences in these behaviors — at school versus at home and across ages.

The Research Objectives

This study explores aspects of elementary students' learning processes in a science web-based learning environment at school and at home. Its specific aims were to:

- Explore and characterize the students' behaviors while engaging with the online environment at school versus at home.
- Explore and characterize the students' behaviors while engaging with the online environment at school versus at home in each specific age group.
- Demonstrate the potential of log file analysis and data mining to evaluate students' learning processes during online learning.

Methods

This framework concerns the characteristics and consequences of the actual usage of the web features within the learning processes in a specific science module. The investigation conducted is of a descriptive nature using quantitative methods (data mining), which have been used in order to explore the students' learning behaviors at school and at home and compare the both. We evaluated students' behaviors by analyzing their log files.

Participants

Participants were 1,671 3rd–6th grade students from different elementary schools in Israel, who learned in a science web-based learning module, as a part of their curriculum. 903 participants (54%) used the module at school, while 768 participants (46%) used it at home. As for the age groups of the students — 316 participants (19%) are in third grade, 325 participants (20%) are in fourth grade, 555 participants (33%) are in fifth grade, and 475 participants (28%) are in sixth grade.

The Learning Environment

We used an Earth Science web-based learning module dealing with the moon phases.[3] This module is a part of OFEK, a web-based learning environment[4] for

[3] http://www.cet.ac.il/ofek/eng/unit11.asp
[4] OFEK was developed by CET— the Center of Educational Technology.

elementary school, in Science, Math and Language. The module includes six different activities implementing four instructional tools: simulation (see Figure 1), three drill and practice exercises, a game and a self-test. The moon phases' simulation appears as the first activity. Following are the three drill and practice activities and a game. In the next page, the students may choose to do an on-line self-test. All of the activities (except for the simulation) offer the students an immediate feedback.

Figure 1: "The Rising of the Moon" Simulation in the Module

Procedure

Log files of a large population (N = 2,643) for six months (September 2008–February 2009) were collected and preprocessed. The study was carried out in four consecutive phases:

Phase I: Data exploration. This phase focuses on understanding the meaning of the metadata, collecting, describing, and exploring the data. In this phase we have examined the dataset and the format of the basic variables.

Phase II: Data preprocessing. This phase focuses on cleaning and formatting the data. The original dataset was much larger at first and was consisted out of N = 2,643 cases. The first stage was to filter all cases who were not students (e.g.

teachers, administrators). At the second stage a filter was applied for keeping students from third to sixth grade (we have decided to focus on students who are learning the subject as a part of their curriculum). At the third cleaning stage a filter was applied for screening students who only entered the environment but did not use the module (spent less than 5 minutes), students who did not log out from the environment (spent more than an hour) and those who spent time in the module at school and at home. The dataset was preprocessed and the final set of cases to be analyzed was defined (N = 1,671).

Phase III: Computing variables. The compatibility of the variable to previous empirical research in this field was taken into consideration, as well as their association to our framework. Algorithms for calculating the variables were formally written and implemented using EXCEL.

Phase IV: Descriptive statistics. Finally, independent *t tests* and ANOVA were performed using SPSS.

Variables

The variables that were extracted and computed from the log file (based on the students' behaviors) in Phase III are described in Table 1.

Table 1: Extracted and Computed Variable List

Variable name	Variable description	Remarks
Locus of Learning	Student's location	8:00-14:00 – School 14:00-8:00 - Home
Time on Task	Time on Task (entire module or activity)	Total learning time in the module or in activity (seconds)
Clicks	Actions taken in the module or in an activity	Number of mouse clicks in the module
Pace	Pace (entire module or activity)	Number of clicks divided by the time on task (clicks/sec)
Completed Activities	Completed Activities	Number of successfully completed activities
Incorrect answers	Errors in all activities	Number of negative feedbacks given to the student in all activities

Results

Students' Online Behaviors at School versus at Home

The analysis of the log file's extracted variables provided a summary of the students' behaviors (Time on task, Clicks and Pace) in the different activities at school and at home. As shown in Table 2, students tend to spend 699.9 seconds at home (SD = 530.6) on all the tasks in the module opposed to 553.3 seconds (SD = 369.4) at school, 27% more time at home. When we check their behavior in each activity we can observe the same tendency (except for the game). They also tend to learn in a slower pace, 0.134 (SD = 0.098) click per seconds, while the number of clicks (actions) stays the same (which means that the amount of activities learnt by the students remains the same at school and at home). As shown in Figure 2, the distribution of the students' Time on task in the module is very similar at school and at home.

