COMPUTER-AIDED TESTING: ASSESSMENT OF AUTOMATIC ITEM GENERATION TO CREATE MULTIPLE CHOICE TEST ITEMS

M. Panczyk, J. Belowska, J. Gotlib

Division of Teaching and Outcomes of Education, Faculty of Health Sciences, Medical University of Warsaw (POLAND)

Abstract

Introduction:

Computer-aided testing is an alternative to traditional paper-and-pen tests. Despite the fact that computer-based examination methods are well-established in academic centres, continuous monitoring of their quality is still needed.

In the academic year 2013/14, doctoral degree students of the Medical Faculty at Medical University of Warsaw (MUW) took part in the "*Reliability in research*" course for the first time conducted in the form of blended-learning. To be awarded a credit for the course, the students were supposed to pass a final test of the e-learning course that was made available on the Moodle platform.

Aim of study:

Evaluation of usefulness of automatically generated computer-interactive multiple choice test for assessment of achievements of blended-learning students.

Materials and Methods:

A total of 96 PhD students, including 45 first year (25 physicians) and 51 second year (24 physicians) students of a doctoral degree course. It was assumed that both groups of students are comparable and may constitute control groups for each other. The e-learning test results comprised a total of 43 multiple-choice questions (four options to choose from) and were subgrouped into the following categories: (1) *Ethical aspects of scientific unreliability*, (2) *Scientific misconduct*, (3) *Copyright and research activity*, (4) *Conflict of interest in research*, (5) *Rules of "Good Research Practice"*.

A test set was generated individually for each student out of all questions from the database. The easiness of particular versions and the frequency of using particular questions from the entire pool of questions as well as within the thematic areas were compared to assess the quality of particular question sets. The significance of differences in results was assessed and the mean time necessary for completing the test in both groups was evaluated. Non-parametric Mann-Whitney U test was used for analysis. For all analyses, the a priori level of significance was 0.05.

Results:

The questions included in automatically generated tests reflected the proportion of questions within the thematic subgroups. Deviation from representativity within the fields was not larger than 1.5% and frequency of using test questions ranged between 1.35 and 3.13% (mean: 2.33% ± 0.45). Total test easiness was high and amounted to 0.854 (0.755 – 1.000), and both groups of students did not differ significantly with respect to this ($P > 0.05$). No significant differences were also found with respect to the time spent on completing the test by first and second year students (460.9 s ± 124.260 versus 436.9 s ± 135.974, $P > 0.05$).

Conclusions:

Reliable computer-based examination methods are used to meet the requirement of uniform rules and criteria of assessment of students' achievements. Automatic generation of question sets with the use of the Moodle platform tools may ensure fair and unbiased assessment of educational progress. The quality of computer-aided testing is comparable to that of traditional paper-and-pen test with reference to assessing the achievement of selected outcomes of education.

Keywords: educational measurement, computer-aided testing, multiple-choice questions, e-learning.

1 INTRODUCTION

Apart from the traditional methods of teaching, such as lectures or seminars, computer-assisted learning is beginning to play a more and more important role. It is a frequently used form of education and such teaching is an integral part of educational system also at university level. Besides benefits resulting from, e.g. lowering the costs of e-classes in comparison with the traditional ones, an increase in students' knowledge and understanding of a particular subject and allowing students to direct their own learning [1] is an important argument that supports the use of modern forms of teaching. Apart from the obvious benefits resulting from using computer-aided teaching an appropriate question arises, namely is using automatic verification techniques an efficient form of evaluating a student's progress?

Multiple-choice questions (MCQs) test is a commonly used and accepted way of a student's progress verification in academia. Tests are widely used in various areas of medical education – from general sciences through basic ones and to clinical ones. Moreover, measuring methods of educating in skills with the use of MCQs may concern very narrow and specialised areas of knowledge, and may also include several fields of medicine as it is the case in professional examinations, such the United States Medical Licensing Examination (USMLE) or National Council Licensure Examination for Registered Nurses (NCLEX-RN). Due to such popularity of this form of examining, there were numerous guidelines for constructing MSQs prepared that concentrated on good practices [2-4]. Also, many universities and schools have their own inner rules and regulations concerning examining with the use of tests constructed in a manner similar to MCQs.

