AUTHOR(S): FERNANDEZ-ALEMAN, J.L., LOPEZ-GONZALEZ, L., GONZALEZ-SEQUEROS, O., JAYNE, C., LOPEZ-JIMENEZ, J.J. and TOVAL, A

TITLE: The evaluation of i-SIDRA - a tool for intelligent feedback - in a course on the anatomy of the locomotor system.

YEAR: 2016

Publisher citation: FERNANDEZ-ALEMAN, J.L., LOPEZ-GONZALEZ, L., GONZALEZ-SEQUEROS, O., JAYNE, C., LOPEZ-JIMENEZ, J.J. and TOVAL, A. 2016. The evaluation of i-SIDRA - a tool for intelligent feedback - in a course on the anatomy of the locomotor system. International journal of medical informatics [online], 94, pages 172-181. Available from: https://doi.org/10.1016/j.ijmedinf.2016.07.00

OpenAIR citation: FERNANDEZ-ALEMAN, J.L., LOPEZ-GONZALEZ, L., GONZALEZ-SEQUEROS, O., JAYNE, C., LOPEZ-JIMENEZ, J.J. and TOVAL, A. 2016. The evaluation of i-SIDRA - a tool for intelligent feedback - in a course on the anatomy of the locomotor system. International journal of medical informatics, 94, pages 172-181. Held on OpenAIR [online]. Available from: https://openair.rgu.ac.uk.

The evaluation of i-SIDRA - a tool for intelligent feedback - in a course on the anatomy of the locomotor system

José Luis Fernández Alemán[1]*, Laura López González[2], Ofelia González Sequeros[2], Chrisina Jayne[3], Juan José López Jiménez[1], Ambrosio Toval[1]

[1] *Faculty of Computer Science, Department of Informatics and System, University of Murcia, Spain*

[2] *Faculty of Medicine, Department of Human Anatomy, University of Murcia, Spain*

[3] *Coventry University, UK*

*Corresponding author. Faculty of Computer Science, Campus of Espinardo, Murcia, Spain. Telephone: (+ 34 868) 884621; fax: (+ 34 868) 884151

Email addresses: aleman@um.es (José Luis Fernández-Alemán), lalogo@yahoo.com (Laura López González), sequeros@um.es (Ofelia González Sequeros), ab1527@coventry.ac.uk (Chrisina Jayne), juanjoselopez@um.es (Juan José López Jiménez), atoval@um.es (Ambrosio Toval)

KEYWORDS:

E-LEARNING, LOCOMOTOR SYSTEM, NEURAL NETWORK, EXPERIMENT

Abstract

Objective: This paper presents an empirical study of a formative mobile-based assessment approach that can be used to provide students with intelligent diagnostic feedback to test its educational effectiveness.

Method: An audience response system called SIDRA was integrated with a neural network-based data analysis to generate diagnostic feedback for guided learning. A total of 200 medical students enrolled in a General and Descriptive Anatomy of the Locomotor System course were taught using two different methods. Ninety students in the experimental group used intelligent SIDRA (i-SIDRA), whereas 110 students in the control group received the same training but without employing i-SIDRA.

Results: In the students' final exam grades, a statistically significant difference was found between those students that used i-SIDRA as opposed to a traditional teaching methodology (T(162) = 2.597; p = 0.010). The increase in the number of correct answers during the feedback guided learning process from the first submission to the last submission in four multiple choice question tests was also analyzed. There were average increases of 20.00% (Test1), 11.34% (Test2), 8.88% (Test3) and 13.43% (Test4) in the number of correct answers. In a questionnaire rated on a five-point Likert-type scale, the students expressed satisfaction with the content (M=4.2) and feedback (M=3.5) provided by i-SIDRA and the methodology (M=4.2) used to learn anatomy.

Conclusions: The use of audience response systems enriched with feedback such as i-SIDRA improves medical degree students' performance as regards anatomy of the locomotor system. The knowledge state diagrams representing students' behavior allow instructors to study their progress so as to identify what they still need to learn.

1. Introduction

The use of mobile technology is becoming an increasingly more important area of study in health science education [1, 2], with disease classification, medical dictionaries, medical calculators, clinical guidelines, classroom assessment systems and teaching evaluation being just some of their application domains [3]. Classroom response technologies are currently being used in many educational institutions in general [4-6] and in health science faculties [1, 7-11] in particular. A classroom response system (CRS) is a handheld piece of equipment which allows students to respond electronically to questions that lecturers pose in class. A response receiver that plugs into a USB port or a server connected to the Internet receives the students' choices. A software polling program installed in a PC allows instructors to prepare and launch tests.

CRS has been used for various purposes in the medical classroom: (1) to evaluate classmates [8], (2) to evaluate current or previous knowledge [12], (3) to check understanding [11], (4) to maintain attention [13], (5) to facilitate participation, and (6) for post hoc analysis [10]. Experiences and empirical studies in different medical fields such as cardiology [10], radiology [14], gynecology [13], physiology [15], family medicine [16], physical basis of medicine [8], clinical pharmacology [17] and the prescription of safe medication [11] have been conducted by using this educational technology.

CRS usually employs multiple-choice questions (MCQs) which are useful as regards marking in a reliable and objective manner and allow instructors to give feedback in a short amount of time [18]. The fact that MCQs can be used to measure complex abilities and understanding [19] makes CRS an invaluable tool for medical education.

Response-driven feedback for multiple-choice questions (MCQs) based on intelligent analysis is a relatively new teaching approach that has been applied in few disciplines to date: computer science programming [20], medical-surgical nursing [21, 22] and English language teaching [23]. The responses to MCQs are automatically analyzed to understand the students' learning behaviors and to design diagnostic feedback that is customized to each student.

In this paper, a formative mobile-based assessment approach to provide students with intelligent diagnostic feedback is proposed and empirically evaluated. To the best of the authors' knowledge, no other studies have combined CRS and an intelligent analysis of students' responses. Our empirical study showed evidence of the benefits of the use of audience response systems enriched with feedback in a course on the anatomy of the locomotor system. In particular, we found a positive difference in the performance of those students taught using i-SIDRA when compared with those taught using a traditional teaching methodology. Moreover, i-SIDRA promoted the students' involvement in class activities and allowed the instructors to maintain the students' attention and motivation during the learning process. The results of this statistical study will be used to improve medical degree students' anatomy of locomotor system education, and to address future directions as regards designing the feedback associated with the groups of responses of the MCQs formulated.

2. E-learning theories

Three learning theories are commonly applied to the e-learning domain: Behaviourism, Cognitivism, and Constructivism [24]. Behaviourists recommend a structured, deductive approach in online learning. The learning material is usually broken down into small instructional steps using the concept of drill and practice, thus assessing the learner's achievement levels, and providing external feedback [25]. The cognitive school focuses on learners' receiving and processing of information which is transferred to the long-term memory through different cognitive processes. Instructional designers therefore have to prepare online learning materials and provide adequate support for students with different learning styles [24]. The constructivist school suggests that students construct own knowledge from their learning experience [26]. e-learning is therefore an active process in which learners control their learning process. When using learning online by adopting Constructivism, students should have time to reflect on the learning content. Questions on the course contents can be used throughout each lesson to encourage reflection. Instructors should focus on interactive learning activities to promote higher-level learning. Communal constructivism is an extension to socio-constructivism in which learners construct personal knowledge as a result of their experiences and interactions with others [27].