Figure 2: Distribution of Time on Task at School versus at Home (N = 1671)

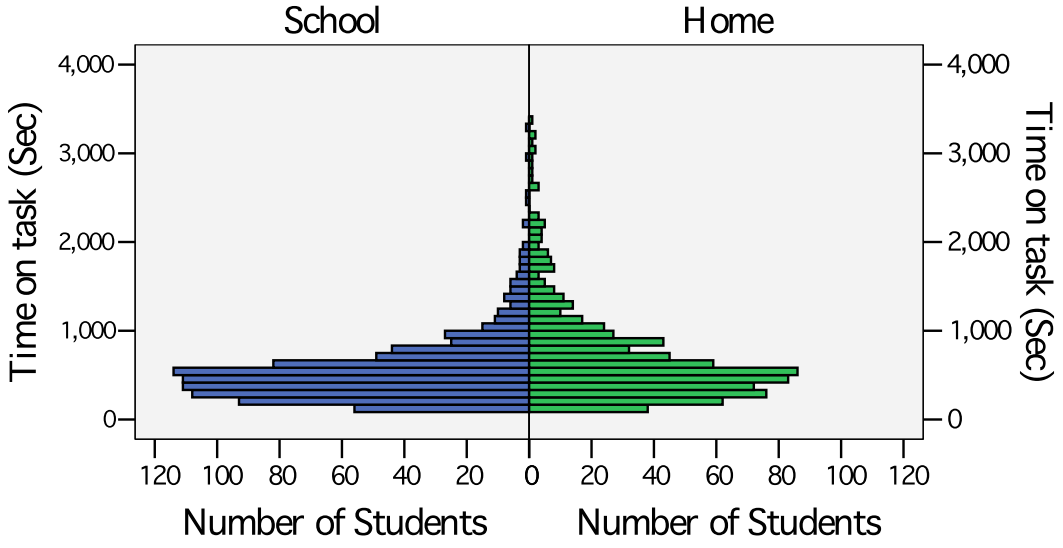


Table 2: Distribution of the students' activities at school and at home (N = 1671)

t	SD	M	Locus of Learning	Behavior	Activity
-3.80**	95.7	84.7	School	Time on Task	**Drag Quest**
	131.6	107.2	Home		
-2.32*	7.87	11.05	School	Clicks	
	11.47	12.23	Home		
2.90**	.142	.192	School	Pace	
	.127	.172	Home		
-2.99**	97.8	91.6	School	Time on Task	**Pull Down Menu**
	163.4	112.6	Home		
-.77	7.7	12.9	School	Clicks	
	10.6	13.2	Home		
2.85**	.117	.177	School	Pace	
	.122	.159	Home		
-2.33*	113.4	81.4	School	Time on Task	**Quiz**
	89.8	94.7	Home		
1.03	9.1	10.8	School	Clicks	
	10	10.3	Home		
4.12**	.148	.178	School	Pace	
	.130	.147	Home		
-3.28**	170.9	100.2	School	Time on Task	**Simulation**
	227.5	135.1	Home		
-1.70	24.5	14.7	School	Clicks	
	27.8	17.1	Home		
4.43**	.317	.238	School	Pace	
	.183	.176	Home		
-5.62**	116.7	200.1	School	Time on Task	**Test**
	201.8	251.9	Home		