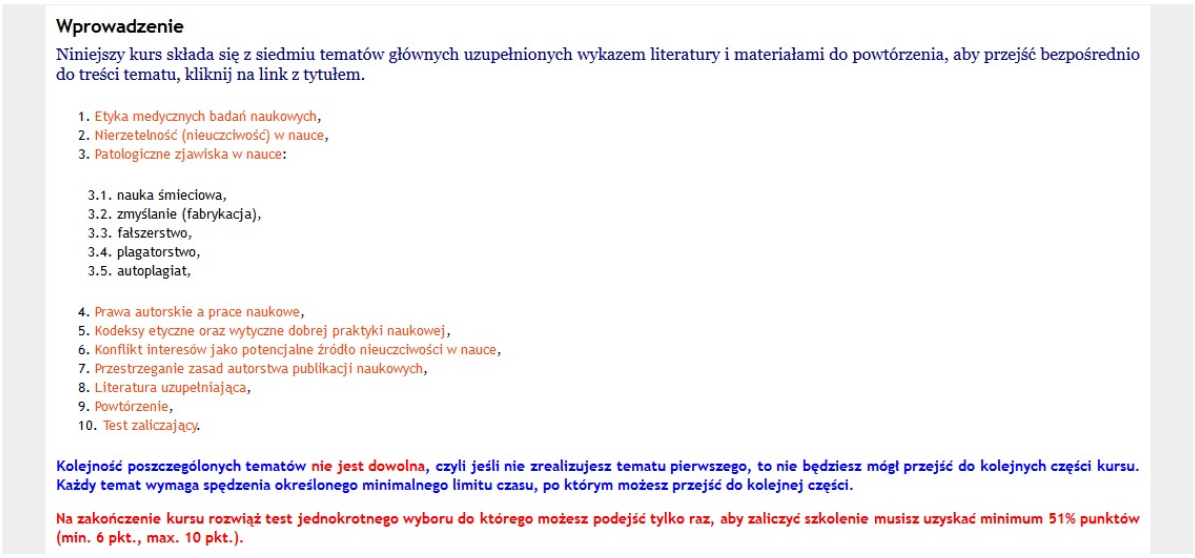
The increase in the importance of tests in educational measurement caused the increase of interest in the computer-aided techniques that help evaluate students' achievements. If in the case of small groups, it is possible to use traditional paper-and-pen tests (e.g. a 50-item ones or end-of-unit tests), then in a situation when a final exam includes hundreds or sometimes even thousands of examinees, it becomes necessary to introduce a more automatic method of testing. Computer-aided testing (e-test) is one form of examining that allows to generate testing sets that would meet the expected parametres automatically. Its main disadvantage, however, is the necessity to prepare a vast pool of questions, the so-called bank of testing questions. While generating individual variations of a test, questions drawn from the pool must be equivalent in quality in every version of the test so that the obtained results would be reliable. Maintaining domain cohesion also requires designing appropriate sub-tests prepared on the basis of assumptions of a planned test. Subtests must have defined knowledge and skills domains in their range. Apart from mere generation of a set of questions, computer-aided testing can also automatically create computerised adaptive tests (CATs) [5]. However, this form of testing has one important cost: CATs require large banks containing thousands of multiple-choice items. The use of large banks permits continuous testing while minimising item exposure so that test security can be maintained [6].

Dissemination of computer-aided tests was accompanied by intensive development of teaching that would include various e-learning tools, such as computer assisted learning [7] or computer-aided learning [8]. As early as in the mid-60s there appeared the first computer based techniques in teaching methodology [9] and then gradually, from mid-80s, various forms of computer-based learning have been introduced to universities [10].

A great progress observed in the last decade in the development of IT technologies can also be reflected in the methods used in educational measurement. Advanced systems designed to be used in distance learning can meet high requirements that are placed before methods used in competence evaluation, also on-line [11, 12]. In addition to the development of new teaching approaches, utilization of computer technology and multiple-choice test formats has significantly altered evaluation practices in medical schools. Therefore MCQ testing, the most frequently used method in education, is equally often used in e-learning [13 – 15].