Innovations in e-learning technologies have brought about a revolution in education. Some examples of this are collaborative learning, individualized learning, and the new role of the teacher as a facilitator of learning and appraiser of competency [28]. Multiple factors which can affect the e-learning adoption have also been studied. The authors of a study based on the Technology Acceptance Model found that women's opinions were more significantly determined by perceptions of ease of use, whereas men were more strongly influenced by their perception of the usefulness of e-learning [29]. The authors therefore suggest that gender should be considered as a factor in the development and testing of e-learning theories.

In another study, the 33 influence factors included in the five aspects of innovation adoption according to the theory of Rogers were used to investigate the influence factors of e-learning adoption [30]. Questionnaires were administered to 160 undergraduates in order to identify the key influence factors of e-learning adoption intention. Cost was the largest influence factor of perceived innovative attributes. In contrast, the factor with the least influence was the general advantages perceived which allow people to believe that the new method is better than the traditional one. Involvement for learning has been also linked to learning outcomes [31].

Nevertheless, a number of gaps have been identified between the bodies of knowledge relating to learning theories, instructional design principles, and student learning [32], and there is thus a great need to provide pedagogically effective e-learning environments [33].

3. The e-learning system i-SIDRA

The intelligent SIDRA (Sistema De Respuesta Inmediata de la Audiencia in Spanish) is a free and publicly available application with a client-server architecture. This tool allows a survey taker to create, collect and analyze responses to multiple choice questions (MCQs). The intelligent SIDRA (i-SIDRA) system has different interfaces for students and instructors. Users can access it via the web or mobile devices. The i-SIDRA mobile version can currently be downloaded from the Apple App Store and from Google Play.

In i-SIDRA, a set of MCQs related to a specific topic is called a test. A student can read the MCQs, respond to the questions, and view the percentage of correct answers to each question at the end of or during the online session. In contrast, a teacher can create an MCQ test, run a test, export the results of a test, and obtain the information related to students' answers. A test can be set up in such a way that the participating students' responses are identifiable or anonymous.

The e-learning process applied in our study consists of the following steps:

Step 1. Collect data for neural network. MCQ tests are created by teachers and taken by students to obtain their responses. In our experiment, responses were collected from four ten-MCQ exams taken during the 2012-13 course. The input patterns used to train SDNN are obtained by encoding each of the students' responses into a binary form. For a five-choice test, with the answers A), B), C), D) and E), a possible codification is: A- 00001, B- 00010, C- 00100, D- 01000 and E- 10000. Table I shows five examples of encoded responses for a six-MCQ exam.

Table I. Example of five input patterns and two groups for SDNN. Id: Response identifier; G: Group generated by SDNN

Id	G	Response	Encoded response
R1	G1	[A, A, B, C, D, E]	[0,0,0,0,1,0,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,0]
R2	G1	[A, A, D, A, E, A]	[0,0,0,0,1,0,0,0,0,1,0,1,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,1]
R3	G1	[A, B, E, D, E, A]	[0,0,0,0,1,0,0,0,1,0,1,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,1]
R4	G2	[B, C, E, A, B, C]	[0,0,0,1,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,1,0,0]
R5	G2	[B, A, D, A, B, C]	[0,0,0,1,0,0,0,0,0,1,0,1,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,1,0,0]

Step 2. Train the neural network. For each test, the SDNN learns to group the input patterns based on their similarities and differences. The groups or clusters generated represent different states of knowledge with regard to the questions formulated in the test. The groups of responses share the same or similar mistake(s) and hits. For example, one group may be formed of R1, R2 and R3 (Table I), in which every response has in common the answer A to question 1. The other answers to the other questions vary within the group. A good training of the neural network relies upon the use of representative data as input patterns. In our experiment, the number of responses from the aforementioned cohorts of students used to train the SDNN was 96 for all the four tests. The number of groups formed for each test depends on the variation in student responses. A relatively small and therefore manageable number of groups of similar answers are usually obtained. In our experiment, the number of groups or knowledge states obtained was 14, 8, 11 and 11 for Test 1, Test 2, Test 3 and Test 4, respectively. A more detailed description of the SDNN can be found in Appendix A.

Step 3. Introduce the MCQs tests into i-SIDRA by using the instructors' Web interface. The instructor logs into the i-SIDRA system and creates each test using the i-SIDRA interface. In our experiment, four ten-MCQ tests were entered in this step.

Step 4. Write and include feedback in i-SIDRA. In i-SIDRA, the teachers are presented with the MCQs. An edit window is used to introduce particular feedback for each possible response to a question. In our experiment, the medical teachers designed and completed the feedback based on each incorrect answer to the questions. The feedback was formulated in such a way that it did not reveal the correct answer, or even which question the feedback referred to. The i-SIDRA system then automatically generated a different feedback for each group or knowledge state obtained in Step 2. To carry out this task, the i-SIDRA system used the templates of student responses associated with each group and the explanatory text prepared by the instructors. For example, in the case of the template "A A/B B/D/E A/C/D D/E A/E" for group G1 in Table I, the students responded A (e.g. "*Between the Subscapularis, Latissimus dorsi and Serratus anterior*") to question 1 (e.g. "*Scapulothoracic sinsarcosis has two sliding areas:*"), and A (e.g. "*The Extensor carpi ulnaris*") or B (e.g. "*The Extensor carpi radialis longus*") to question 2 ("*Which of these muscles is attached to the carpal bones?*"), and so on. The feedback texts introduced by the medical teachers for incorrect response A to question 1 (e.g. "*Remember that the Latissimus dorsi is located superficially at the back of the trunk and does not belong to shoulder girdle muscles*"), and incorrect responses A (e.g. "*The Extensor carpi ulnaris rises from the lateral epicondyle of the humerus, by the common Extensor tendon. It is inserted on the ulnar side of the base of the fifth metacarpal bone*") and B (e.g. "*The Extensor carpi radialis longus rises from the lateral supracondylar ridge of the humerus, from the lateral intermuscular septum, and from the common tendon of the origin of the Extensor muscles of the forearm. It is inserted into the dorsal surface of the base of the second metacarpal bone, on its radial side*") to question 2 were finally combined and included in the feedback of the pattern grouping G1. Note that in our study, feedback is only created for incorrect options. In some questions, feedback might also be needed for correct options if instructors identify this possibility. Atomic feedback for correct options could therefore be included by adopting a defensive position. However, the feedback for each pattern grouping could be overloaded. On the other hand, if the atomic feedbacks associated with each incorrect response are similar, the feedback formed for a pattern grouping could contain redundant information. The instructor should therefore revise the feedback obtained to identify any problem and to confirm its validity. The intention of the feedback is therefore to address misunderstood concepts which may have caused the incorrect answers common to each pattern grouping. A full example including a questionnaire, atomic feedback and feedback associated with pattern groupings is provided in English as supplementary material.

