

Online System for Knowledge Assessment Enhances Students' Results on School Knowledge Test

Benjamin Kralj and Saša Aleksej Glažar

University of Ljubljana, Faculty of Education, Kardeljeva ploščad 16, 1000 Ljubljana, Slovenia

* Corresponding author: E-mail: benjamin.kralj@pef.uni-lj.si

Received: 25-01-2013

Abstract

Variety of online tools were built to help assessing students' performance in school. Many teachers changed their methods of assessment from paper-and-pencil (P&P) to online systems. In this study we analyse the influence that using an online system for knowledge assessment has on students' knowledge. Based on both a literature study and our own research we designed and built an online system for knowledge assessment. The system is evaluated using two groups of primary school teachers and students (N = 686) in Slovenia: an experimental and a control group. Students solved P&P exams on several occasions. The experimental group was allowed to access the system either at school or at home for a limited period during the presentation of a selected school topic. Students in the experimental group were able to solve tasks and compare their own achievements with those of their coevals. A comparison of the P&P school exams results achieved by both groups revealed a positive effect on subject topic comprehension for those with access to the online self-assessment system.

Keywords: Online assessment, computer assessment, primary school, chemistry education

1. Introduction

Information communications technology (ICT) provides users with useful tools to achieve their educational goals by creating a suitable e-learning environment. The function of such an environment is to provide the following: (1) delivery of e-learning materials; (2) recording of experience and creation of a user profile; (3) management of e-learning materials and e-courses.¹ Acknowledging these particular requirements led us to design and construct an online system for knowledge assessment that enables students to learn while solving tasks. Traditionally we can track students' knowledge by monitoring their results in written paper-and-pencil (P&P) tests,² asking questions during class, and grading assignments of student's individual work. Evolved online tools enabled various online self-assessment systems to be developed. For example Ibabe and Jauregizar³ used an online self-assessment as part of an e-learning process so that students had access to self-assessment exercises wherever and whenever they liked.³ It is also possible to monitor students' previous knowledge and their progress in subject comprehension. In addition, by analysing the test scores, potential trouble areas can be identified and individual students can exercise control over their own progress.

Researchers reported how achievement monitoring had a positive effect on students' learning.⁴

Students who use this kind of system can assess their knowledge either on their own to discover their weaknesses or with the help of their teachers recoup any academic loss. Information about students' results can also positively influence teachers' instructional approaches. For example, Beck and Davidson⁶ described how having an online "early warning system" was useful for detecting high-risk students before low grades jeopardize their college careers.⁷ It can also have a positive effect on students' opinion about their own knowledge. Students, who are unaware that certain abilities or factual and procedural knowledge is insufficient, are unlikely to make sufficient effort to acquire and construct new knowledge.³ Studies showed that students who solved self-assessment exercises improved their final grades and "In most academic settings, more exposure to course content typically translates to better overall understanding of the content".⁸ Exposure to content can also influence students' attitude towards taking school knowledge tests. When students track their own progress (using graphically displayed data or tables), their progress is associated with higher achievement.⁴ Pen et al. observed that after implementing a self-assessment system, a positive change occurred in the classroom envi-

ronment resulting in increasing students' confidence when taking knowledge tests which resulted in higher test scores.⁹ Students can advantage during the online testing if they have familiarized themselves with previous practice on the system.¹⁰

Observing students' progress is helpful for teachers. Having the appropriate data means that teachers can identify and focus their attention on those students most likely to receive a low grade point average.¹¹ It also allows them to calibrate their estimation of students' achievements. Ideally an online system should provide professionals with a tool capable of yielding more information in less testing time.¹² When an e-learning environment is incorporated in homework assignments, students are able to organize themselves into proximity groups to collectively solve complex tasks.⁵

There is no guarantee that any new system will have a positive influence on learning and there are many factors including motivation, possession of a PC, computer skills, access to the internet and its perceived usefulness, which can explain why some students use the system while others do not.³ School knowledge test results cannot be "*significantly improved only by replacing the paper-and-pencil test with Web-based test*".¹³ After analysing the results of many self-assessment strategies and innovations, Marzano concluded that anywhere from 20% to 40% of the studies in any given subject area will report negative results.

1. 1. Description of the Online System

The »TikTakTest™«ⁱ online system was built on the Apache HTTP Server¹⁵ and uses web technologies relational database management system MySQL for storing and extracting data. The MySQL stores data in tables similar to spreadsheet programs like Microsoft Excel or LibreOffice Calc. Each table stores one type of data e.g., task questions, task answers, chapter names or chapter goals. Different types of data are stored in different tables. In any table, each line represents an individual data unit specific to that table such as one task, one answer, one chapter, one chapter goal. The same data is stored on the same table and every data unit is uniquely identified so that it can be connected and cross-referenced to data in other tables. Commands and scripts are executed using the server script language PHP.¹⁷ The Hyper Text Markup Language is used to structure text and multimedia documents and create hypertext links between documents.¹⁸ Cascading Style Sheets (CSS), is a simple mechanism for applying style to web documents, to generate the final look of the online system.¹⁹

TikTakTest™ is available online, so that all users (teachers and students) can access it either from school, home or any suitable location. There are no limitations regarding when the system can be accessed or the number of tasks that the teacher can store or students solve. All that is required to access the system is a username and password. Users will receive their secure login data by e-mail. First, the system administrator registers a teacher and the login data is automatically sent out to them. Then a teacher will register their own students into the system which sends out automatically generated usernames and passwords. Once logged in, teachers and students have various options available to them; teachers are able to manipulate tasks and knowledge tests, while students' focus on solving them. The system is based on a database of tasks and student responses. The very construction of the system along with the collected data enables a user to: (1) enter tasks; (2) build knowledge tests using specific criteria; (3) solve tasks from a specific chapter or already prepared knowledge tests and (4) monitor their achievement/progress.

The structure is designed to imitate that of the school and reflects the level of education, its curriculum and its educational goals. Schools are sorted by their level of education: primary school, secondary school and higher education. In this way teachers can save tasks under a particular subject e.g. chemistry allowing teachers, in schools of the same level of education, to create, co-organize and share tasks within a single database.