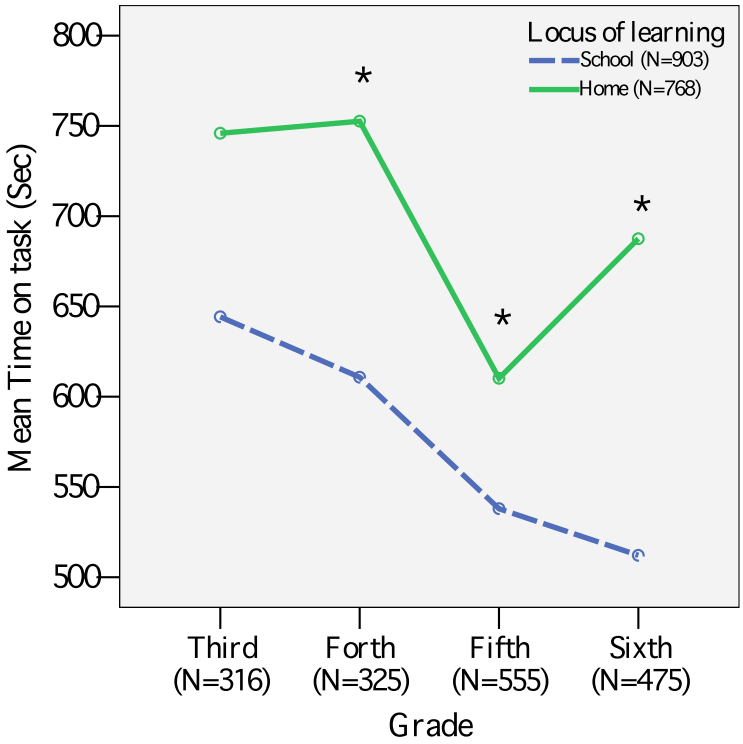
-.33	12.3	19.8	School	Clicks	
	17.2	20.1	Home		
4.16**	.092	.120	School	Pace	
	.076	.100	Home		
-1.51	148.2	138.5	School	Time on Task	**Game**
	181.5	152.4	Home		
1.30	21.9	23.8	School	Clicks	
	22.7	22.1	Home		
4.67**	.087	.189	School	Pace	
	.080	.167	Home		
-6.62**	369.4	553.3	School	Time on Task	**Total Activities**
	530.6	699.9	Home		
-1.41	45.5	74.4	School	Clicks	
	62.6	78.2	Home		
4.47**	.099	.156	School	Pace	
	.098	.134	Home		

*p<0.05, **p>0.01

Students' Online Behaviors at School versus at Home in Different Age Groups

The module's subject matter is a part of the 3rd – 6th grade science curriculum. We distinguished the students by their grade and their locus of learning and characterized their behaviors. The time on task at school and at home distinguished by grade is described in Figure 3. As also shown in previous results (Table 1), students tend to spend more time learning at home than at class (except for 3rd grade). Findings indicate that across grades, younger students spend more time than older students at school and at home ($F(1670) = 11.28$, $p < 0.01$). For example, 3rd grade students spend approximately 650 seconds at home while 6th grade students spend 525 seconds.

Figure 3: Mean Time on Task (Sec) at School and at Home in each Grade (Significant differences of $p < 0.05$ between locus groups are marked with an asterisk).



When examining the students' incorrect answers (Figure 4), 3rd, 5th and 6th grade students tend to make fewer errors when they use the module at home. The figures show that 3rd, 4th and 5th grade students have a similar amount of errors while 6th grade students have lass ($F(1578) = 8.13$, $p < 0.01$).

The students' completion of activities is shown in Figure 5. Only 5th grade students tend to complete more activities at home than at school. When examining the trends across ages, 3rd, 4th and 5th grade students behave differently while 6th grade students behave similarly to 5th grade students and tend to complete more activities. Older students tend to complete more activities than younger students ($F(1670) = 27.82$, $p < 0.01$).

Figure 4: Mean Incorrect Answers (%) at School and at Home in each Grade (Significant differences of $p < 0.05$ between locus groups are marked with an asterisk.)

