Undergraduate Students Assessment on Materials Chemistry Topic using an Auto-Calibrated Online System

Lorentz JÄNTSCHI[1], Carmen Elena STOENOIU[1], Sorana Daniela BOLBOACĂ[2]
[1]Technical University of Cluj-Napoca, 15 Constantin Daicoviciu Street, 400020 Cluj-Napoca, Romania, E-mails: lori@academicdirect.org, carmen@webmail.academicdirect.ro
[2]"Iuliu Hatieganu" University of Medicine and Pharmacy, 13 Emil Isac Street, 400023 Cluj-Napoca, Romania, E-mail: sbolboaca@umfcluj.ro

ABSTRACT - Development of communication and information technology has an important impact on teaching, learning and knowledge assessment. According with national trends in objective evaluation of undergraduate students' knowledge, starting with experiences obtained by creation of the multiple choices examination system for general chemistry, an auto-calibrated online evaluation system was developed. The aim of the research was to assess the first year undergraduate students' knowledge on materials chemistry topic at the Faculty of Materials Science and Engineering, the Technical University of Cluj-Napoca, Romania by the use of the developed auto-calibrated online evaluation system. The testing and evaluation methodologies are presented. The students' performances in terms of individual testing points and the average time needed to give a correct answer were collected and analyzed. The proposed auto-calibrated online evaluation system proved to offer a stable and valid evaluation environment of undergraduate students' knowledge assessment on materials chemistry topic.

KEYWORDS - Computer Assisted Evaluation; Undergraduate Students; Knowledge Evaluation; Materials Chemistry Topic

INTRODUCTION

Development of communication and information technology had an important impact on teaching, learning and assessment at all levels of formal education: kindergarten [1], primary education [2], secondary education [3,4], post secondary education [5]. The teaching profession is seen in the new technology era as *evolving from an emphasis on teacher-centered, lecture-base instruction to student-centered, interactive learning environment* [6]. According with the development of information and communication technology, the educational system all over the world is confronting on one hand with challenge of transforming the curriculum and educational process to provide students with the skills needed in order to function effectively in the actual dynamic, information-rich, and continuously changing environment and on other hand with development of objective, fast and equitable students' knowledge assessment methodologies and environments.

The advantages offered by the information and communication technologies in training and educational process can be summarizing as follows:

- Facilitate organizational learning, by transforming tacit knowledge into explicit knowledge [7];
- Provide across the country and across the globe educational access [8];
- Provide new tools to support collaborative learning [9];
- Emphasis the instructional process based on relationships, inquiry and intervention [8];
- Provide an environment which allows the teacher to give students more options and responsibilities for their own learning [10];
- Promote active learning in the learning process, producing and sharing knowledge, and participating at times as expert [11].

At higher education level, computers are used also as instruments in the students' knowledge assessment, being found as common type of evaluation in sciences and applied sciences domains: computer science [12], engineering [13], medicine [14], health sciences [15], pharmaceuticals [16], etc.

According with the necessity of development of an objective system for students' knowledge evaluation at university level, starting with experiences obtained by creation of the multiple choice examination system for general chemistry [17,18], an auto-calibrated online system was developed [19] and assessed [20]. The materials chemistry knowledge assessment of the undergraduate first year students' was study by the used of the auto-calibrated online system and the obtained results are presented.

MATERIAL AND METHOD

At Technical University of Cluj-Napoca, Romania, the Materials Sciences and Engineering Faculty, the curriculum contains as core course for the first year of study the Materials Chemistry topic. According with course description and with the subject matter, the Materials Chemistry course contains tutorials and laboratory sessions; at the end of the course there is an exam for students' knowledge assessment. In the university year 2005-2006, there were enroll in the first year of study at the Materials

Sciences and Engineering Faculty a number of ninety-nine students. All students participated at the lectures (PowerPoint presentations) and laboratory sessions that included practical activities and/or computer aided learning environments. Full copies of materials were available and the students had the possibility to enroll voluntarily into the team responsible for creation of the multiple-choice questions (MQCs) banking.

The imposed rules in multiple-choice banking creation were as follows:

- Each question has a statement or a situation (steam) and a list of five suggested solutions (options);
- Each question has at least one and no more than four correct options;
- On the received subject matter, each student enroll in the MQCs banking creation was reasonable by creation of proportional number of questions with one, two, three and respectively four correct options.