In the academic year of 2013/14, doctoral degree students of the Medical Faculty at Medical University of Warsaw (MUW) took part in the "*Reliability in research*" course for the first time conducted in the form of blended-learning. As part of "in-class" teaching, during the five-hour introductory lecture, students had a chance to become familiar with issues concerning copyrights and rules of good practices in the field of scientific research. The remaining 25 hours of self-study were carried out in form of computer-aided learning as a multi-module e-learning course accessible at the Moodle educational platform (http://www.nzd.moodle.wum.edu.pl/course/index.php?categoryid=59) (Fig. 1). Moreover, students could contact the lecturers on-line both in synchronous mode (Skype, chat) and asynchronous one (e-mail, forum). The condition under which they would pass the course,

was a positive evaluation of the course conducted in form of a test comprising of 10 MCQs test selected randomly from the pool of questions. In order to ensure a domain representativeness, questions were divided into 5 thematic subtests that were the equivalents of consecutive modules in e-learning course.



Wprowadzenie

Niniejszy kurs składa się z siedmiu tematów głównych uzupełnionych wykazem literatury i materiałami do powtórzenia, aby przejść bezpośrednio do treści tematu, kliknij na link z tytułem.

1. Etyka medycznych badań naukowych,
2. Nierzetelność (nieuczciwość) w nauce,
3. Patologiczne zjawiska w nauce:

   3.1. nauka śmieciowa,
   3.2. zmyślanie (fabrykacja),
   3.3. fałszerstwo,
   3.4. plagatorstwo,
   3.5. autoplagiat,

4. Prawa autorskie a prace naukowe,
5. Kodeksy etyczne oraz wytyczne dobrej praktyki naukowej,
6. Konflikt interesów jako potencjalne źródło nieuczciwości w nauce,
7. Przestrzeganie zasad autorstwa publikacji naukowych,
8. Literatura uzupełniająca,
9. Powtórzenie,
10. Test zaliczający.

Kolejność poszczególnych tematów nie jest dowolna, czyli jeśli nie zrealizujesz tematu pierwszego, to nie będziesz mógł przejść do kolejnych części kursu. Każdy temat wymaga spędzenia określonego minimalnego limitu czasu, po którym możesz przejść do kolejnej części.

Na zakończenie kursu rozwiąż test jednokrotnego wyboru do którego możesz podejść tylko raz, aby zaliczyć szkolenie musisz uzyskać minimum 51% punktów (min. 6 pkt., max. 10 pkt.).

Figure 1. Moodle platform: a view of a module that introduces a participant into the subject of e-learning course together with explanations of passing conditions for the subject of "*Reliability in research*".

Considering the advantages and disadvantages of automatic tools that check the students' achievements as well as the possibilities of implementation of such solutions in virtual environment, the quality of functioning of the integrated distant learning systems should constantly be under control. That is why, within the evaluation of educational quality of students in the subject of "*Reliability in research*", assessment of automatically generated computer-interactive multiple choice test that checked the achievements of students taught with blended-learning method was performed.

1.1 Materials and Methods

The resulting data used for analysis were obtained from computer generated test performed on a group on 96 PhDs, including 45 (25 GPs) from the first and 51 (24 GPs) from the sophomore year of the third degree studies at MUW. It was assumed that both groups of students are comparable and may become a control sample. The obtained results as pending e-learning completion included 43 four-option MCQs and were sub-grouped into the following categories: (1) *Ethical aspects of scientific unreliability*, (2) *Scientific misconduct*, (3) *Copyright and research activity*, (4) *Conflict of interest in research*, (5) *Rules of "Good Research Practice"*. Test set was generated individually for each student from the pool of questions collected in the database. In order to carry out the on-line test, a *Quiz* tool was used that was accessible as an e-learning Moodle platform module (Modular Object-Oriented Dynamic Learning Environment) which helps to support distant learning (LCMS, Learning Content Management System).