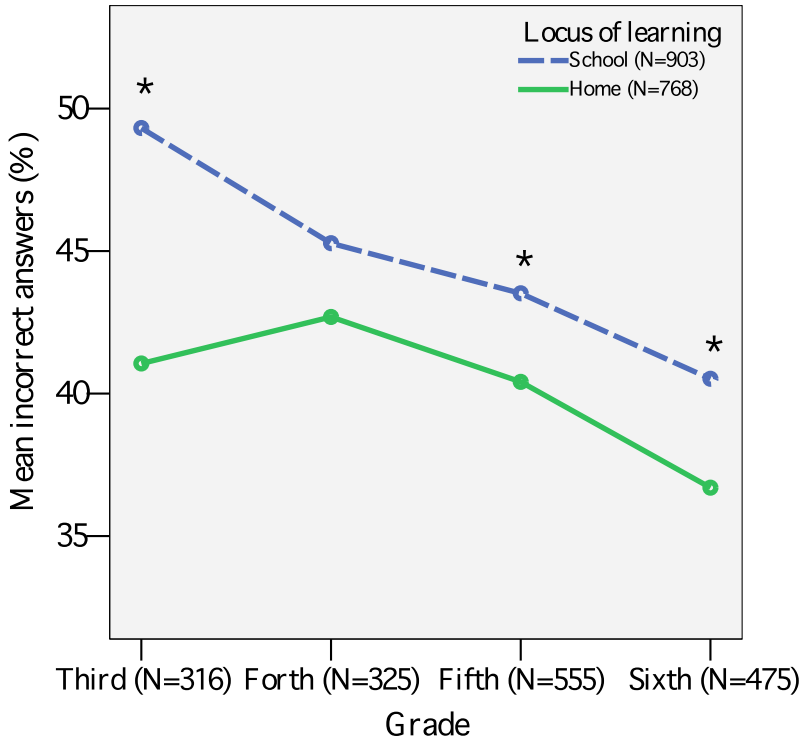
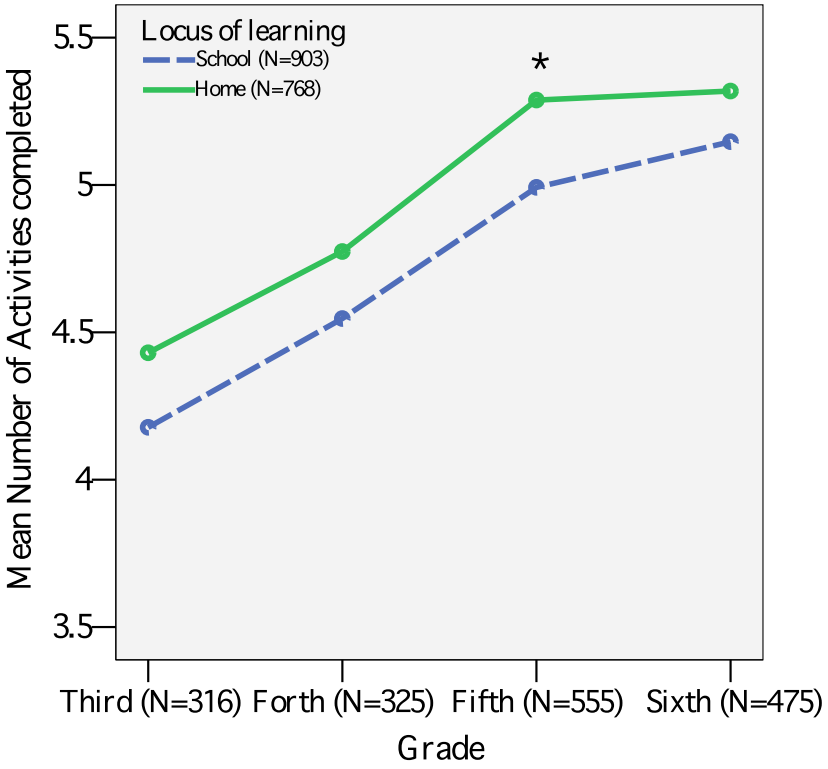


Figure 5: Number of Activities Completed at School and at Home in each Grade (Significant differences of $p < 0.05$ between locus groups are marked with an asterisk.)



Discussion and Future Work

Web-based learning environments require students to be autonomous in their learning, and even more so when they learn independently at home. Many cognitive, meta-cognitive and affective aspects of learning which are relevant to the way students learn online can be realized differantly depanding on the locus of learning and age among others. These aspects might be reflected by the hidden traces students leave in log files. A very challenging task is to reveal distintive behaviours and to infer from them on the learning processes. In this study, we have demonstrated the potential of using log file analysis for enhancing our understanding of the online learning process and also to learn about the differences in the learning behaviors — at school versus at home and across ages.