Step 5. Launch the MCQs tests. The teachers run each MCQ test, which is available on the students' Web or mobile interface. MCQ tests present problems in the form of questions related to the topics addressed on the course. In our experiment, four MCQ tests were launched during the Fall term of 2013. Observe that the students involved in this step are different from those in Step 1.

Step 6. Receive students' answers. Students using i-SIDRA are asked to complete each test as many times as required in a time interval determined by the instructor (45 minutes in our study) or until all the right answers are provided. When a test is submitted by a student, i-SIDRA's server receives the students' answers and compares them with the answer expected. The i-SIDRA system records each student's response times for further analysis, while simultaneously classifying the responses into a pattern grouping that returns the specific associated feedback, provided that one or more questions were answered incorrectly. Note that students are not told how many or which questions were answered correctly, but only whether all of the questions were answered correctly. Fig. 1 shows an example of the MCQs and feedback in Test 2.

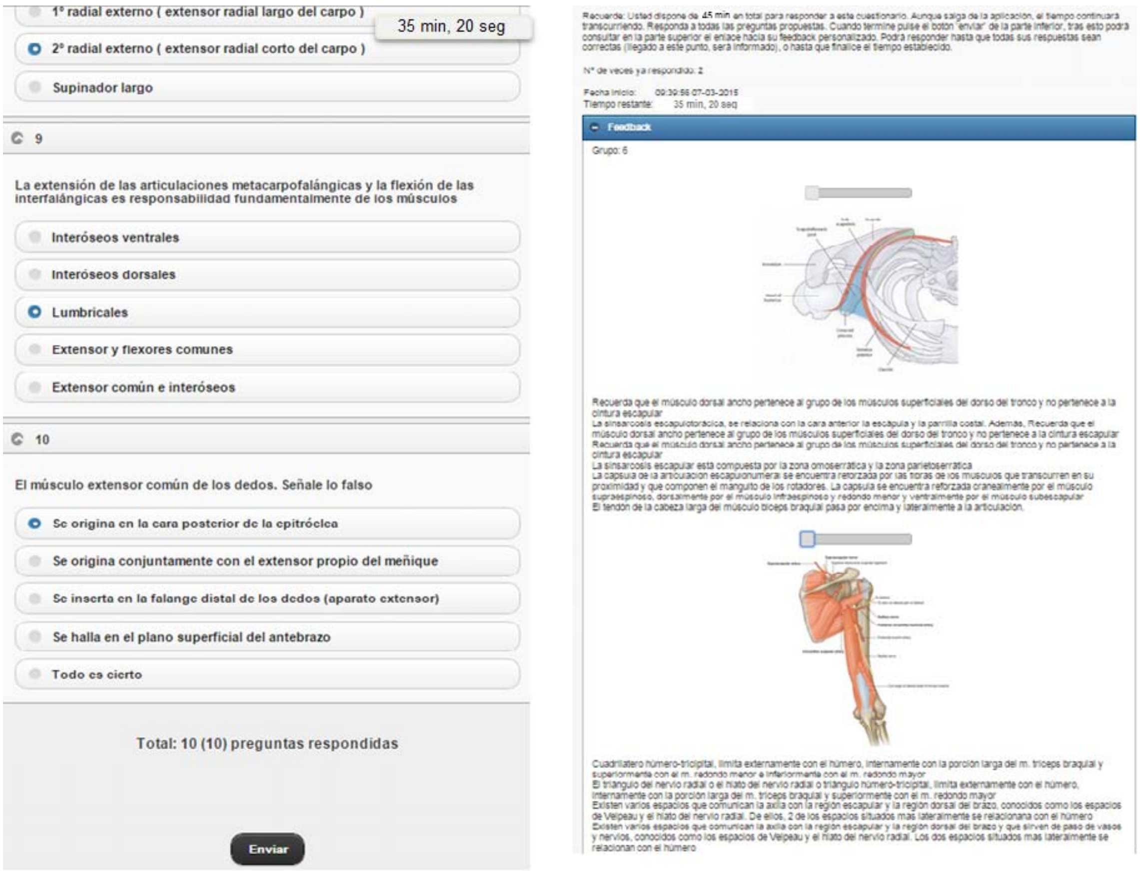

Fig. 1 - Example of questions and feedback associated with knowledge group 6 in Test 2. Images extracted from the course reference book [34].

Step 7. Show the results. At the end of the session, i-SIDRA shows to the teacher for each question the response time by each student, average response time and percentage of correct answers, based on each student's last submission. The instructors may show the correct answers to the MCQs and solve doubts or misconceptions in discussion groups. Statistical information related to the student's answers, such as number of submissions, transitory intermediate knowledge states and average time per submission, can be used to analyze the students' progress and identify any misconceptions that can be addressed in subsequent face-to-face sessions. The data collected can also be used to analyze the effectiveness of each knowledge state's feedback by identifying *sink* states, that is to say, knowledge states in which the number of correct answers does not increase after reading the feedback. The learning of individual students can also be studied by following a particular student's progress using the knowledge state transition diagram over time. The experiment is completed in lab or theory sessions using mobile smartphones or PCs connected to i-SIDRA.

The steps followed in the e-learning process using I-SIDRA are outlined in Fig. 2.

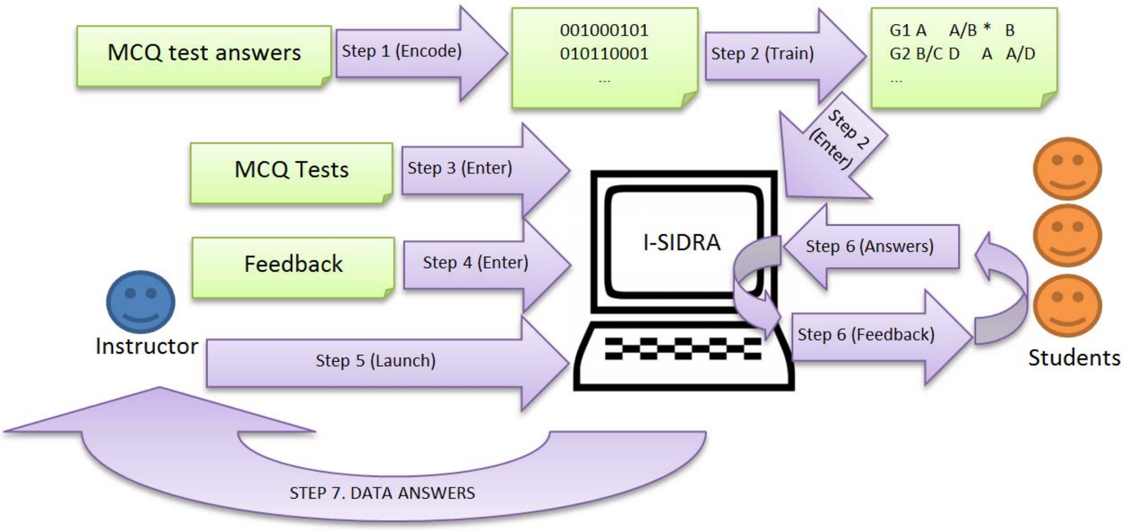

Fig. 2 - The e-learning process using i-SIDRA steps.