Every subject is organized under subject topics based on the school curriculum. For every topic the level of achievement is measured by goals that present guidance for grading students' knowledge and tasks are linked to individual goals, which enable the system and its users to filter out tasks of a specific level of education, curriculum chapter or specific chapter goal. This allows the system to be adapted to any subject such as physics, biology or mathematics at any level of education.

1. 1. 1. Entering Tasks

A teacher when creating new tasks is guided, step-by-step by the system, to choose a topic, subtopic and the type of task. Different types of tasks can be chosen from a menu including a true/false task, task with one correct answer and a task with more than one correct answer. The system also offers the possibility to insert pictures so that the students must answer questions by recognizing visual elements. Another type of task demands a numerical answer, where students answer the question by entering their results into a

ⁱ TikTakTest™ is a registered trade mark. It is an online system designed by Benjamin Kralj teacher of chemistry and physics and assistant at the Faculty of Education University of Ljubljana, Slovenia. He presented the TikTakTest™ as part of his diploma thesis prepared while at the Faculty of Education University of Ljubljana in Slovenia under the supervision of prof. dr. Saša A. Glažar in 2008 for the purpose of improving the learning process and students' self-assessment. The TikTakTest™ was demonstrated in numerous Slovenian primary schools and is regularly used as an assessment tool for students. Because of its flexibility it is used in many non-school related projects. This research was conducted as a part of the author's doctoral studies at the Faculty of Natural Sciences and Engineering University of Ljubljana, Slovenia in The Department of Chemical Education and Informatics.

textbox. When entering a new question, it is necessary to define its taxonomic level. Four different cognitive levels are available according to Bloom's taxonomy: (i) remember, (ii) understand, (iii) apply and (iv) higher levels.²⁰

The system includes topic chapters that students are expected to comprehend at the end of the school yearⁱⁱ and within each chapter specific goals define what needs to be taught and learnt. Teachers are able to assign each task to a specific chapter and appoint a task to the appropriate goals. The connection task-goal-chapter enables students to select the chapter for which they would like to solve tasks. The system will then search the database for the appropriate tasks and assigns them to the students.

Finally, a teacher can define whether or not the task is private or open to everyone. Settings can be changed at any time. The system remembers which user entered a particular task and records the time and date. The system also represents the first step in establishing quality control, since regular overview of tasks can be made and revision is not bound to any geographical unit.

1. 1. 2. Building a Knowledge Test

Knowledge tests are constructed from the available tasks stored in the system. When creating a new test the teacher determines its title and assigns any number of tasks to it. Tasks are accessible through a search engine

embedded in the system. This allows teachers to select tasks using specific concepts, which can be found anywhere in questions, answers or in keywords attached to images. Since tasks are assigned to the goals of the curriculum and through them to chapters, a teacher can find appropriate tasks by using a chapter filter. It is possible for a knowledge test to be assigned to be solved by a specific school class, and a time limit set in which the students can access a specific test. In addition, the teacher has the option to either predetermine the order of solving the tasks or allow the system to randomly select the order of tasks each time a student chooses to solve it. Figure 1 shows the path taken in constructing a knowledge test.

1. 1. 3. Solving Tasks

The system offers students different options to solve tasks stored in its database. There are four ways for students to start solving tasks.

- (1) Students can solve tasks that their teacher has set in prepared knowledge tests. Each time a student chooses to complete a knowledge test the same tasks are offered to be solved either in a predetermined or random order. Teachers are able to save knowledge tests and mark them as accessible to all students. The students can then choose which ones to solve.

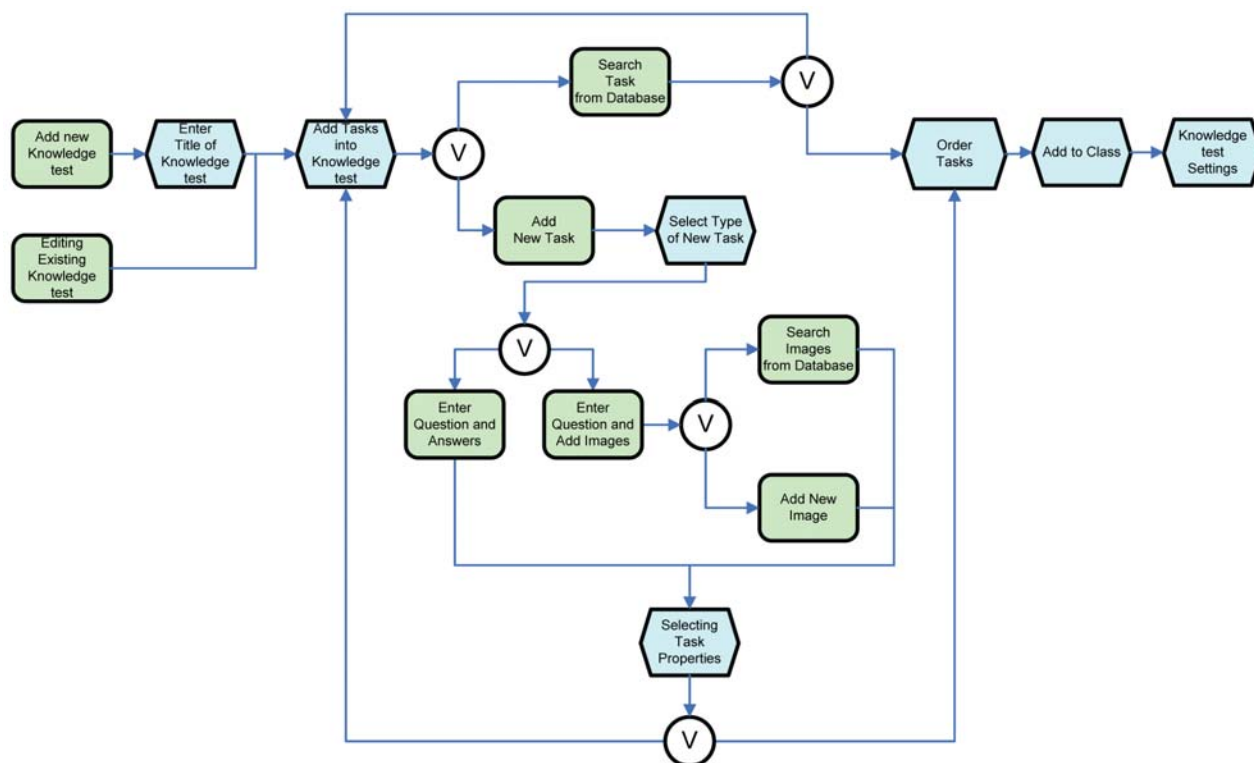


Figure 1. Building a Knowledge Test

ⁱⁱ School year in Slovenia starts on September 1st and lasts for 10 months until June 24th.