Each student that completes the assumed activities received bonus points to the final mark, according with the quality of work. For example, if the student respect the proportion between questions with one, two, three and four correct option receives a number of points as bonus. Penalties were applied (a number of point were subtract from the bonus points) when the imposed rules in MQCs creation were not respected and when the created questions were wrong (errors in statement and/or in option(s)).

The methodology of undergraduate students' knowledge assessment was included:

- Place of the examination: at test center;
- Type of examination: computer- and teacher-assisted;
- Period and time of examination: according with the structure of academic year and with the students and teacher schedule;
- Number of MCQs: thirty;
- Generation of the MCQs tests: double randomization from MCQs banking (randomization of the steatement and randomization of the options order);
- Number of tests: as many time as the student wanted in the imposed period of time and according with the schedule;
- Penalties when the student gives up shortly after the test begun.

The applied score and results methodologies contains the following specifications:

- Each question received one point if all the correct answers and none of the distractors (the incorrect answers presented as a choice) are selected, and zero points otherwise: *all-or-none rule*;
- At the end of the test, student identification data, the time when the test begun and ended, the number of correct answers and the average time needed to gave a correct answers are displayed;
- Starting from the distribution of the individual scores (which include the number of correct answer and the average time needed to give a correct answer) the system assign to the lower score the mark equal with four (the exam is failed) and to the highest score the mark equal with ten (ten being the best mark). The students' marks were auto-calibrated each time when a student performed a test.

All students from the Faculty of Materials Science and Engineering were included into the study. They had the possibility to familiarize with the evaluation environment before final examination as many time as they wished. A number of five variables were collected from each evaluation during the imposed examination period: student first and second name, data and time when the test begin and end (as yy.mm.dd hh.mm.ss format), the number of correct answers, and the time needed to give a correct answer. Data were store into a database and were summarized and analyzed with Statistica at a significance level of 5%. The 95% confidence intervals for proportions were calculated by the use of an original method, based on the binomial distribution hypothesis [21].

RESULTS

The auto-calibrated online evaluation system on materials chemistry was developed and is available via the address: http://vl.academicdirect.org/general_chemistry/materials_chemistry/, the access being restricted to the test center by checking the IP addresses.

Each student performed at least one time the materials chemistry test by the use of the auto-calibrated online evaluation system. For each student the system generated a test with thirty MQCs out of eight hundred sixty-one. The MCQs banking contains: 34.4% questions with one correct option (95%CI [31.3, 37.7]), 23.2% questions with two correct options (95% CI [20.5, 26.2]), 20.9% questions with three correct options (95%CI [18.3, 23.8]) and 21.4% with four correct options (95%CI [18.7, 24.3]).

The distributions of the number of tests express as relative frequency and associated 95% confidence intervals are in table I.

The range of days between two tests, when was applicable, varied from zero (the student performed the test twice in the same day) to 21 days; in 75.75% cases were less than or equal with 10 days. Nine students out of ninety-nine (95%CI [4.05-16.15]) performed the test twice in the same day.

TABLE I.
DISTRIBUTION OF THE NUMBER OF TESTS

Test	Students		
	No.	f_r (%)	95% CI
One	12	12.12	[6.07-20.19]
Two	55	55.56	[45.46-65.65]
Three	20	20.20	[13.14-29.28]
Four	9	9.09	[4.05-16.15]
Five	1	1.01	[0.01-5.04]
Six	1	1.01	[0.01-5.04]
Seven	1	1.01	[0.01-5.04]
Total	99	100	

In conformity with the specification of the auto-calibrated online system, when the students had more than one test, the less performing results (in terms of the number of correct answers and/or the average time needed to give a correct answer) was not took into consideration in computing the final mark. Statistical characteristics (express as average – mean, median, mode, standard deviation – StdDev, minimum – Min and maximum – Max) of the testing points (C_{tp}) and of the average time needed to give a correct answer (C_{tm}) are summarized in table II.