4. Method

4.1. Participants and Data Collection

This study involved a sample of first-year medical students enrolled in a General Anatomy of Human Musculoskeletal System (GAHMS) course during the academic year 2013/2014 at the University of Murcia (Spain). GAHMS is a first-term course which focuses on the study of general anatomy with an emphasis on the morphology of bone, joint and muscle systems. GAHMS was divided into 4 thematic blocks: Block 1: Introduction to gross anatomy and musculoskeletal anatomy of the trunk (thorax, abdomen and pelvis); Block 2: Musculoskeletal anatomy of the upper limbs; Block 3: Musculoskeletal anatomy of lower limbs; Block 4: Musculoskeletal anatomy of the head and neck. Students attend 4 h/week of lectures and 2 h/week of skills practice during a period of 15 weeks, completing a total of 6 ECTS credits.

A recruitment session was organized on the first class day during which potential participants were invited to participate. The potential participants received a detailed verbal presentation that highlighted the purpose of the study, a description of the procedures used, and an estimate of the time required to complete the tasks. A total of 200 students of mixed gender, age, computer experience and educational background gave individual oral consent. Students could withdraw from the study at any time without prejudicing their academic results.

4.2. Design

Two teaching methods on acquisition and retention of GAHMS skills and knowledge in the context of medical practice were designed for comparison. The selected sample was randomly divided into two groups. An experimental group comprising 90 students, which used i-SIDRA, and a control group of 110 participants received the same training concerning GAHMS but a neural network based tool was not employed.

Data were collected at four time points during the study: (1) demographic data, obtained from an institutional on-line tool for administration facilities, were collected from all the participants immediately prior to the teaching intervention; (2) four MCQ review tests consisting of 10 questions concerning the content covered in class were taken by the experimental group students using i-SIDRA, and by the control group using a discussion group; (3) a system evaluation questionnaire rated on a five-point Likert-type scale was administered to discover and appraise the students' experiences during the use of i-SIDRA; (4) a final examination took place immediately after the instructional period.

Note that the experimental group students took individual MCQ review tests by using i-SIDRA in a time interval of 45 minutes. The instructors then gave an anonymous i-SIDRA test containing the same questions, in the form of an audience response system. In this last test, the correct answers to the MCQs were shown to avoid the frustration that students may feel when all of the questions have not been responded to correctly [35], and doubts and misconceptions were resolved in discussion groups in a time period of 15 minutes. In contrast, after each thematic block the control group students took part in guided discussion groups in which the instructor analyzed different questions in the classroom to review the content covered in the block (including the knowledge tested in the i-SIDRA MCQs) and resolve any doubts in a time period of 60 minutes.

All of the questions were formulated by taking into consideration recommendations concerning accepted item-writing guidelines [36]. Moreover, the questionnaires included a maximum of 10 questions to avoid the fatigue effect. Up-to-date literature regarding current recommended medical practice was used to ensure relevant course content.

4.3. Hypotheses

The following hypotheses were investigated:

H1: The final exam grades of the students who use i-SIDRA will be higher than those of the other students. One independent variable (Teaching method) and one dependent variable (Performance measured by the final exam grades) were defined to test the statistical hypothesis.

H2. The feedback designed by instructors and delivered by the i-SIDRA system helps to clarify misunderstandings and helps students identify what they still need to learn. One independent variable (Time point) and one dependent variable (NumberCorrectAnswers, measured using the difference between the scores at the beginning of the i-SIDRA session and the scores at the end of the i-SIDRA session) were defined to test the statistical hypothesis.

H3. Students are satisfied when using i-SIDRA.

5. Results

5.1. Performance

Table II shows the descriptive statistics (median, mean, and standard deviation) for Performance. On average, the students in the experimental group (using i-SIDRA) achieved better results in the final exam than those in the control group. Observe that 14 students in the experimental group and 22 students in the control group were discarded because they did not take the final exam.

Table II. Descriptive statistics for Performance. “N”: Number of students; “M”: Mean; “Md”: Median; “SD”: Standard deviation.

Group	N	M	Md	SD
Experimental (i-SIDRA)	76	5.945	6.250	1.630
Control	88	5.235	5.310	1.838

A t-test for unpaired samples for the Performance variable was applied. The results revealed that there was a statistically significant difference when the students used i-SIDRA versus a traditional learning methodology ($T(162) = 2.597$; $p = 0.010$). The data therefore suggest that the use of i-SIDRA allowed the students to gain knowledge, thus leading us to confirm hypothesis H1.

Descriptive statistics about students, attempts and feedback’s effectiveness in i-SIDRA are provided in Table III. A total of 50, 38, 33 and 29 students in Test1, Test2, Test3 and Test4, spent more than the average time per student. It is noteworthy that one student increased his score by 233% (from 3 to 10 correct answers) in Test3.

Table III. Detail of the classification of the students and the attempts by experiment.

Description	Test1	Test2	Test3	Test4
Students				
Total number of students	87	82	72	67
Number of students attaining the “state of perfect knowledge” (see Section Discussion)	34	28	16	39
Number of students not attaining the “state of perfect knowledge”	53	54	58	28
Number of students with only one attempt or submission (students did not receive feedback)	4	5	9	30
Number of students with more than one attempt or submission (students received feedback)	83	77	63	37
Number of students attaining the “state of perfect knowledge” with only one attempt (no feedback)	3	3	2	22
Number of students attaining the “state of perfect knowledge” with more than one attempt (feedback)	31	25	14	17
Total number of attempts	846	469	406	572
Average time taken per attempt (minutes)	2,12	2,26	2,25	2,05
Number of attempts per student	9,72	5,72	5,63	4,14
Average time taken per student (minutes)	20,67	12,94	12,71	8,53
Maximum number of attempts made by a student	38	29	25	20
Feedback’s effectiveness				
Number of students with more than one attempt (improvement after feedback)	75	51	38	34
Number of students with more than one attempt (worsening after feedback)	2	7	9	0
Number of students with more than one attempt (neutral after feedback)	7	5	9	7

5.2 Analysis of the knowledge improvement in using i-SIDRA

The students were allowed to do each test several times within a specified time window (45 minutes). The evolution in the number of correct answers was studied to see how the students progressed towards a flawless test. The increase in the number of correct answers in the four tests was identified to check the effectiveness of i-SIDRA. There were average increases of 20.00% (Test1), 11.34% (Test2), 8.88% (Test3) and 13.43% (Test4) in the number of correct answers. The results of the t-test for paired samples for the NumberCorrectAnswers variable are shown in **Table IV**. In the four MCQs tests, there was a statistically significant difference between the scores in the first submission (at the beginning of the test) and the scores in the last submission (at the end of the test). These data suggest that the feedback written by teachers and provided by i-SIDRA allowed the students to learn new concepts and clarify misunderstandings, thus leading us to accept hypothesis H2.