- (2) Students can choose to solve tasks by selecting a particular chapter from the curriculum. The system will filter tasks from its database according to the student's selection. From this group, the system will randomly select tasks for the students to solve. Tasks in the database are sorted according to cognitive level. The system chooses tasks according to the highest level a student has attained during task solving. Initially, the system offers random tasks from the first cognitive level (remember). For a student to reach the next cognitive level (understand) they must answer correctly a certain percentage of tasks from their highest achieved cognitive level. It is not a decision tree as Murthy describes, but it is determined by the percentage of correct answers from the highest reached cognitive level.²¹
- (3) The student has the option to select tasks set by their teacher. The system filters tasks that their teacher has stored in the database. The system will then randomly select from this group specific tasks for the student to solve. Again, the highest cognitive level constrains the selection.
- (4) A student can select 'new tasks only'. The system filters tasks that were recently saved into the system and randomly selects few of them for the student to solve. Again, student's achieved cognitive level is the criteria for task selection. The system presents each question to the student individually, so that an answer is required before a student can move on to the next. In the multiple choice option, possible answers are displayed randomly, next to a letter of the alphabet i.e., A, B, C and so on. The system enables the random display of answers which makes memorising a letter before a correct answer irrelevant. When all the questions have been answered and the student has reviewed their selections they can then submit their answers for evaluation.

1. 1. 4. Achievements

During task solving student's answers are stored in the database. Every answer is automatically checked and evaluated. The system calculates the number of correct and incorrect answers and presents the result to the student as an absolute number and as a percentage of correct answers. This data is also presented graphically so that students can compare their achievements to their past results, to their classmates or their coevals from other schools. When solving a knowledge test, the best results from all the students involved are also presented graphically, on which the student's personal achievement is marked for comparison. All data are kept anonymous and cannot be linked to a specific user. Teachers have the option to compare students' achievements on a selected knowledge test or chapter marked with a student's name. This allows teachers to monitor an individual's results and achievements.

1. 2. The Purpose of the Research

In Slovenian primary schools chemistry is part of the 8th and 9th grade (age 12 to 15) curriculum. Learning topics are grouped as follows: *Substances and its Changes, Substances and its Properties, Pure Substances and Compounds, Substance Construction, Chemical Reactions, Atom and Periodic Table, Particle Connections, Electrolytes, Quantity Relationships, Elements in Periodic Table, Family of Carbohydrates, Family of Oxygen Organic Compounds, Family of Nitrogen Organic Compounds, and Polymers*. The topic from organic chemistry 'Oxygen Organic Compounds' was chosen.

The aim of this study is to determine the influence that an online system for knowledge assessment has on 9th grade (age 13 to 15) chemistry students' learning process by measuring the influence that a self-assessment system has on knowledge improvement, knowledge comprehension and knowledge sustainability. TikTakTest™ is a useful tool for this since it can be easily implemented into the existing learning process.

1. 3. Research Question

Research questions were formed into two hypotheses:

- H1:** Students that use an online system for knowledge assessment for learning chemistry achieve better results than students who do not use the system.
- H2:** Using an online knowledge assessment system influences the students' knowledge sustainability.

2. Method

2. 1. Participants

For the purpose of this research, we chose primary school chemistry course (ages 13 to 15) since the principle author is a doctoral student of chemical education. The research took place during the 2010/2011 school year. An invitation was sent to 178 primary school head teachers and chemistry teachers in the Republic of Slovenia. Twenty chemistry teachers (4.43 % out of 451) agreed to participate. Since the research topic 'Oxygen Organic Compounds' is included in the year nine curriculum, 9th grade chemistry students aged 13 to 15 ($M = 13.92$; $SD = 0.352$) were selected as test subjects. In total, 686 students (3.8 % out of 17.854 in the country) including 357 females (52 %) and 329 males (48 %) took part. Parental consent was obtained.

2. 2. Instruments

We designed several measurement instruments, paper-and-pencil (P&P) tests, named the "Pre-Test", "Test" and "Post-Test". Initially, we used the Pre-Test results to place students into one of two groups: experimental or

control group. After the experiment the students were then asked to solve the Test and then after one month the Post-Test. The tasks were the same in the Test and Post-Test. Figure 2 shows a timeline of the research.

At the beginning of each P&P test, the students wrote their unique code on the paper for tracking purposes. Students were allowed to use a calculator and a copy of the periodic table was provided. Question topics were chosen from the primary school curriculum for chemistry. The students were then given 30 minutes to complete the P&P test.

All three P&P tests included 20 multiple-choice questions. Most of the tasks were text only, and students were instructed to mark the correct answer from the four possible answers. Certain tasks included images, and students were instructed to recognize elements and choose the correct answer.

students to carry out or use a procedure i.e., apply (tasks: 2, 10 and 17).

2. 3. Research Design

After agreeing to participate and the parents had given their consent, the teachers received a schedule of chemistry topics to be taught to the students over 10 school hours i.e., 5 weeks in those schools that have 2 hours of chemistry per week. Both experimental and control groups then had 10 school hours to elaborate on the prescribed chapter and its content. Only the students and teachers in the experimental research group were allowed to use the online system. In order not to interrupt normal school processes, specific instructions were given neither to the teachers nor students about when or where to use the system. Nevertheless, a help service