Educational Moodle platform is one of the most popular tools supporting distant learning (>49 000 registered sites in >200 countries), that allows relatively easily to create and manage e-learning courses. One of the most important components of this platform are modules that check students' progress, i.e. *Quiz*, *Lesson* and *Exercise*. Module *Quiz* makes it possible to create, conduct and assess students using on-line tests built, e.g. on the basis of MCQs. Questions that are created are later stored in the database and may be used on numerous occasions within the framework of a course. Questions in the pool are categorised and the module enables automatic generation of a test set, where individual MCQs can be randomly placed items. Checking of students' answers takes place immediately after they have done the test. Thanks to this, a student may immediately obtain information about a grade as well as detailed data concerning mistakes he or she has made (the so-called feedback). Moreover, a time limit may be set for each test as well as a range of MCQs that were

used previously. Also the available security options are important, such as additional password, required net address (a unique IP of a computer), and the date and hour of accessing the test [16].

In order to evaluate the quality of individual sets of questions, simplicity of individual variations, frequency of the used questions in the whole pool and thematic subgroups were compared. Relevance of differences was assessed in the obtained results as well as mean time needed to do the test in both groups of students. A non-parametric Mann-Whiney U test was used in comparisons. For all analyses, the a priori level of significance was 0.05.

2 RESULTS

Equinumerosity analysis shows that a slight difference in the number of students between the groups should not impact the possible statistical relevance of the observed differences (equipotence, chi-squared test, *P*> 0.05).

The overall result in points obtained by the examinees was similar in both groups (Mann-Whitney U test, *P*> 0.05). Also, parametres of spreading results, such as standard deviation and interquartile range were similar in both groups of students (Levene test for equality of variances, *P*> 0.05) (Fig. 2).

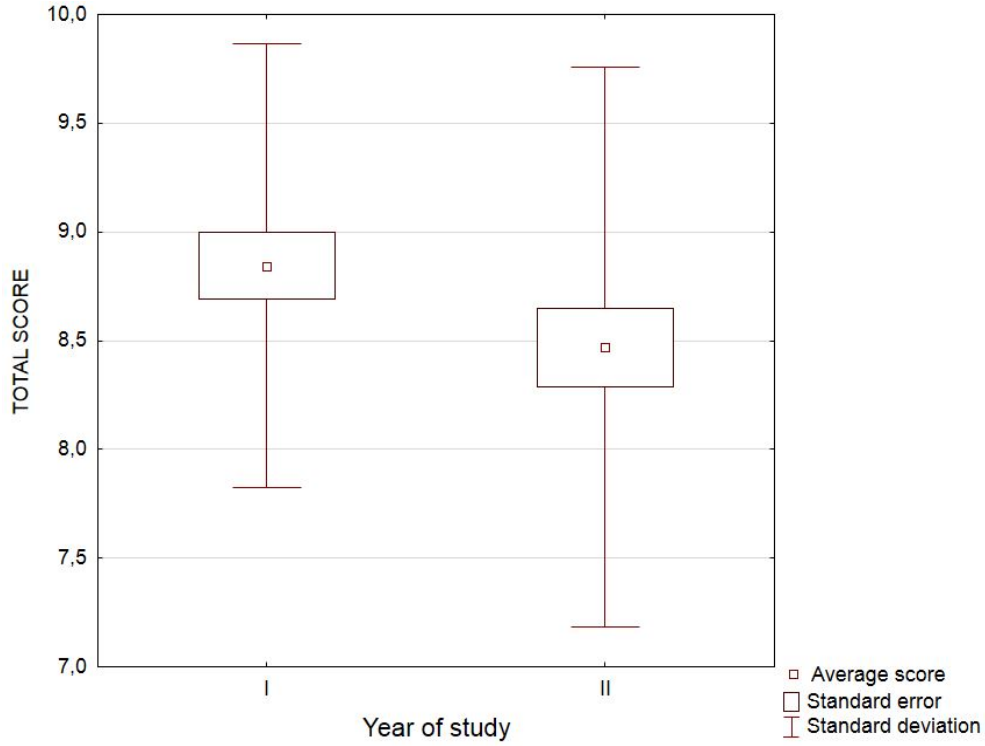


Figure 2. Box-and-Whisker Plots comparing students' results of tests in the subject of "*Reliability in research*".