Table IV. Descriptive statistics and paired Student's t-test results for NumberCorrectAnswers in four time points. "N": Sample size; "M": Mean; "SD": Standard deviation; "T": T-Student; "P": p value.

Time point	N	M	SD	T	P
FirstTestTimepoint1	87	6.59	1.789	12.420	0.000
LastTestTimepoint1	87	8.59	1.402		
FirstTestTimepoint2	82	6.79	2.077	5.899	0.000
LastTestTimepoint2	82	7.93	2.059		
FirstTestTimepoint3	72	5.68	2.466	4.722	0.000
LastTestTimepoint3	72	6.57	2.705		
FirstTestTimepoint4	67	7.25	2.470	6.268	0.000
LastTestTimepoint4	67	8.60	2.060		

In all the time-points, the scores obtained in the last test carried out using i-SIDRA are better than those obtained in the first test. Instructors can therefore have confidence in i-SIDRA's learning potential, since the feedback had a positive impact, thus supporting hypothesis H2.

5.3 Survey

The students then filled in a questionnaire concerning their participation in the experiment, the principal object of this survey being to collect feedback about the students' perceptions as regards their experience with i-SIDRA during the experiment execution. A total of 11 questions were formulated using a five-point Likert-type scale (5 = very high; 4 = high; 3 = medium; 2 = low; 1 = very low). **Table V** shows the means and standard deviations for the students' answers. A total of 71 students took part in the survey (81%). The use of i-SIDRA was positively evaluated by the students, indicating that hypothesis H3 is supported. Moreover, the students expressed satisfaction with the content provided and the methodology used when learning about anatomy. They generally found that the feedback helped them to learn concepts of anatomy, and most of them wanted to repeat the experience in another subject. Note that these perceptions confirm the findings obtained in Subsections 4.1 and 4.2, in which the scores obtained by the students in the final exam were significantly improved after using i-SIDRA during the academic course.

Table V. Means and standard deviations of students' perceptions. "M": Mean; "SD": Standard deviations; "MD": Median.

Id	Question	M	SD	M
Q1	You are satisfied with the use of i-SIDRA	4.1	0.8	4
Q2	The time allotted to each activity was sufficient	4.2	0.8	4
Q3	You found the use of i-SIDRA during the course useful	3.9	0.8	4
Q4	The i-SIDRA system promotes clarification and understanding of concepts	3.9	0.8	4
Q5	The i-SIDRA system favors and improves the learning process	3.9	0.7	4
Q6	The i-SIDRA system is useful as a self-assessment tool	4.3	0.6	4
Q7	The feedback received has been useful to improve my learning	3.5	1.1	3
Q8	I would like to use this system again in other subjects	3.9	0.8	4
Q9	I believe that the use of this system will positively influence my grades	3.5	0.8	4
Q10	I found that the content provided and the methodology used in the subject are suitable when learning about anatomy	4.2	0.7	4
Q11	The instructors used and incorporated i-SIDRA appropriately in the classroom	4.6	0.5	5

6. Discussion

6.1. Learning outcomes

Audience response technology has been identified as an education technology that may be invaluable in the monitoring of and improvement to medical education [10]. The hypothesis testing in our study has shown that the experimental group obtained significantly better marks than the control group in the final exam. These findings confirm previous research on the use of CRS [7, 15, 37] and intelligent feedback [20]. This can be attributed to the fact that the formative assessment improves students' performances in subsequent summative assessments [38]. Nevertheless, a study on the adoption of CRS without intelligent data analysis on a medical course regarding screening for breast cancer and cervical cancer reported no change in exam performance [39].

6.2. Feedback

Feedback is a key element in i-SIDRA. While answering the MCQs, the students were provided with a variety of different images and videos related to the locomotor system, with a direct application of theory into practice. The students received immediate feedback that was customized to their knowledge. A concept-oriented diagnostic feedback approach was used to encourage them to reflect on misunderstood concepts, while the feedback provided them with hints as to what they had not understood. This feedback was not over-detailed to avoid confusion, and came with a brief explanation or a reference to educational material, thus allowing the students to easily interpret and understand its purpose [23]. After reading, reflection and cognition, the students had the opportunity to try the test again.

The average increase in correct answers was between 8.88% (test 3) and 20.00% (test 1), which is significantly higher than in previous experiments [20] conducted on a computer science course (between 0.30% and 5.00%). This may be justified by the enhancer effect of CRS. Note that since the feedback did not state exactly which answer was wrong, as has occurred in other studies [40], the score increases, with a total of 10 questions in each test, and four or five possible answers, this could not have occurred by chance.

6.3. Participation and interactivity

Learner engagement is one of the benefits most commonly identified in literature [1]. Brainwave analysis has shown that the use of smartphones to carry out mobile polling in class is an effective means to increase student attention [41]. The active participation of students is a typical feature of CRS lecture delivery, especially for large classes in which the interactivity can be increased [38]. The interaction between faculty and students is also improved by using CRS [42]. Moreover, a summative assessment combined with CRS can be an incentive for attendance [43]. Although enthusiasm should be cautious, an enjoyable tool helps students maintain attention and motivation during the learning process [13]. Since students' privacy can also be maintained in the anonymous operation mode supported by i-SIDRA, anxiety and feelings of embarrassment and intimidation are alleviated [44], thus promoting and encouraging participation in the classroom. In our experiment, the use of anonymous i-SIDRA was accompanied by dialogues in addition to in-depth and broad-ranging discussions at the end of each test, and students felt connected to their other classmates who share common goals. Although data on participation and interactivity was not collected, the medical teachers observed a very high level of involvement in these post-test activities.

In the experimental group, the number of students involved on the course changed from 87 in the first test to 67 in the last test. This can be attributed to a very high workload in the form of seminars, lectures, and a final examination. Nevertheless, the dropout rate was higher in the control group (20.00%) than in the experimental group (15.55%). The experimental group students were more encouraged and had more energy to complete the course.

6.4. Satisfaction

A previous study reported that medical students on a clinical pharmacology course even preferred CRS to computer based testing [17]. Students' satisfaction with CRS use is positively documented in a majority of the related literature [1]. The reason for this phenomenon may be explained by the fact that formative assessment greatly improves students' satisfaction with a course [38]. Our survey was designed to cover the most frequently used scales in CRS literature [45]: learning and understanding. Students positively viewed the incorporation of i-SIDRA into lectures (Q1, mean, 4.1). We believe that the instructors' experience in CRS approaches had a positive effect on the students' perceived satisfaction. As observed in previous literature [46], the benefits do not depend simply on the technology but on how well it is used to promote reflection in the learners. In our experiment, a computer science technician was always present to cover any eventuality. Moreover, the medical instructors overcame their lack of technical knowledge with great willingness and showed interest in exploiting the approach as shown in the responses to Q11 (mean, 4.6).