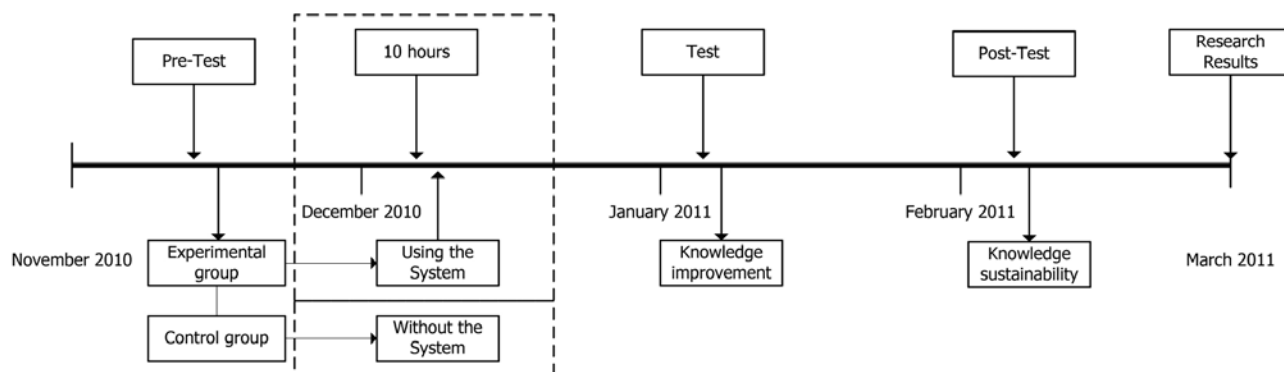


Figure 2. Time plan of the research

2. 2. 1. Pre-Test

In total, 669 students solved the Pre-Test. The purpose of the Pre-Test was to determine the level of students' knowledge and to place students between the experimental and control groups so that students in each group had a comparable level of knowledge.

2. 2. 2. Test and Post-Test

The Test results were used to measure the effect of system usage on the students' knowledge and the Post-Test results to determine the system's influence on knowledge sustainability. In total 636 students solved the Test and 624 solved the Post-Test. Because of a printing error one question was eliminated from the tests but since its deletion had no effect on the quality of the research the results of all 19 tasks were analysed.

According to Bloom's taxonomy nine tasks require students to retrieve relevant knowledge from their long-term memory i.e., remember (tasks 1, 5, 6, 8, 11, 12, 13, 16 and 19), seven tasks required students to construct meaning from instructional messages i.e., understanding (tasks: 3, 4, 7, 14, 15, 18 and 20) and three tasks required

contactable either *via* e-mail or by telephone was made available.

When all the students had discussed the same topic, both groups completed the Test. A comparative analysis was then made to identify any differences between the achievements between the two groups. After one month the students were asked to complete the Post-Test containing the same tasks as in the Test.

Because the study was designed to be nonintrusive there was no control over whether or not the teacher provided the students with the entire requested information on a given subject. All users were free to choose how, where, when and in which manner to use the system. However every time they logged in, the system registered their work. This allowed the teachers to monitor their students' achievements and the amount of work done (solved tasks and online tests).

For this research, specific tasks were designed and saved into the system's database. There was no explicit knowledge test built into the system nor were there any other tools required. The system randomly searched the database for different tasks depending on the selected menu options. Since there were many tasks entered into the sys-

tem from teachers from all over the country it was impossible to predict how many times the system would select a task to be solved.

3. Results

The number of students who solved the Pre-Test, Test and Post-Test were 669, 636 and 624, respectively. Number variants because not all students were always present at the time when P&P tests took place. In the experimental group there were 453 students: 235 females (51.9 %) and 218 males (48.1 %) with a mean age of 13.90 years (SD = 0.346) while the control group comprised of 233 students: 122 females (52.4 %) and 111 males (47.6 %) with a mean age of 13.95 years (SD = 0.362). We used SPSS to analyse our data.

A Cronbach alpha coefficient was used to measure internal consistency of the Pre-Test, Test and Post-Test. The calculated Cronbach alpha for the Pre-Test was 0.655. Since 0.7 is reported as an acceptable value,²² three Pre-Test tasks (numbered: 4., 6. and 9.) were removed from any further analysis. The new Cronbach alpha was 0.700. For the Test and Post-Test, the Cronbach alpha was 0.705 and 0.734, respectively. According to these results the scale is reliable.

An independent-samples t-test was then applied to evaluate the *control and experimental group* results for the Pre-Test, Test and Post-Test results. There was no significant difference in students' achievements in the Pre-Test between the experimental (M = 11.11; SD = 3.24) and control group results (M = 11.12; SD = 3.01; $t(667) = -0.049$; $p = 0.96$). Results reveal a significant difference between the experimental (M = 12.27; SD = 3.51) and control group's Test results (M = 11.47; SD = 3.13; $t(483.373) = 2.945$; $p = 0.003$), and no significant difference for the Post-Test between the experimental (M = 12.07; SD = 3.586) and the control group results (M = 11.68; SD = 3.420; $t(622) = 1.289$; $p = 0.198$). The results are presented in Table 1.

A paired-samples t-test was then used to compare the results obtained by the *experimental group* for the Test and Post-Test. There was no statistical significant decrease in the results between the Test (M = 0.644; SD = 0.183) and Post-Test (M = 0.640; SD = 0.185; $t(399) = 0.421$; $p = 0.674$) scores. The paired-samples t-test was also used to evaluate the results obtained by the *control group* for the Test and Post-Test. The results reveal no statistical significant increase between the Test (M = 0.604; SD = 0.164) and Post-Test (M = 0.615; SD = 0.181; $t(184) = -0.855$; $p = 0.394$) scores. Described results are presented in Table 2.

3.1. Test Scores

The appendix contains the full table of questions and answers for all 19 tasks. Table 3 gives a typical example of a task. After each question the answers are given with the percentage of how many students chose that specific answer in the experimental (E) and in the control (C) group. The difference between the groups is also calculated (D) together with the Bloom taxonomy level (B). A number one (1) denotes the correct answer.

In 14 out of the 19ⁱⁱⁱ tasks the experimental group achieved better results than the control group (Figure 3 positive bar). In six out of the 14 tasks the t-test revealed a statistically important difference in the students' achievements (tasks 2, 5, 11, 12, 15 and 16). According to Bloom's taxonomy, tasks 5, 11, 12 and 16 require students to remember information, task 15 requires understanding and task 2 requires students to apply their knowledge.

There were five tasks (Figure 3: negative bar) where the students in the control group achieved better results than in the experimental group. Task 20 was designed to measure the students' understanding of the topic. The t-test revealed a significant difference between the two groups in terms of students' achievement (see Appendix).