Frequency distribution analysis of the test results shows that in case of students of the first year a smaller left-sided skewness is obtained as well as less marked leptokurticity than it is the case with the students of the sophomore year.

There were no significant differences in mean time spent by students of the first and second year on solving the test (460,9 s ± 124,260 *versus* 436,9 s ± 135,974, median comparison U Mann-Whitney test, *P*> 0.05). Detailed juxtaposition of the test results for both groups of students are presented in Table 1.

Table 1. Results of test in the subject of "*Reliability in research*" carried out in two groups of PhDs.

	PhD students of the 1st year	**PhD students of the 2nd year**
*N**	45	51
*Averagescore (±SD**)*	8.8 (±1.021)	8.5 (±1.286)
*Median score***	9.0	9.0
Interquartile range ($Q_{25\%}$ - $Q_{75\%}$)	8.0 – 10.0	8.0 – 9.0
Coefficient of variation	11.5%	15.2%
Skewness	–0.879	–1.037
Kurtosis	0.894	2.492
*Average time of the test (*seconds*)*	460.9 (±124.260)	436.9 (±135.974)
*Median time of the test (*seconds*)****	495.0	467.0

* chi-squared test, equipotence (*P*> 0.05);

** Levene test for equality of variances (*P*> 0.05);

*** Mann-Whitney U test (*P*> 0.05)

Automatically generated variants of e-tests reflected proportions of individual test items included in thematic sub-groups. Deviation from representativeness in domains were no greater than 1.5% and frequency of using test questions remained within the range of 1.35 and 3.13% (2.33% ± 0.45 on average). Overall simplicity of an e-test was high and remained at 0.854 (value for subtests from 0.755 to 1.000). Both groups of students were similar in the level of simplicity of the questions, both in the whole pool of test questions and in individual thematic groups (Mann–Whitney U *test, P*> 0.05). A detailed characteristic of the used base of questions together with the mean values of the index of test difficulty were presented in Table 2.

Table 2. Characteristic of the used base of questions in order to generate variants of e-tests in the subject of "*Reliability in research*".

Subtest	**Ethical aspects of scientific unreliability**	**Scientific misconduct**	**Copyright and research activity**	**Conflict of interest in research**	**Good Research Practice**
Number of questions in the database	10	17	10	3	3
Average share of the generated test	23%	39%	23%	8%	7%
Index of test difficulty (total)	0.769	0.911	0.835	0.755	1.000
Index of test difficulty (1st year)	0.838	0.928	0.881	0.718	1.000
Index of test difficulty (2nd year)	0.714	0.932	0.822	0.821	1.000

2 DISCUSSION

Tests performed in a conventional pen-and-paper form which need preparation, evaluation and result analysis can occasionally be time-consuming. The opportunities to create a test item bank and to obtain test results immediately after the examination offer advantages over conventional test methods. Moreover, in case of traditional teaching with no computer tools used in distant learning, self-control and self-evaluation of progress through independent work on exercises or tests done by a student at

home is not effective enough and sometimes even impossible to perform. The above flaws may be successfully eliminated by enriching educational classes with e-learning elements, an excellent example of which are various forms of blended-learning that apply for example computer-based examination method. For this purpose, many institutions are now adopting open-source applications (such as Moodle), which carry benefits in terms of cost and functionality [17]. In accessible literature, there are plenty of examples which make use of Moodle in teaching pharmacology [18, 19], surgery [20], radiology [21], dermatology [22] or emergency medicine [23].

Introducing e-learning courses using the Moodle platform is generally welcomed by students [24-26] and complementing traditional classes using elements of distant learning positively reinforces the efficiency of both teaching and learning; it also is connected with receiving higher final grades [27]. Observations prove that using blended teaching and learning techniques can promote an active, deeper approach for learning that enhances student learning outcomes [28, 29]. Moreover, results of several correlation analyses show that teaching using blended techniques and e-courses favours better passing rate in students during their final exams performed using paper-and-pen method [28-30]. Obviously, the above observations only confirm the fact that good students generally handle e-learning courses better and they also do better in final exams than students who are average or weaker, which is reflected in positive correlations. However, it is not yet clear whether high activity of weaker students in e-courses impacts positively the final exam results they achieve [29, 31, 32]. Generally, the above observations allow to assume that using tools offered by the Moodle platform may have a positive influence on the quality of teaching medical students.