TABLE II.
STATISTICAL CHARACTERISTICS OF THE TESTING POINTS AND OF THE AVERAGE TIME NEEDED TO GIVE A CORRECT ANSWER

Test	Characteristic	Mean	Median	Mode	StdDev	Min	Max
First	C_{tp}	12.00	11.40	5.3	6.07	0.8	24.7
	C_{tm}	1.64	1.24	1.23	1.13	0.15	4.76
Second	C_{tp}	16.07	16.25	17.4	6.45	3.8	30.2
	C_{tm}	2.51	2.17	2.13	1.49	0.29	6.64
Third	C_{tp}	17.63	17.75	N.A.	6.70	7	27.6
	C_{tm}	2.86	2.57	N.A.	1.52	0.71	5.76
Forth	C_{tp}	20.33	19.90	N.A.	1.69	18.9	22.2
	C_{tm}	3.38	3.48	N.A.	0.52	2.81	3.84
Sixth	C_{tp}	21.65	21.65	N.A.	5.87	17.5	25.8
	C_{tm}	3.55	3.55	N.A.	1.87	2.23	4.87
Seventh	C_{tp}	23.00	23.00	N.A.	N.A.	N.A.	N.A.
	C_{tm}	3.98	3.98	N.A.	N.A.	N.A.	N.A.

StdDev = standard deviation; Min = minimum; Max = maximum; N.A. = not applicable
C_{tp} = testing points; C_{tm}= average time needed to give a correct answer

Twenty students out of ninety-nine (20.20%) performed the examination for three times. The less performing test in terms of number of corrected answer or the average time needed to give a correct answer was withdrawn. The distributions of the testing points give by the students which performed the test three times are in figure 1 (the students' id are from left to right: std_6, std_11, std_15, std_17, std_18, std_19, std_20, std_23, std_27, std_35, std_37, std_40, std_48, std_58, std_67, std_74, std_78, std_81, std_88, and std_99).

Nine students out of ninety-nine (9.09%) performed the test for four times. For this sample of student, the graphical representation of the testing points after the less performing test were withdrawn are in figure 2.

The comparisons between the average of testing points and of the average time needed to give a correct answer according with the number of test are summarized in table III. Note that the comparisons were performed between results obtained at the initial and at the final evaluation.

Comparison of the testing points obtained by students at the initial and final test (after withdrawn of the less performing test) shown that the average of the testing points were statistical significant higher at the final (18.741) evaluation comparing with the initial (8.675) evaluation ($p < 0.0001$, $n_{valid} = 32$, see figure 3).

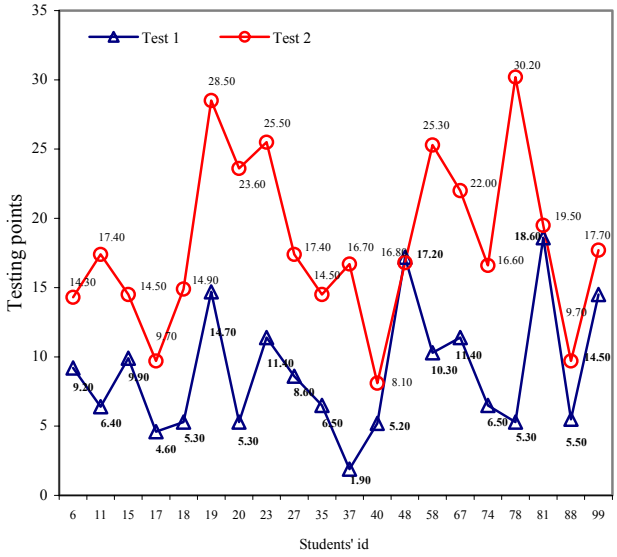

Figure 1. Distribution of the testing points for sample of students that performed the test three times

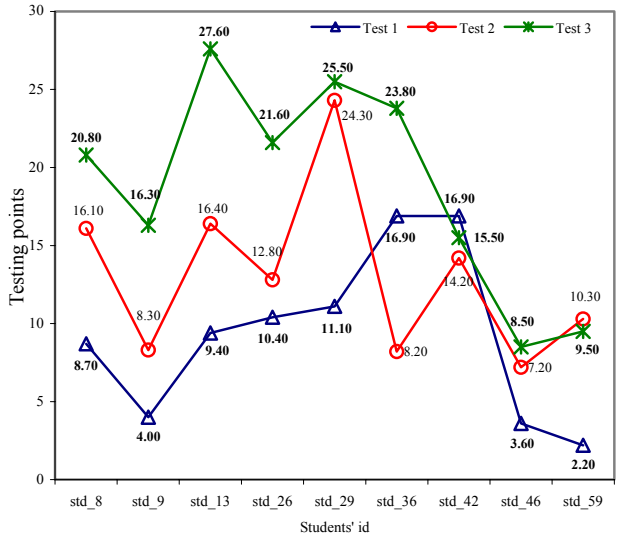

Figure 2. Distribution of the testing points for sample of students that performed the test four times