ⁱⁱⁱ Task 9 was excluded from analysis due to print error so instead of 20, only 19 tasks were used for analysis.

Table 1. Results of independent-sample t-tests for experimental and control group students' Pre-Test, Test and Post-Test results.

	Experimental group			Control group			dt	t	p
	N	M	SD	N	M	SD			
Pre-Test	445	11.11	3.244	224	11.12	3.011	677	-0.049	0.961
Test	419	12.27	3.512	217	11.47	3.130	483.373	2.945	0.003
Post-Test	424	12.07	3.586	200	11.68	3.420	622	1.289	0.198

Table 2. Results of paired-samples t-test for experimental and control group students.

	N	Test		N	Post-Test		dt	t	p
		M	SD		M	SD			
Experimental group	400	0.644	0.183	400	0.640	0.185	399	0.421	0.674
Control group	185	0.604	0.164	185	0.615	0.181	184	-0.855	0.394

Table 3. Tasks of Test and Post-Test with answers selection percentage of students form Experimental (E) and Control (C) group.

Tasks	E	C	D	B
1. Evaluate the statement: An alcohol that has a longer chain of carbon atoms also has a higher boiling point.				
1 A) Yes	88%	84%	4%	Remember
B) No	11%	13%	-2%	
no answer	1%	3%	-2%	

Legend: E: Experimental group selection percentage (N = 419) C: Control group selection percentage (N = 217) D: Difference in percentage between the Experimental and Control group (E – C) B: Bloom cognitive level

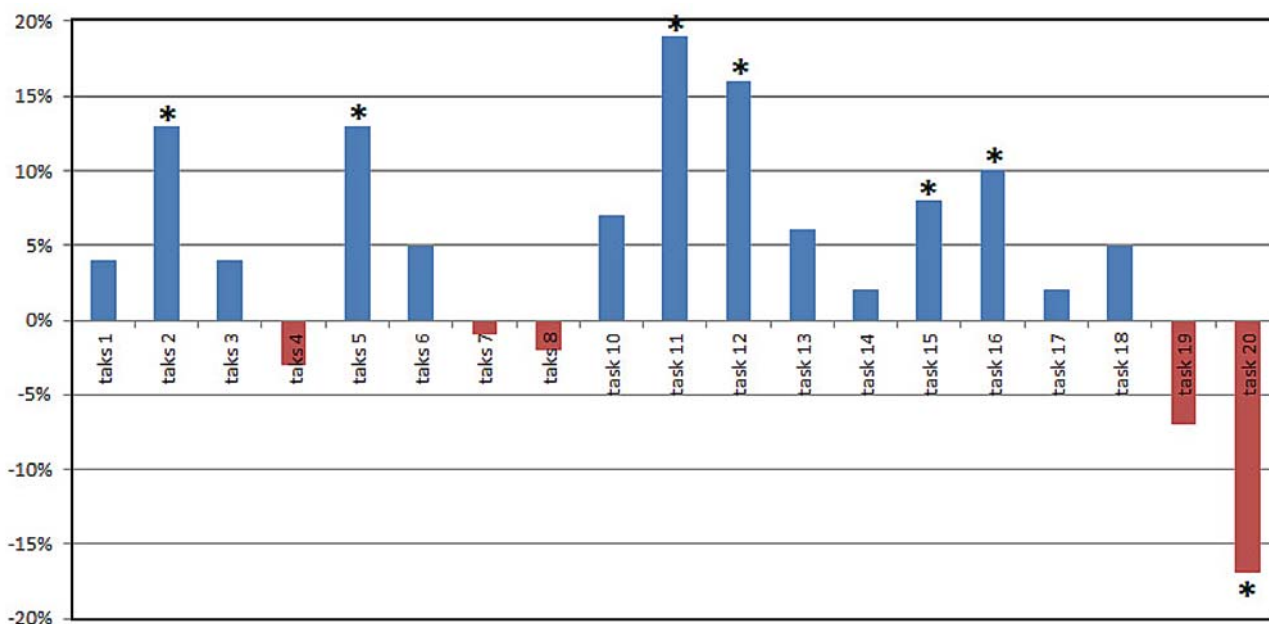


Figure 3. Difference in Test achievement between Control and Experimental group students. Statistically important differences are marked with asterisk (*).

4. Discussion

Analysis of the Test results revealed that the experimental group achieved significantly better results than the control group and confirms our initial hypothesis (H1). It is important to note that specific instructions on how to use the system were given to neither teachers nor students. Since only the teachers of the experimental group and their students had access to the system, and there was no influence on the daily school routine, we can argue that the system does influence students' knowledge comprehension with minimal intrusion. Furthermore, the system can act as a valuable tool for knowledge assessment without the need for major adaptations to established classroom processes. Our findings agree with other researches looking at other age groups.^{3,11,4,9}

Time is a major factor in how much of learnt material is retained i.e., as time passes the amount of information originally learnt diminishes.⁷ Creating a system that reduces this effect would be a significant step towards quality knowledge sustainability. Analysis of the Post-Test results reveals a positive difference in favour of the

experimental group albeit the difference is not statistically significant. And further analysis of the data is required before we can either confirm or reject hypothesis (H2), since there are many factors that can influence a negative results.^{3,14}

A comparison was made inside each group (control and experimental) to evaluate the effects on students' achievements for the Test and Post-Test using a paired-sample t-test. A slight but statistically insignificant reduction in the experimental group's results between the Test (64.4%) and Post-Test (64.0%) was observed. The same comparison was made for the control group. Analysis reveals that the results for the Test (60.5%) improved in the Post-Test (61.5%), but the difference is not statistically significant.

Students in the control group had lower Test results (60.5%) than students in the experimental group (64.4%). The same is also true for the Post-Test where the experimental and control group achieved 64.0% and 61.5%, respectively.

In 14 out of 19 Test tasks the experimental group achieved better average results than the control group. In 10

out of the 14 tasks the experimental group students achieved scores >5 % higher than in the control group. We also observed the largest difference for task 11 where there was a 19% improvement over the control group. There were 5 tasks at which the control group's performance was better than experimental group. Task 20 revealed the biggest difference with results on average 17% higher than the experimental group's results. This task asked students to compare rational formulas of two compounds and determine which of the four possible statements was correct (see Appendix).