The following can be included into the properties of educational measurement that allow its evaluation and optimisation: a) lack of bias, b) precision in scoring, c) reliability, d) validity, e) objectivity [33]. These properties are listed hierarchically together with the increase in their complexity and lowering observability. Assessment of impact of various types of cultural, personal and organisational factors on the results obtained by the examinees were omitted in the presented research. Also the assessment of scoring precision was excluded, recognising the fact that it is of little importance in case of using MCQ format for which zero-one score is applied. Additionally, considering the assessment of objectivity, a principle was assumed that the areas of skills and knowledge represented in e-tests are in accordance with the accepted curriculum assumptions for the subject of "*Reliability in research*", which is reflected in the applied test content outlines.

Because the used MCQs have not been standardised and normalised, a problem of repetitiveness and comparability arises for the students' marks in consecutive editions. Differences not relevant statistically in the results obtained by the first and second year PhDs prove that the used tool of educational measurement provides a reliable and repetitive assessment of competences in students who complete the e-course. Moreover, slight deviations from the representativeness of individual MCQs that create the computer generated variants of e-tests allow to assume that the random algorithm created for the *Quiz* module fulfilled its function. The obtained results of the MCQ e-tests analysis show that the used testing tool is appropriate when evaluating the progress of students taught within the frameworks of the subject "*Reliability in research*" using a blended –learning method.

The *Quiz* module provides a possibility of analysis of individual MCQs considering their quality parametres, such as range of difficulty or differentiating ability [16], which is very important in the light of evaluation of an exam question pool and improvement of MCQs for the needs of consecutive e-learning course editions. On the basis of the above possibilities of the Moodle platform, it was determined that the quality parametres of questions that comprise the MCQs database did not differ significantly between comparable groups of PhDs. It needs to be underlined that using good quality MCQs when structuring e-tests may significantly influence the results a student will achieve in the final test if such is a prescribed form of evaluation for a given subject [30, 34].

Assessment in the subject of "*Reliability in research*" is carried through only on the basis of results of the final test (the so-called summing-up grade). It may well be assumed that using the division suggested by Susan Brookhart [35], the only measure of the achieved results of teaching in the subject of "*Reliability in research*" is fulfilling the requirements of the final test ("evaluation based on the requirements") excluding a student's input during the process of learning ("evaluation based on input"). Applying the above measurement remains in accordance with the concept of analytical evaluation which is a three-stage process comprising of the following: specifying performance requirements (these are stipulated in education standards), determining requirements (manner of concluding on the basis of the obtained measurement results, e.g. using a test) and creating a scale of student's achievements (operationalization of the scale of grades using requirements and norms). When referring to passing the subject of "*Reliability in research*" it is the norm of performance

standards that applies, because there was no standardisation of a test on a selected sample so as to establish empirical standard. As far as the results of an individual MCQ test are concerned, the most important, from practical point of view, is to determine quantity standard, i.e. the lowest score gained, which would allow to conclude whether a student fulfils the requirements for a given level [36].

Quantity standard assumed *a priori* for the subject of "*Reliability in research*" has undergone certain corrections *a posteriori* after obtaining the test results. Quantitative analysis of the obtained results of an e-test allowed to establish a cut-off point of passed/failed at the level of 51%. For both groups of PhDs, their passing rate was 100% using this criterion of evaluation. As Seluakumaran et al. (2011) [29] stress, overall evaluation of students' progress should also include their input into the learning process during the course time. As the authors note, having a well-constructed e-test allows to include its results into the score obtained by the students in their final exam, thanks to which the final mark will reflect the overall progress a student made during the course of learning [29]. It was noted that introducing additional forms of encouragement such as credit points that would increase the weight of a final mark achieved in the exam, had a positive influence on students' activity during e-learning courses [30]. Since, in case of the subject of "*Reliability in research*" curriculum did not include a final exam, introducing an e-test as well as additional score for a student's activity during the course of learning is worth considering, e.g. in form of using interactive checking tools, such as *Lesson* and *Exercise,* and social network activity*: Wiki* and *Workshop*. Presently, efforts are underway to develop a more comprehensive on-line quiz section and to include it as part of a formative assessment of "*Reliability in research*" course.