TABLE III
COMPARISON BETWEEN TESTS: TESTING POINTS (C_{tp}) AND AVERAGE TIME NEEDED TO GIVE A CORRECT ANSWER (C_{tm})

No. of tests	First test			Last test			p-aC_{tp}	p-aC_{tm}
	aC_{tp}	aC_{tm}	n_{valid}	aC_{tp}	aC_{tm}	n_{valid}		
One	10.68	1.322	12	N.A.	N.A.	N.A.	N.A.	N.A.
Two	14.22	2.052	55	N.A.	N.A.	N.A.	N.A.	N.A.
Three	8.915	1.103	20	18.145	2.891	20	< 0.0001	< 0.0001
Four	9.244	1.070	9	18.789	3.158	9	< 0.005	< 0.005
Five	4.30	0.40	1	18.90	2.81	1	N.A.	N.A.
Six	9.60	0.97	1	25.8	4.87	1	N.A.	N.A.
Seven	2.20	0.15	1	23.00	3.98	1	N.A.	N.A.

N.A. = not applicable

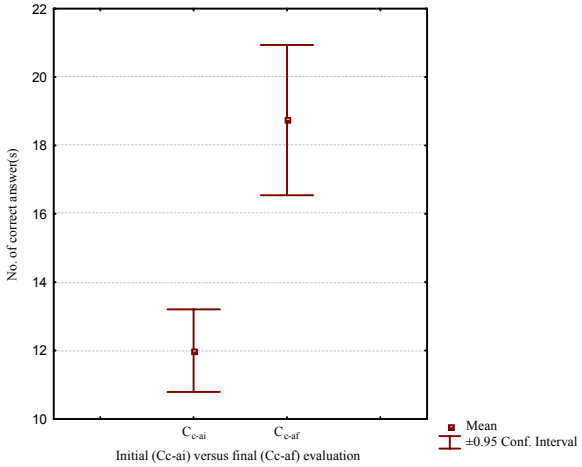

Figure 3. Distribution of the number of correct answers gave by students at initial and final evaluation

Comparison of the average time needed to give a correct answer for the initial and final evaluation reveled that the average time was significantly higher at the final (3.059) evaluation comparing with the initial (1.038) evaluation ($p < 0.0001$, $n_{valid} = 32$, see figure 4).

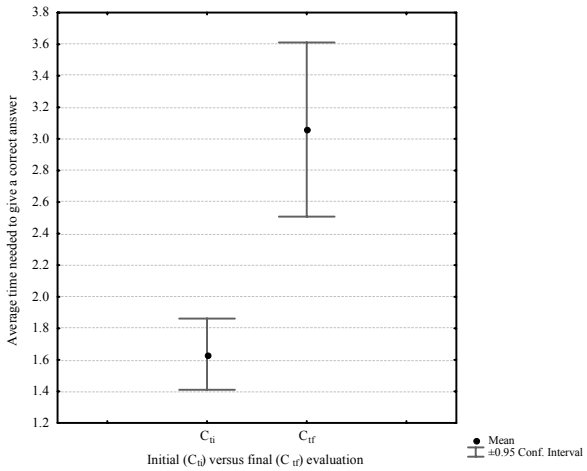

Figure 4. Distribution of the average time needed to give a correct answer at the initial and final test

DISCUSSION

The evaluation of the students' knowledge is an obligatory task at the end of a course for undergraduate students. The classical essay examination is not the proper type when the class is of ninety-nine students. According with speed, accuracy, objectiveness and fairness [22], testing methods which contains multiple-choice questions are frequently used [23,24].