With regards to Bloom's taxonomy we find statistically important differences for certain Test tasks. At the first level (remember) in 4 out of 9 tasks the experimental group results were significantly higher, at the second level (understand) it was 1 out of 7 tasks, and for the third level (apply) 1 out of 3.

It can be argued that the system provides students with a tool to help them measure their knowledge at all taxonomic levels. However because there exist different teaching methods, different approaches to learning, numerous tasks, various possibilities for solving tasks (some students may have only solved the first level tasks) and that for task 20 the control group achieved statistically better results, one must be cautious about stating that the system has a positive influence over all fields.

The next step is to investigate how using such a system can affect teaching, learning and knowledge assessment. What role does the students' motivation has? Is it possible to use the system to detect the influence of teachers' motivation? Are there any measurable connections between teachers' motivation and motivation of their students? Could the system be used to provide teachers with valuable information regarding the quality of their tasks?

It must be stressed that the experimental group used the system with tasks not only prepared by their own teachers but they also had the option to solve random tasks. Acknowledging this, the content i.e., topics, themes and images, and qualities such as discrimination, difficulty, and the time required to solve these tasks require further analysis to explain the observed results. After analysing fully the system in its present form it is the intention of the authors to upgrade the system and perform further research.

Teachers are enabled to use the system to conduct their own researches. Continuous data flow from their students' responses can help them regularly self-evaluate their teaching process. Future upgrades of the system will provide teachers with extended variety of use and research options since there are many opportunities to collect data in school – for example students' individual work or in groups, homework or excursion evaluation, and even students' discussion over specific tasks especially those that appear harder to solve. Technical research will explore new ways to enable users many different task types and offer students more accurate tool for measuring their subject comprehension progress.

5. References

1. C.-P. Chu, Y.-C. Chang and C.-C. Tsai, PC2PSO: personalized e-course composition based on Particle Swarm Optimization, <http://www.springerlink.com/content/361wk01346w34gx8/?p=a5d09996873c4ebcbc3871c99da303e2&pi=85>, (assessed: 23.2.2010)
2. A. J. Arce-Ferrer and E. Martínez Guzmán, *Educational and Psychological Measurement* **2009**, *69*, 855–867.
3. I. Ibabe and J. Jauregizar, *Higher Education: The International Journal of Higher Education and Educational Planning* **2010**, *59*, 243–258.
4. R. J. Marzano, *Educational Leadership* **2010**, *67*, 86–87.
5. L. Fisher and T. Holme, *The Chemical Educator* **2000**, *5*, 269–276.
6. H. P. Beck and W. D. Davidson, *Research in Higher Education* **2001**, *42*, 709–723.
7. C. H. Grenwelge, *Journal of Psychoeducational Assessment* **2009**, *27*, 345–350.
8. B. K. McFarlin and M. D. Jackson, *The Diabetes Educator* **2008**, *34*, 766–775.
9. J. H. Penn, V. M. Nedeff and G. Gozdzik, *J. Chem. Educ.* **2000**, *77*, 227–231.
10. O. L. Liu, H.-S. Lee and M. C. Linn, *Journal of Research in Science Teaching* **2011**, *48*, 1079–1107.
11. D. Kennepohl, M. Guay and V. Thomas, Using an Online, Self-Diagnostic Test for Introductory General Chemistry at an Open University, <http://pubs.acs.org/doi/abs/10.1021/ed900031p>, (assessed: 16.9.2010)
12. H.-s. Lin, S. T. Lee and D. Treagust, *J. Chem. Educ.* **2005**, *82*, 1565–1569.
13. T.-H. Wang, *Computers & Education* **2008**, *51*, 1247–1263.
14. R. J. Marzano, *Phi Delta Kappan* **2009**, *91*, 30–37.
15. Apache, The Apache HTTP Server Project, <http://httpd.apache.org/>, (assessed: 30.3.2010)
16. MySQL, The world's most popular open source database, <http://www.mysql.com/>, (assessed: 31.3.2010)
17. PHP, Hypertext Preprocessor, <http://www.php.net/>, (assessed: 31.3.2010)
18. HTML, The American Heritage® Science Dictionary, <http://dictionary.reference.com/browse/html>, (assessed: 22. 2. 2012)
19. W3C, What is CSS?, <http://www.w3.org/Style/CSS/>, (assessed: 22. 2. 2012)
20. Bloom's Taxonomy Revised: A Taxonomy for Learning, Teaching, and Assessing, <http://www.transitionmathproject.org/partners/wcp/wcp.asp>, (assessed: 17. 8. 2012)
21. S. K. Murthy, *Data Mining and Knowledge Discovery* **1998**, *2*, 345–389.
22. J. Pallant, in: *SPSS Survival Manual in the Allen & Unwin*, **2005**,–318.

Povzetek

Merjenje šolskih učnih dosežkov dosegamo z uporabo različnih spletnih orodij. Učitelji v vse večji meri prilagajajo načine vrednotenja znanja ter s papirne oblike prehajajo na uporabo različnih spletnih sistemov. V tej raziskavi ugotavljamo učinke, ki jih ima uporaba spletnega sistema za vrednotenje znanja, na znanje učencev. Na podlagi literature in lastnih opažanj smo oblikovali spletni sistem za vrednotenje znanja. Učinke delovanja sistema smo vrednotili s pomočjo dveh skupin učiteljev in učencev (N = 686) iz Slovenije: eksperimentalno in kontrolno. Učenci so v času raziskave ob različnih priložnostih v šoli na tradicionalen način (s papirjem in svinčnikom) reševali preizkuse znanja. Med obravnavo izbrane vsebine je le eksperimentalna skupina uporabljala spletni sistem, kjer so lahko reševali naloge ter primerjali svoje dosežke z dosežki vrstnikov. Primerjava rezultatov preverjanja znanja s papirjem in svinčnikom je pri učencih, ki so imeli dostop do spletnega sistema za samostojno preverjanje znanja, pokazala pozitivne spremembe pri poznavanju vsebin.

Supplementary Information

Online System for Knowledge Assessment Enhances Students' Results on School Knowledge Test

Benjamin Kralj and Saša Aleksej Glažar

The table of questions and answers are given for all 19 tasks that were used to assess students' knowledge and an example is provided in Table 4. After each question answers are given with percentage, how many students choose specific answer in experimental (E) or in the con-

trol (C) group. The subtraction of both percentages for each answer is calculated (D) and the Bloom taxonomy level (B) stated. A one (1) in front of the answer marks the correct answer.