3 CONCLUSIONS

Necessity to fulfil a requirement of uniform rules and evaluation criteria of students requires using a tried and tested computer-based examination method. Test sets that are automatically generated using tools accessible on the Moodle platform may ensure a fair and objective evaluation of educational progress. Reduction in teaching time required to administer and score examinations is a major advantage of computer-aided testing compared with written tests. Taking the above into consideration, it may be assumed that an MCQs e-test is a good quality tool that checks whether a student achieved the assumed educational level, a tool comparable to traditional pen-and-paper test at the same time devoid of certain flaws of conventional written tests.

REFERENCES

[1] Feng, J.Y., et al. (2013). Systematic review of effectiveness of situated e-learning on medical and nursing education. Worldviews Evid Based Nurs 10(3),pp. 174-83.

[2] Haladyna, T.M. and M.C. Rodriguez. (2013). Developing and validating test items. New York: Taylor & Francis.

[3] Case, S.M. and D.B. Swanson. (2002). Constructing Written Test Questions For the Basic and Clinical Sciences. 3rd ed. National Board of Medical Examiners.

[4] Boland, R.J., N.A. Lester, and E. Williams. (2010). Writing Multiple-Choice Questions. Acad Psychiatry 34(4),pp. 310-316.

[5] van der Linden, W.J. and C.A. Glas. (2010). Elements of adaptive testing. New York: Springer.

[6] Breithaupt, K., A.A. Ariel, and D.R. Hare. (2010). Assembling an inventory of multistage adaptive testing systems, in Elements of adaptive testing, W. van der Linden and C. Glas, Editors. New York: Springer, pp. 247-266.

[7] Baig, S. (2010). Computer assisted learning programs (CALP) versus conventional teacher centered methods for undergraduate medical education. J Coll Physicians Surg Pak 20(5),pp. 352.

[8] Seabra, D., et al. (2004). Computer aided learning versus standard lecture for undergraduate education in urology. J Urol 171(3),pp. 1220-2.

[9] Bitzer, M. (1966). Clinical nursing instruction via the PLATO simulated laboratory. Nurs Res 15(2),pp. 144-50.

[10] Newbern, V.B. (1985). Computer literacy in nursing education. An overview. NursClin North Am 20(3),pp. 549-56.

[11] Gierl, M.J., H. Lai, and S.R. Turner. (2012). Using automatic item generation to create multiple-choice test items. Med Edu 46(8),pp. 757-65.

[12] Gierl, M.J. and H. Lai. (2012). Using Weak and Strong Theory to Create Item Models for Automatic Item Generation, in Automatic Item Generation: Theory and Practice, M. Gierl and T. Haladyna, Editors. 2012, Routledge: New York,pp. 47-63.

[13] Blaum, W.E., et al. (2013). Towards Web 3.0: taxonomies and ontologies for medical education -- a systematic review. GMS Z Med Ausbild 30(1),pp. Doc13.

[14] Han, H., D.S. Resch, and R.A. Kovach. (2013). Educational technology in medical education. Teach Learn Med 25 Suppl 1,pp. S39-43.

[15] Bains, M., et al. (2011). Effectiveness and acceptability of face-to-face, blended and e-learning: a randomised trial of orthodontic undergraduates. Eur J Dent Educ 15(2),pp. 110-7.

[16] Cole, J. and H. Foster. (2007). Using Moodle: Teaching with the popular open source course management system. Birmingham: O'Reilly Media, Inc.