The students reported that i-SIDRA was useful as regards improving their learning (Q5) with a similar average mean (3.9) to that obtained in three previous studies: 38 medical residents using a CRS at resident conferences (mean, 4.0) [47]; 175 undergraduate science students using a CRS in large-group physiology lectures (mean, 4.0) [15], and 46 family medicine professionals, mainly physicians, using a CRS in one family medicine lecture (mean, 3.8) [16]. The students also expressed their expectation that i-SIDRA would be used on other courses, as confirmed in other studies focused on CRS [48].

6.5. How i-SIDRA provides an insight into learning behavior

In our experiment, better results were obtained in tests on introduction to gross anatomy and musculoskeletal anatomy of the trunk and musculoskeletal anatomy of the head and neck (Test1 and Test4) than in those on musculoskeletal anatomy of the upper and lower limbs (Test2 and Test3). The root of these differences may be the lower complexity of introduction to gross anatomy musculoskeletal anatomy of the trunk, the novelty effect of Test1 and the students' high motivation in Test4 when the deadlines for the exam sessions were approaching.

The data collected were used not only to identify the common mistakes in the groups of student answers, but also to analyze how the feedback influenced the learning of individual students. Each student had recorded a track of his/her progress over time in the database. This allowed the instructors to analyze how that student's answers changed after reading the feedback.

Fig. 3 presents a knowledge state diagram representing the students' behavior in Test4. There are eleven states of knowledge which capture some commonality in the set of questions responses. These states are self-organized into the following layers: Start, Layer 1, Layer 2, Layer 3, and Layer 4. For example, a student in state 3 may go via state 5 in the following layer, before reaching the 'state of perfect knowledge' which represents correct answers to all the questions. The average score for each layer is increased from start state to final state. Average scores are 60.00% for the beginning level (Layer 1), 70.00% for the intermediate level (Layer 2), 70.30% for the advanced level (Layer 3), and 100% for the final state. The knowledge state transitions represent reclassifications into a new (or the same) state, and the students therefore receive new (or the same) feedback. This occurs when students submit a new set of answers, having received some feedback associated with their previous state. Note that the knowledge state 4 does not have exit transitions and instructors can therefore deduce that the current feedback is insufficient to move beyond this state. Another clue that can be used to identify problematic feedback is self-transition (transition in which the target state is the same as the source state). The knowledge states 5, 9 and 11 generated between 11 and 80 triggering events in a self-transition. The average time taken per attempt in these self-transition events was 1.68 minutes, as opposed to 2.05 minutes in the whole of Test4. These values may have resulted from the students' anxiety in the form of multiple submissions when they had only one or two incorrect answers and repeatedly received the same feedback. We also observed that the exit transition of knowledge state 9 towards the 'state of perfect knowledge' included more triggering events than those of knowledge states 5 and 11. This can be attributed to the fact that feedback associated with the knowledge state 9 included 6 pictures whereas the knowledge states 5 and 11 included only 4 pictures. Previous research has shown the educational potential of pictures and movies in the learning of anatomy [49].

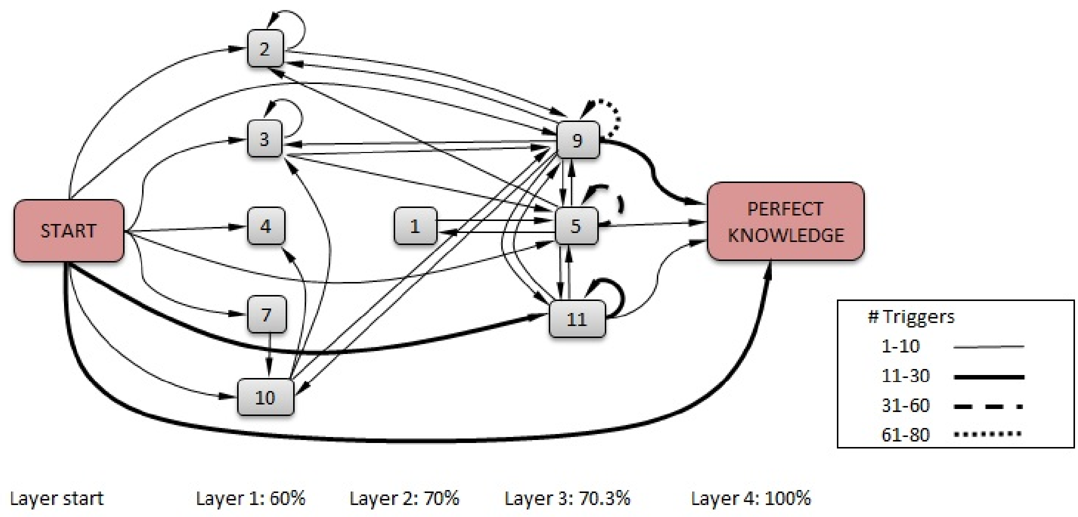

Fig. 3 - Knowledge state transitions in Test4

When globally analyzing the results of the tests in i SIDRA, we found that the questions with the least correct answers in all of the tests were related to areas of special topographic complexity such as the inguinal canal, posterior wall of the axilla or supra or infra-piriform foramen. Despite the great effort made during lectures or in the discussion group, students perceive these regions as very difficult to learn, because of the large number of anatomical structures involved and their complicated spatial organization.

Secondly, we also found mistakes in tests on more conceptual issues, such as the composition of the spinal nerves and their function (perhaps because of their novelty or the fact that they had little to do with topics learned before), or about the actions of the muscles of the leg on the ankle joint. Observe that kinematics is also always difficult to learn. In contrast, the students were able to solve more specific aspects regarding origins and muscle insertions, although these sometimes involved the attainment of a great deal of knowledge about the anatomical language.

6.6. Limitations

One internal threat to validity is the influence of confounders. The effect of the confounding variable related to previous knowledge of GAHMS was mitigated during the students' selection process since all of the students were enrolled in GAHMS, which is a first year course in the first term, for the first time. These students had studied Health Sciences at high school, and had no prior knowledge of anatomy. Moreover, the effect of this and other potential confounding variables (i.e., student motivation, student personality, student tiredness, and so on) were controlled by randomly forming the student groups (control and experimental). Moreover, no subject identified any serious problem with either i-SIDRA or the tests carried out during the experiment, as is shown in the results of the post-experiment questionnaire.

Although not imposing restrictions on the time spent carrying out the experimental tasks can be a valid approach in controlled experiments [50], a time limit was established for each test so as to maintain the ability to reveal differences between the scores at the beginning and at the end of each test.

The number of students involved in the experiment and the higher dropout rate in the control group than in the experimental group may be external threats to validity. On the one hand, empirical research has been successfully conducted with a fewer students [20, 51, 52]. On the other hand, a higher final examination attendance rate in the experimental group than in the control group can be explained by the fact that the experimental group students were more motivated to complete the course, since the active participation of students as regards using i-SIDRA and the formative assessment were an incentive [53, 54].

The external validity of the results might also be threatened by the small size of the experimental objects (four 10-question tests on GAHMS). More extensive tests could overload novice medical students and produce a fatigue effect on students. In spite of this, we are of the opinion that the findings of this study may be of interest to other people outside the case investigated since the skills and knowledge assessed are relatively common in general anatomy.