Table 4. Tasks of Test and Post-Test with answers selection percentage of students from experimental (E) and control group (C).

Tasks	E	C	D	B
1. Evaluate the statement: An alcohol that has a longer chain of carbon atoms also has a higher boiling point.				
1 A) Yes	88%	84%	4%	Remember
B) No	11%	13%	-2%	
no answer	1%	3%	-2%	
2. The formula of a compound consisting of four carbon atoms connected in a chain with a hydroxyl group that is not at the end of the chain was wrongly named by John as butan-3-ol. What is the correct name of the compound?				
A) Butan-1-ol	7%	6%	1%	Apply
1 B) Butan-2-ol	78%	65%	13%	
C) Butan-4-ol	7%	11%	-4%	
D) Butanol	8%	17%	-9%	
no answer	1%	1%	0%	
3. Butan-1-ol has boiling point of 118 °C and hexan-1-ol has boiling point at 156 °C. Evaluate the boiling point of pentan-1-ol.				
A) 95 °C	2%	2%	0%	Understand
B) 104 °C	3%	3%	0%	
1 C) 138 °C	87%	83%	4%	
D) 187 °C	9%	12%	-3%	
no answer	0%	0%	0%	
4. Select the alcohol which has the lowest solubility in water.				
A) Butanol	5%	8%	-3%	Understand
B) Hexanol	5%	7%	-2%	
1 C) Octanol	72%	75%	-3%	
D) Propanol	15%	8%	7%	
no answer	2%	2%	0%	
5. Select the general formula of alcohols.				
A) C_nH_{2n+2}	9%	14%	-5%	Remember
1 B) $C_nH_{2n+2}O$	49%	36%	13%	
C) $C_nH_{2n+1}O$	21%	24%	-3%	
D) $C_nH_{2n}O$	20%	24%	-4%	
no answer	1%	1%	0%	
6. Name the compound with the rational formula.				
	$CH_3-CH_2-CH_2-CH_2-\overset{OH}{\underset{ }{CH}}-CH_2-CH_3$			
A) Heptane	2%	1%	1%	Remember
1 B) Heptan-3-ol	84%	79%	5%	
C) Heptan-5-ol	11%	18%	-7%	
D) Heptanol	3%	2%	1%	
no answer	1%	0%	1%	

7. Evaluate the statement: Hexan-3-ol is a tertiary alcohol.					
	A) Yes	32%	31%	1%	Understand
1	B) No	67%	68%	-1%	
	no answer	0%	1%	-1%	
8. During alcoholic fermentation we can produce a solution with up to 15 % of ethanol. Why can't we get a higher percentage of ethanol using this method?					
	A) We have to warm up the solution, because yeast needs more energy to produce a solution with greater than 15 % of ethanol.	9%	5%	4%	Remember
	B) Because alcohol evaporates from the solution at a higher concentration.	16%	13%	3%	
1	C) Because yeast die at a higher ethanol concentration.	45%	47%	-2%	
	D) Because the ratio of sugar and water always results in a 15 % alcohol solution.	26%	32%	-6%	
	no answer	4%	2%	2%	
10. Which is the correct sequence of compounds during the gradual oxidation of propanol to carbon dioxide and water?					
	A) propanol → carbon dioxide and water	11%	11%	0%	Apply
	B) propanol → propene → carbon dioxide and water	8%	8%	0%	
1	C) propanol → propanal → propanoic acid → carbon dioxide and water	73%	66%	7%	
	D) propanol → propanone → propanoic acid → carbon dioxide and water	7%	13%	-6%	
	no answer	2%	1%	1%	
11. At the oxidation of which alcohol we get aldehyde?					
1	A) hexan-1-ol	58%	39%	19%	Remember
	B) hexan-2-ol	16%	26%	-10%	
	C) hexan-3-ol	9%	16%	-7%	
	D) 2-methylhexan-2-ol	15%	15%	0%	
	no answer	2%	4%	-2%	
12. How do we name a homologous series with the formula R-CHO?					
1	A) Aldehyde	79%	63%	16%	Remember
	B) Alkane	8%	7%	1%	
	C) Alcohol	5%	10%	-5%	
	D) Phenol	8%	17%	-9%	
	no answer	0%	3%	-3%	
13. Which rational formula represents the molecule of hexan-2-one?					
	A) $\text{CH}_3\text{-CH}_2\text{-CH}_2\overset{\text{O}}{\parallel}\text{C-CH}_3$	3%	4%	-1%	Remember
1	B) $\text{CH}_3\text{-CH}_2\text{-CH}_2\text{-CH}_2\overset{\text{O}}{\parallel}\text{C-CH}_3$	81%	75%	6%	
	C) $\text{CH}_3\text{-CH}_2\text{-CH}_2\text{-CH}_2\text{-CH}_2\overset{\text{O}}{\parallel}\text{C-CH}_3$	6%	9%	-3%	
	D) $\text{CH}_3\text{-CH}_2\text{-CH}_2\text{-CH}_2\overset{\text{OH}}{\text{CH}}\text{-CH}_3$	10%	12%	-2%	
	no answer	0%	0%	0%	
14. From the gradual oxidation of which alcohols do we obtain ketones?					
	A) from primary alcohols	10%	10%	0%	Understand
1	B) from secondary alcohols	69%	67%	2%	
	C) from tertiary alcohols	14%	15%	-1%	
	D) from quaternary alcohols	6%	6%	0%	
	no answer	1%	2%	-1%	
15. Which of the compounds below can we classify as an organic acid?					
	A) $\text{CH}_3\text{-CO-CH}_3$	4%	5%	-1%	Understand
	B) H_2SO_4	17%	20%	-3%	
	C) $\text{CH}_3\text{-CH}_2\text{-CH}_2\text{-CHO}$	6%	8%	-2%	
1	D) $\text{CH}_3\text{-CH}_2\text{-CH}_2\text{-CH}_2\text{-COOH}$	73%	65%	8%	
	no answer	1%	2%	-1%	