The developed auto-calibrated online system had advantages and disadvantages. One of the main advantage results from the active implication of the students in creation of multiple-choice banking. This activity motivates students to ask questions and to find answers, implying them into an active learning process and an active interaction with the teacher, processes useful in acquiring materials chemistry knowledge. Another advantage refer the speed of examination: on a class of ninety-nine students, the time needed to assess each essay is significantly greater comparing with the use of auto-calibrated online system. At the end of the examination, according with data stored into database, the system display the number of correct answers, and allows to the student to view his/her performances in terms of number of correct answers and the average time needed to give a correct answer reported to the results obtained by his/her colleagues. Regarding the objectiveness and fairness, the proposed auto-calibrated online system discourages cheating. On one hand, with a database of eight-hundred sixty-one multiple choice questions is a little bit hard to learn just the questions (note that the randomization process is double: a randomization of the question and a randomization of the order of the five options). On the other hand, the number of distinct tests generated based on the questions store into database is of C_{861}^{30} which is equal with $2.54 \cdot 10^{55}$, thus the probability that two students to have the same test is very small. The system has its disadvantages. The most important disadvantage is represented by the assessment environment: each student is able to write on a paper but there is no certitude that every student had minimum computers skills. In order to withdraw this disadvantage, the students had possibility to use before the examination the evaluation environment in order to familiarize with interface, modality of choosing the answer and of navigation. With these occasions, the students had also the possibility to assess their knowledge and to identify personal gaps.

Each student had the possibility to test his/her knowledge as many times as desired into the imposed period of time. Twelve out of ninety-nine students performed the evaluation once, being content with obtained performances. Fifty-five students decide to performed the assessment for the second time and the average of testing points was greater with four points comparing with first evaluation (see table II). Analyzing the results in assembly it can be observed that the average of the testing points increase with the number of examination (see table II, figure 1 and 2). The greatest difference between testing points it can be observed on sample of students which performed the test three times (see figure 1, std_78) when the difference is of 24.90 points. In the sample of students which performed the test four times, the greatest difference was of 18.20 points (see figure 2, student id std_13, difference between last and first examination). There were also observed the decreasing of performances from the first test to the final test: -0.40 points (see figure 1, student id std_48), respectively -1.40 points (student id std_42, figure 2). With few exceptions, the students presented to the third or fourth

examination after they acquired more knowledge on materials chemistry topics with two exceptions: std_48, respectively std_42.

Analyzing the average time needed to give a correct answer (see table II) it can be observed that the value increase with the number of test. In other word, at first evaluation, the average time needed to give a correct answer was of 1.64 seconds and increase with evaluations until 3.98 seconds (seventh evaluation). The greatest difference is observed between first and second evaluation; the time needed to give a correct answer in the second evaluation is greater with 0.87 seconds comparing with first evaluation. These increases can be explained as follows: the students realized that they need to read more carefully the questions and associated options in order to make de correct chooses.

Comparing the average of testing points and respectively the average time needed to give a correct answer on the samples of students which performed the test more than two times (see table III), it can be observed that the performances were significantly better (p-value always less than 0.005). Further study are necessary in order to analyze the answer gave by each student to each question, study that can provide useful information about the materials where students had difficulties in understanding and learning. The practical activities, the seminaries and the courses on materials chemistry topic could be improved based on obtain information.

On assembly, it can be observed that the performances of students were significantly better at the final evaluation comparing with initial evaluation in terms of testing points (see figure 3) as well as in terms of average time needed to give a correct answer (see figure 4).

CONLCUSIONS

The proposed auto-calibrated online evaluation system proved to offer a stable and valid evaluation environment of undergraduate students' knowledge assessment on materials chemistry topic.

Students' performances in terms of testing points and of average time needed to give a correct answer revealed to be improved at final evaluation comparing with first evaluation, showing an improvement in acquired materials chemistry knowledge.

ACKNOWLEDGMENT

The research was supported in part by UEFISCSU Romania through projects ET108/2006.

REFERENCES

[1] E. Segers, and L. Verhoeven, "Long-term effects of computer training of phonological awareness in kindergarten", *J. Comput. Assist. Lear.*, vol. 21, issue 1, pp. 17-27, February 2005.

[2] E. H. Khan, "Use of computer-assisted learning in a primary school", *Computers & Education*, vol. 13, issue 4, pp. 355-362, 1989.

[3] A. Albirini, "Teachers' attitudes toward information and communication technologies: the case of Syrian EFL teachers", Computers *& Education*, vol. 47, issue 4, pp. 373-398, December 2006.