Internal consistency, criterion validity, construct validity and reliability were not quantified in the questionnaires used in the experiment. To mitigate the threat to this construct validity, a content validation of the questionnaires used in the experiment was carried out by medical professionals. Misunderstandings related to some questions were identified and corrected. Moreover, we have avoided writing negatives or double negatives in the survey questions, since it takes a long time for surveyors to figure out whether they would actually agree or disagree with these questions. However, there is a general tendency towards assent rather than dissent (acquiescence), so the average of all answers may tend to respond on the agree side.

To reduce consistency issues, the same teachers designed the four tests. Moreover, a guideline document describing the experiment design was followed by the teachers with the aim of reproducing the same conditions in the four tests.

7. Conclusions

This paper has shown the effects of a CRS combined with an intelligent analysis of the students' responses on student performance in an anatomy of the locomotor system course. A snap–drift neural network (SDNN) approach was used to obtain an efficient means to discover a relatively small and therefore manageable number of groups of similar answers. These groups were used to gain insights into the students' learning needs and generate appropriate diagnostic feedback. Evidence of the feedback's effectiveness has been provided and the results of the statistical study will be used to improve medical degree students' anatomy of the locomotor system education. The enormous amount of data collected using i-SIDRA provides instructors with a wide range of possibilities during the post hoc analysis [10]. Student responses can also be employed to retrain the neural network which may result in new groups. The feedback from the new groups must be revised by the instructors if new refined groupings are to be created.

With regard to future directions, we intend to adopt an iterative algorithm based on a Bayesian approach to produce a precise indication of course competencies associated with MCQs [19]. A better identification of the diagnostic competencies will allow the instructors to improve the design of the feedback customized to each student. We also plan to integrate i-SIDRA into a course management system such as Moodle or Sakai to facilitate the adoption of this type of thought-provoking environment in other medical disciplines. Recent literature has reported that the delivery of an e-learning course using Moodle increments performance and satisfaction when compared to traditional didactic lectures [55].

Authors' contributions

José Luis Fernández-Alemán, Laura López González and Ofelia González Sequeros contributed to the following: the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article and approval of submitted version. The authors Chrisina Jayne, Ambrosio Toval and Juan José López Jiménez made the following contributions to the study: analysis and interpretation of data, drafting the article and approval of submitted version.

Acknowledgments

This research is part of the project GINSENG (TIN2015-70259-C2-2-R) supported by the Spanish Ministry of Economy and Competitiveness and European FEDER funds.

Conflicts of interests

There are no conflicts of interests.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at

Summary points
What was already known about the topic: • CRS (classroom response system) has been successfully used for different purposes in the medical classroom. • CRS allows instructors to create safe and active learning environments, to offer a fun atmosphere during class and to collect and analyze student feedback quickly (and anonymously) during class. • Research into response-driven feedback for multiple-choice questions (MCQs) based on intelligent analysis is limited.
What does this study adds to our knowledge: • The use of CRS combined with the intelligent analysis of the students' responses has a positive effect on student performance on an anatomy of the locomotor system course. • The diagnostic feedback of i-SIDRA results in a statistically significant increase in medical students' knowledge. • Medical students expressed satisfaction and reported that i-SIDRA was useful as regards improving their learning.

References

[1] J.C. De Gagne, The impact of clickers in nursing education: A review of literature, Nurse Education Today, 31 (2011) e34-e40.
[2] A.B. Sánchez García, M.J. López Montesinos, J.L. Fernández Alemán, Wireless devices in nursing education, Investigación y Educación en Enfermería, 31 (2013) 95-106.
[3] R. Luanrattana, K. Win, J. Fulcher, D. Iverson, Mobile Technology Use in Medical Education, J Med Syst, 36 (2012) 113-122.

[4] K. Moss, M. Crowley, Effective learning in science: The use of personal response systems with a wide range of audiences, Computers & Education, 56 (2011) 36-43.
[5] J.E. Caldwell, Clickers in the Large Classroom: Current Research and Best-Practice Tips, CBE—Life Sciences Education, 6 (2007) 9-20.
[6] J.R. MacArthur, L.L. Jones, A review of literature reports of clickers applicable to college chemistry classrooms, Chemistry Education Research and Practice, 9 (2008) 187-195.
[7] T.E. Schackow, M. Chavez, L. Loya, M. Friedman, Audience Response System: Effect on Learning in Family Medicine Residents, Family Medicine, 36 (2004) 496-504.
[8] A. Nájera, J.M. Villalba, E. Arribas, Student peer evaluation using a remote response system, Medical Education, 44 (2010) 1146-1146.
[9] M. J Hashim, Standard setting using an audience response system with 'clickers', Medical Education, 47 (2013) 530-530.
[10] P. Schick, S. Abramson, J. Burke, Audience response technology: under-appreciated value of post hoc analysis, Medical Education, 45 (2011) 1157-1158.
[11] J.M. Garbutt, T.M. DeFer, G. Highstein, C. McNaughton, P. Milligan, V.F. Fraser, Safe Prescribing: An Educational Intervention For Medical Students, Teaching and Learning in Medicine, 18 (2006) 244-250.
[12] J.W. Turban, The audience response system: a modality for course evaluation, Medical Education, 43 (2009) 488-489.
[13] A. Pradhan, D. Sparano, C.V. Ananth, The influence of an audience response system on knowledge retention: An application to resident education, American Journal of Obstetrics and Gynecology, 193 (2005) 1827-1830.
[14] P. Bhargava, A.E. Lackey, S. Dhand, M. Moshiri, K. Jambhekar, T. Pandey, Radiology Education 2.0—On the Cusp of Change: Part 1. Tablet Computers, Online Curriculums, Remote Meeting Tools and Audience Response Systems, Academic Radiology, 20 (2013) 364-372.
[15] S.A. Gauci, A.M. Dantas, D.A. Williams, R.E. Kemm, Promoting student-centered active learning in lectures with a personal response system, Advances in physiology education, 33 (2009) 60-71.
[16] R. Latessa, D. Mouw, Use of an audience response system to augment interactive learning, Fam Med, 37 (2005) 12-14.
[17] J. Antoun, R. Nasr, N.K. Zgheib, Use of technology in the readiness assurance process of team based learning: Paper, automated response system, or computer based testing, Computers in Human Behavior, 46 (2015) 38-44.
[18] R.M. Harden, Student feedback from MCQ examinations, Medical Education, 9 (1975) 102-105.
[19] S. Manning, A. Dix, Identifying students' mathematical skills from a multiple-choice diagnostic test using an iterative technique to minimise false positives, Computers & Education, 51 (2008) 1154-1171.
[20] J.L. Fernández-Aleman, D. Palmer-Brown, C. Jayne, Effects of Response-Driven Feedback in Computer Science Learning, IEEE Transactions on Education, 54 (2011) 501-508.
[21] J.L. Fernández-Alemán, C. Jayne, A.B. Sanchez, J.M. Carrillo-de-Gea, A. Toval, Neural network-based data analysis for medical-surgical nursing learning, in: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 2012, pp. 6036-6039.
[22] J.L. Fernández-Alemán, C. Jayne, A. García, J. Carrillo-de-Gea, A. Toval Alvarez, Knowledge Clustering Using a Neural Network in a Course on Medical-Surgical Nursing, in: C. Jayne, S. Yue, L. Iliadis (Eds.) Engineering Applications of Neural Networks, Springer Berlin Heidelberg, 2012, pp. 385-394.
[23] R. Guo, D. Palmer-Brown, S.W. Lee, F.F. Cai, Intelligent diagnostic feedback for online multiple-choice questions, Artificial Intelligence Review, 42 (2014) 369-383.