Our findings indicate that students learn autonomously at home without teacher supervision. More over, their learning behaviors are manifested differently according to the locus of learning (school vs. home) and grade (age). Students tend to spend more time (30% more) learning at home than at school, at a slower pace,

get higher scores and have less incorrect answers. It may suggest that while learning at home students are less stressful, are not constraint by time and can be more focused on task. These findings are compatible with the evidence that students appear to benefit from smaller amounts of activities at home (less than 1 hour per night) and have a positive relationship with achievement (Cooper, Robinson, & Patall, 2006).

When comparing the age factor, older students tend to spend less time learning and complete more activities. Both older and younger students tend to spend more time learning at home while younger students tend to spend more time learning than older students. In contrary to other studies who have shown that younger students have less-effective study habits and are more easily distracted (Hoover-Dempsey et al., 2001; Muhlenbruck et al., 2000), our findings indicate the opposite.

School-home gap is starting to fade away (learning wise) and a school-home learning continuum can be established. Part of the learning can be transferred to after school hours and can be evaluated similarly by means of data mining tools. Our future work will focus on gathering all these variables and employing data mining techniques in order to find learning patterns at home and at school across ages.

Educational Data Mining is an emerging research field, serving a range of educational goals within web-based educational systems, such as: evaluation of learning and effectiveness of instructional designs, development of adaptive environments for students based on their actual behaviors, provision of feedback to both students and educators, or identification of irregular learning behaviors.

References

Barker, B. O. (2000). Anytime, anyplace learning. *Forum for Applied Research and Public Policy*, *15*(1), 88–92.

Besser, H., & Bonn, M. (1997). Interactive distance-independent education: Challenges to traditional academic roles. *Journal of Education for Library and Information Science, 38*, 35–42.

Cavanaugh, C., Gillan, K. J., Kromey, J., Hess, M., & Blomeyer, R. (2004). *The effects of distance education on K–12 student outcomes: A meta-analysis.* Naperville, IL: Learning Point Associates.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research. *Review of Educational Research*, *76*, 1–62.

Gershaw, D. A. (1989). *Line on life: Locus of control.* Retrieved December 5, 2004, from http://www.azwestern.edu/psy/dgershaw/lol/ControlLocus.html.

Hoover-Dempsey, K. V., Battiato, A. C., Walker, J. M. T., Reed, R. P., De Jong, J. M., & Jones, K. P. (2001). Parental involvement in homework. *Educational Psychologist*, *36*, 195–209.

Keegan, D. (1996). *Foundations of distance education.* London: Routledge.

Linn, M. C., Clark, D. & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, *87*, 517–538.

Muhlenbruck, L., Cooper, H., Nye, B., & Lindsay, J. J. (2000). Homework and achievement: Explaining the different strengths of relation at the elementary and secondary school levels. *Social Psychology of Education*, *3*, 295–317.

Nachmias, R., & Hershkovitz, A. (2007). *Using web mining for understanding the behavior of the online learner.* The International Workshop on Applying Data Mining in e-Learning (ADML'07*)*, Crete, Greece.

Oliver, R. (2002). The role of ICT in higher education for the 21st Century: ICT as a change agent for education. *Proceedings of the Higher Education for the 21st Century Conference*. Miri, Sarawak: Curtin University.

Osborne, J., & Hennessy, S. (2003). *Literature review in science education and the role of ICT: Promise, problems and future directions.* A Report for NESTA Futurelab (No. 6). Bristol: NESTA Futurelab.

Romero, C., Ventura, S., & Garcia, E. (2008).Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, *51*(1), 368–384.

Sims, R., & Hedberg, J. (1995). Dimensions of learner control: A reappraisal for interactive multimedia instruction. In H. Maurer (Ed.), *Educational multimedia and hypermedia.* Association for the Advancement of Computing in Education, USA.

Voogt, J. (2008). IT and curriculum processes: Dilemmas and challenges. In J. M. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp.122-123). Berlin: Springer.