[4] J. Meijer, and F. Riemersma, "Teaching and testing mathematical problem solving by offering optional assistance", *Instructional Sci.*, vol. 30, issue 3, pp. 187-220, 2002.

[5] N. Hiramanek, "Self directed learning and continuing medical education", *Aust. Fam. Physician.*, vol. 34, issue 10, pp. 879-880, October 2005.

[6] Information and Communication Technologies in Teacher Education [online]. [cited 2006 September] Available from: URL: http://unesdoc.unesco.org/images/0012/001295/129533e.pdf#search=%22Information%20and%20Communication%20Technology%20-%20advantage%20in%20educational%20process%22

[7] L. Falconer, "Organizational learning, tacit information, and e-learning: A review", *Learning Organization*, vol. 13, issue 2, pp. 140-151, 2006.

[8] H. Uzunboylu, "International review", *Educ. Tech. Res.*, vol. 54, issue 2, pp. 201-209, April 2006.

[9] M. G. Madariaga, A. T. Evans, W. Brobbey, M. Phillips, E. Lo, K. Rezai, D. N. Schwartz, G. M. Trenholme, and R. A. Weinstein, "Learning by doing: Developing fellows' academic skills through collaborative research", *Med. Teach.*, vol. 28, issue 1, pp. 77-80, February 2006.

[10] E. De Graaff, and A. Kolmos, "Characteristics of Problem-Based Learning", *Int. J. Eng. Educ.*, vol. 19, issue 5, pp. 657-662, 2003.

[11] T. Newby, D. Stepich, J. Lehman, and J. Russel, "Instructional technology for teaching and learning", Upper Saddle River, Merrill/Prentice Hall, New Jersey, 2000.

[12] S. Encheva, and S. Tumin, "Web-based assessment tests supporting learning", *Lect. Notes Comput. Sc.*, vol. 3807 LNCS, pp. 134-143, 2005.

[13] F. Farinha, E. Portela, C. Domingues, and L. Sousa, "Knowledge-based systems in civil engineering: Three case studies", *Adv. Eng. Softw.*, vol. 36, issue 11-12, pp. 729-739, November 2005.

[14] R. S. Day, "Challenges of biological realism and validation in simulation-based medical education", *Artif. Intell. Med.*, vol. 38, issue 1, pp. 47-66, September 2006.

[15] A. Fekete, J. Kay, J. Kingston, and K. Wimalaratne, "Supporting reflection in introductory Computer Science", *SIGCSE Bull.*, pp. 144-148, 2000.

[16] D. Sibbald, "A student assessment of the Virtual Interactive Case Tool for Asynchronous Learning (VITAL) and other self-directed learning formats", *Am. J. Pharm. Educ.*, vol. 68, issue 1, pp. 1-7, 2004.

[17] H. I. Nascu, and L. Jäntschi, "Multiple Choice Examination System 1. Database Design and Implementation for General Chemistry", *LJS*, issue 5, pp. 18-33, 2004.

[18] H. I. Nascu, and L. Jäntschi, "Multiple Choice Examination System 2. Online Quizzes for General Chemistry", *LEJPT*, issue 5, pp. 26-36, 2004.

[19] L. Jäntschi, and S. D. Bolboaca, "Auto-calibrated Online Evaluation: Database Design and Implementation", *LEJPT*, issue 8, pp. 178-191, 2006.

[20] S. D. Bolboaca, and L. Jäntschi, "Computer-Based Testing on Physical Chemistry Topic: A Case Study", unpublished.

[21] Binomial Distribution [online]. ©VLFS, 2005 [cited 2006 September]. Available from: URL: http://vl.academicdirect.org/applied_statistics/binomial_distribution/.

[22] S. Toby, and R. J. Plano, "Testing, Testing: Good Teaching Is Difficult; So Is Meaningful Testing", *J. Chem. Educ.*, vol. 81, no.2, pp. 180-181, 2004.

[23] J. J. Venkrbec, J. Kousal, R. Berger, and J. Stetina, "Education programmes in materials science and engineering", *Mat. Sci. Eng. A-Struct.*, vol. A199, issue 1, pp. 79-86, August 1995.

[24] M. Jenkins, E. K. Goo, "Concept-based instruction and Personal Response Systems (PRS) as an assessment method for introductory materials science and engineering", *ASEE Annual Conference and Exposition, Conference Proceedings*, pp. 2133-2142, 2005.