[24] F. Mödritscher, e-Learning theories in practice: a comparison of three methods, Journal of Universal Science and Technology of Learning, (2006) 3-18.
[25] M.J. Atkins, Theories of learning and multimedia applications: An overview, Research Papers in Education, 8 (1993) 251-271.
[26] G. McLeod, Learning Theory and Instructional Design, Learning Matters, 2 (2003) 35-43.
[27] B. Holmes, J. Gardner, E-Learning: Concepts and Practice, Sage, London, 2006.
[28] J.G. Ruiz, M.J. Mintzer, R.M. Leipzig, The impact of e-learning in medical education, Academic Medicine, 81 (2006) 207-212.
[29] C.-S. Ong, J.-Y. Lai, Gender differences in perceptions and relationships among dominants of e-learning acceptance, Computers in Human Behavior, 22 (2006) 816-829.
[30] L. Zhang, H. Wen, D. Li, Z. Fu, S. Cui, E-learning adoption intention and its key influence factors based on innovation adoption theory, Mathematical and Computer Modelling, 51 (2010) 1428-1432.
[31] A.W. Astin, Student involvement: A developmental theory for higher education, Journal of College Student Personnel, 25 (1984) 297-308.
[32] L. Siragusa, K.C. Dixon, Closing the gap between pedagogical theory and online instructional design: a bridge too far?, in: G. Chiazzese, M. Allegra, A. hifari, S. Ottaviano (Eds.) Methods and Technologies for Learning, WIT Press, Southampton, 2005.
[33] A. Aljamal, H. Cader, C. Chiemeke, M. Speece, Empirical assessment of e-learning on performance in principles of economics, International Review of Economics Education, 18 (2015) 37-48.
[34] R.L. Drake, W. Vogl, A.W.M. Mitchell, Grays's Anatomy for students, Elsevier Inc, Philadelphia: Churchil Livinstone, 2010.
[35] R. Guo, D. Palmer-Brown, S.W. Lee, F.F. Cai, Intelligent diagnostic feedback for online multiple-choice questions, Artif Intell Rev, 42 (2014) 369-383.
[36] T.M. Haladyna, Developing and Validating Multiple-Choice test items, Lawrence Erlbaum Associates Publishers, 2004.
[37] J.L. Fernández-Alemán, J.M. Carrillo-de-Gea, J.J. Rodríguez-Mondéjar, Effects of competitive computer-assisted learning versus conventional teaching methods on the acquisition and retention of knowledge in medical surgical nursing students, Nurse Educ Today, 31 (2011) 866-871.
[38] L. Lutze-Mann, R.K. Kumar, The formative assessment lecture: enhancing student engagement, Medical Education, 47 (2013) 526-527.
[39] P. Duggan, E. Palmer, P. Devitt, Electronic voting to encourage interactive lectures: a randomised trial, BMC Medical Education, 7 (2007) 25.
[40] A.R. Tait, T. Voepel-Lewis, S.J. Chetcuti, C. Brennan-Martinez, R. Levine, Enhancing patient understanding of medical procedures: Evaluation of an interactive multimedia program with in-line exercises, International Journal of Medical Informatics, 83 (2014) 376-384.
[41] J.C.-Y. Sun, Influence of polling technologies on student engagement: An analysis of student motivation, academic performance, and brainwave data, Computers & Education, 72 (2014) 80-89.
[42] J.K. Knight, W.B. Wood, Teaching more by lecturing less, Cell Biol Educ, 4 (2005) 298-310.
[43] J. O'Flaherty, C. Phillips, The use of flipped classrooms in higher education: A scoping review, The Internet and Higher Education, 25 (2015) 85-95.
[44] D.J. Skiba, Got large lecture hall classes? Use clickers, Nurs Educ Perspect, 27 (2006) 278-280.
[45] J.H. Han, A. Finkelstein, Understanding the effects of professors' pedagogical development with Clicker Assessment and Feedback technologies and the impact on students' engagement and learning in higher education, Computers & Education, 65 (2013) 64-76.

[46] S.W. Draper, M.I. Brown, Increasing interactivity in lectures using an electronic voting system, Journal of Computer Assisted Learning, 20 (2004) 81-94.
[47] J.W. Kung, P.J. Slanetz, P.-H. Chen, K.S. Lee, K. Donohoe, R.L. Eisenberg, Resident and Attending Physician Attitudes Regarding an Audience Response System, Journal of the American College of Radiology, 9 (2012) 828-831.
[48] D.A. Smith, M.M. Rosenkoetter, Effectiveness, challenges, and perceptions of classroom participation systems, Nurse Educ, 34 (2009) 156-161.
[49] C. Adamczyk, M. Holzer, R. Putz, M.R. Fischer, Student learning preferences and the impact of a multimedia learning tool in the dissection course at the University of Munich, Annals of Anatomy - Anatomischer Anzeiger, 191 (2009) 339-348.
[50] J.L. Fernández-Alemán, J.M. Carrillo-de-Gea, J. Vidal Meca, J. Nicolas Ros, A. Toval, A. Idri, Effects of Using Requirements Catalogs on Effectiveness and Productivity of Requirements Specification in a Software Project Management Course, IEEE Transactions on Education, 59 (2016) 105-118.
[51] J.L. Fernández-Alemán, A.B. Sánchez García, M.J. López Montesinos, J.J. López Jiménez, Examining the Benefits of Learning Based on an Audience Response System When Confronting Emergency Situations, Computers Informatics Nursing, 32 (2014) 207-213.
[52] J.L. Fernández-Alemán, Automated Assessment in a Programming Tools Course, IEEE Transactions on Education, 54 (2011) 576-581.
[53] J. Juanes, P. Ruisoto, Computer Applications in Health Science Education, J Med Syst, 39 (2015) 1-5.
[54] T. Gazibara, V. Marusic, G. Maric, M. Zaric, I. Vujcic, D. Kisic-Tepavcevic, J. Maksimovic, N. Maksimovic, L. Denic, S. Grujicic, T. Pekmezovic, A. Grgurevic, Introducing E-learning in Epidemiology Course for Undergraduate Medical Students at the Faculty of Medicine, University of Belgrade: A Pilot Study, J Med Syst, 39 (2015) 1-7.
[55] L.O. Reis, O. Ikari, K.A. Taha-Neto, A. Gugliotta, F. Denardi, Delivery of a urology online course using moodle versus didactic lectures methods, International Journal of Medical Informatics, 84 (2015) 149-154.