16. Select the lethal concentration of ethanol in blood.					
	A) 0,02 ‰	3%	8%	-5%	Remember
	B) 0,5 ‰	16%	8%	8%	
1	C) 2,0 ‰	23%	13%	10%	
	D) 5,0 ‰	58%	70%	-12%	
	no answer	1%	1%	0%	
17. Select the product of the reaction between propanoic acid and ethanol in the presence of H ₂ SO ₄ ?					
	A) CH ₃ -COO-CH ₂ -CH ₂ -CH ₃	9%	8%	1%	Apply
1	B) CH ₃ -CH ₂ -COO-CH ₂ -CH ₃	63%	61%	2%	
	C) CH ₃ -CH ₂ -CO-CH ₂ -CH ₃	14%	16%	-2%	
	D) CH ₃ -COO-CH ₂ -CH ₂ -CH ₃	11%	10%	1%	
	no answer	3%	5%	-2%	
18. What happens when sodium hydroxide is added to a solution of ethanoic (acetic) acid?					
	A) ethanol and a sodium acid forms	11%	10%	1%	Understand
	B) sodium hydroxide combusts and ethanoic (acetic) acid is converted into carbon dioxide and water	15%	16%	-1%	
	C) nothing happens because the reaction does not occur	6%	11%	-5%	
1	D) sodium ethanoate (acetate) and water forms	67%	62%	5%	
	no answer	2%	1%	1%	
19. Name the compound with that has the formula CH ₃ -CH ₂ -CH ₂ -CH ₂ -COO-CH ₂ -CH ₃ ?					
	A) Heptanoate	21%	23%	-2%	Remember
	B) Heptan-3-one	27%	22%	5%	
1	C) Ethyl pentanoate	40%	47%	-7%	
	D) Propyl butanoate	10%	7%	3%	
	no answer	1%	1%	0%	
20. Select the correct statement for the compounds that have the formula CH ₃ -CH ₂ -CH ₂ -CH ₂ -O-CH ₃ and CH ₃ -CH ₂ -O-CH ₂ -CH ₂ -CH ₃ ?					
	A) the compounds are products of the reaction between an acid and a base	14%	10%	4%	Understand
	B) the compounds are functional isomers	25%	24%	1%	
	C) the compounds are ketones	24%	14%	10%	
1	D) the compounds are structural isomers	34%	51%	-17%	
	no answer	3%	1%	2%	

Legend: E: Experimental group selection percentage (N = 419) C: Control group selection percentage (N = 217) D: Difference in percentage between the Experimental and Control group (E - C) B: Bloom cognitive level

An independent-samples t-test was conducted to evaluate the experimental group students' and control group students' results obtained in 19 tasks. Results and significance is reported in Table (below). In 7 out of 19

tasks, the difference in achievements between the control and experimental groups are statistically important. In the Table (below) the tasks are marked with asterisk (*).

Table 5. Independent-sample t-test results for Test tasks with task classification according to Bloom's taxonomic levels.

Task number	Bloom's Taxonomy Revised	Experimental group (N = 419)	Control group (N = 217)	T-test results
1	Remember Recalling (retrieving)	M = 0,881; SD = 0,325	M = 0,843; SD = 0,364	t(395,5) = 1,27; p = 0,204
2	Apply Executing (carrying out)	M = 0,776; SD = 0,418	M = 0,654; SD = 0,477	t(390,0) = 3,17; p = 0,002 *
3	Understand Inferring (predicting)	M = 0,871; SD = 0,335	M = 0,829; SD = 0,377	t(395,2) = 1,37; p = 0,171
4	Understand Classifying (categorizing)	M = 0,721; SD = 0,449	M = 0,747; SD = 0,436	t(448,7) = -0,70; p = 0,484
5	Remember Recalling (retrieving)	M = 0,487; SD = 0,500	M = 0,364; SD = 0,482	t(451,5) = 3,01; p = 0,003 *

Task number	Bloom's Taxonomy Revised	Experimental group (N = 419)	Control group (N = 217)	T-test results
6	Remember Recognizing (identifying)	M = 0,835; SD = 0,371	M = 0,788; SD = 0,410	t(401,4) = 1,42; p = 0,155
7	Understand Classifying (categorizing)	M = 0,673; SD = 0,470	M = 0,677; SD = 0,469	t(438,0) = -0,11; p = 0,911
8	Remember Recalling (retrieving)	M = 0,451; SD = 0,498	M = 0,470; SD = 0,500	t(435,5) = -0,45; p = 0,650
10	Apply Executing (carrying out)	M = 0,726; SD = 0,447	M = 0,664; SD = 0,474	t(415,4) = 1,59; p = 0,112
11	Remember Recognizing (identifying)	M = 0,578; SD = 0,495	M = 0,392; SD = 0,489	t(441,2) = 4,53; p = 0,000 *
12	Remember Recognizing (identifying)	M = 0,790; SD = 0,408	M = 0,627; SD = 0,485	t(377,2) = 4,24; p = 0,000 *
13	Remember Recognizing (identifying)	M = 0,807; SD = 0,395	M = 0,751; SD = 0,433	t(403,6) = 1,58; p = 0,115
14	Understand Classifying (categorizing)	M = 0,690; SD = 0,463	M = 0,668; SD = 0,472	t(429,9) = 0,55; p = 0,583
15	Understand Classifying (categorizing)	M = 0,728; SD = 0,446	M = 0,650; SD = 0,478	t(411,0) = 2,00; p = 0,046 *
16	Remember Recalling (retrieving)	M = 0,227; SD = 0,419	M = 0,134; SD = 0,341	t(521,3) = 3,01; p = 0,003 *
17	Apply Executing (carrying out)	M = 0,625; SD = 0,485	M = 0,613; SD = 0,488	t(434,3) = 0,30; p = 0,761
18	Understand Interpreting (clarifying)	M = 0,666; SD = 0,472	M = 0,622; SD = 0,486	t(426,3) = 1,09; p = 0,278
19	Remember Recognizing (identifying)	M = 0,399; SD = 0,490	M = 0,465; SD = 0,500	t(429,6) = -1,61; p = 0,108
20	Understand Classifying (categorizing)	M = 0,341; SD = 0,475	M = 0,507; SD = 0,501	t(416,9) = -4,02; p = 0,000 *