[17] Uribe-Tirado, A., L.-M. Melgar-Estrada, and J.-A. Bornacelly-Castro. (2007). Moodle learning management system as a tool for information, documentation, and knowledge management by research groups. Profesional de la Información 16(5),pp. 468-474.

[18] Alegret, M., et al. Evaluation of pharmacology competencies through Moodle's questionnaire tool: implementing ongoing evaluation. in METHODS AND FINDINGS IN EXPERIMENTAL AND CLINICAL PHARMACOLOGY. 2008. PROUS SCIENCE, SA PO BOX 540, PROVENZA 388, 08025 BARCELONA, SPAIN.

[19] Pineda, J. Implementation of the virtual platform Moodle in a clinical pharmacology course for medical students. in METHODS AND FINDINGS IN EXPERIMENTAL AND CLINICAL PHARMACOLOGY. 2008. PROUS SCIENCE, SA PO BOX 540, PROVENZA 388, 08025 BARCELONA, SPAIN.

[20] Garcia Urena, M.A., et al. (2009).Application of new technologies to the teaching of surgery in the school of medicine. Cir Esp 85(3),pp. 165-70.

[21] Sparacia, G., et al. (2007). Initial experiences in radiology e-learning. Radiographics 27(2),pp. 573-81.

[22] Ludert, T., et al. (2008). E-learning in the dermatological education at the Charite: evaluation of the last three years. J Dtsch Dermatol Ges 6(6),pp. 467-72.

[23] Shah, I.M., M.R. Walters, and J.H. McKillop. (2008). Acute medicine teaching in an undergraduate medical curriculum: a blended learning approach. Emerg Med J 25(6),pp. 354-7.

[24] Gotlib, J. and M. Panczyk. The use of blended-learning in improving the quality of education in the subject "Methodology of Research" in the Faculty of Health Science in Medical University of Warsaw in Poland - a pilot study. 5th International Conference of Education, Research and Innovation (Iceri 2012), 2012,pp. 5330-5338.

[25] Gotlib, J. (2012). Opinie studentów pielęgniarstwa na temat realizacji zajęć dydaktycznych z przedmiotu „Prawo w ochronie zdrowia” w formie kursu e-learningowego. Zdr Publ122(4), pp. 419-425.

[26] Gotlib, J. and M. Panczyk. (2013). Kształcenie przez Internet w Warszawskim Uniwersytecie Medycznym – doświadczenia własne. Med Dyd Wych 45(5),pp. 30-34.

[27] Costa-Santos, C., et al. (2007). E-learning at Porto Faculty of Medicine. A case study for the subject 'Introduction to Medicine'. Stud Health Technol Inform 129(Pt 2),pp. 1366-71.

[28] Dantas, A.M. and R.E. Kemm. (2008). A blended approach to active learning in a physiology laboratory-based subject facilitated by an e-learning component. Adv Physiol Educ 32(1),pp. 65-75.

[29] Seluakumaran, K., et al. (2011). Integrating an open-source course management system (Moodle) into the teaching of a first-year medical physiology course: a case study. Adv Physiol Educ 35(4),pp. 369-77.

[30] Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. Adv Physiol Educ 31(3),pp. 253-60.

[31] Lovatt, J., O.E. Finlayson, and P. James. (2007). Evaluation of student engagement with two learning supports in the teaching of 1st year undergraduate chemistry. Chem Edu Res Pract 8(4),pp. 390-402.

[32] Ramos, C. and E. Yudko. (2008). “Hits”(not “discussion posts”) predict student success in online courses: a double cross-validation study. Comp Edu 50(4),pp. 1174-1182.

[33] Norman, G.R., C. Vleuten, and D.I. Newble. (2002). International handbook of research in medical education. Vol. 7: Springer.

[34] Dobson, J.L. (2008). The use of formative online quizzes to enhance class preparation and scores on summative exams. Adv Physiol Educ 32(4),pp. 297-302.

[35] Brookhart, S.M. (1993). Teachers' grading practices: Meaning and values. J Edu Measure 30(2),pp. 123-142.

[36] Ben-David, M.F. (2000). AMEE Guide No. 18: Standard setting in student assessment. Med Teach 22(2),pp. 120-130.