PERSONALIZING QUESTIONS USING ADAPTIVE ONLINE KNOWLEDGE ASSESSMENT

MIRAN ZLATOVIĆ
University of Zagreb, Faculty of Organization and Informatics, miran.zlatovic@foi.hr

IGOR BALABAN
University of Zagreb, Faculty of Organization and Informatics, igor.balaban@foi.hr

Abstract: *Current implementation of the Bologna Process in the Republic of Croatia requires that students earn their final grades through continuous effort exerted during the entire semester. To facilitate continuous monitoring and evaluation of students' knowledge and activities, we designed a new adaptive assessment system that incorporates the elements of adaptivity within a series of assessments. Its design allows for a continuous and cumulative knowledge assessment wherein earlier course topics are re-assessed in every subsequent assessment in an adaptive manner based on individual students' achievement of learning goals related to those topics in previous assessments. This paper shows how the system uses a predefined set of rules to adapt questions for each student during the re-assessment of earlier content. The system also facilitates individual post-assessment feedback inferred from previous assessment of each student. Its purpose is to stimulate the emergence of appropriate learning strategies in order to improve the success in achieving learning goals related to re-assessed course topics.*

Keywords: *Online knowledge assessment, Adaptive knowledge assessment*

1. INTRODUCTION

Continuous monitoring and evaluation of students' knowledge and activities, as required by the Bologna Process, can be facilitated by using ICT in instruction. Various forms of knowledge assessment are necessary to check whether the pre-defined learning goals have been achieved. Inclusion of ICT in knowledge assessment, especially in blended education, makes it possible to assess students' knowledge more frequently. However, not all types of online knowledge assessment are suitable for facilitating the more desirable deep learning strategies, which should in turn lead towards better achievement of learning goals [13]. Creating an online system for adaptive knowledge assessment based on the announcement and application of appropriate types of assessment that would be capable of steering the student towards higher levels of learning goals achievement therefore presents a challenge.

This paper is structured as follows: in Chapter 2 we explore the institutional and pedagogical background of the new adaptive assessment system, while fundamental principles of the new system are described in Chapter 3. Adaptive rules used in the system are presented in Chapter 4. Chapter 5 shows an example and the discussion of individual questions adaptation, based on adaptive rules of the system. The conclusion is provided in Chapter 6.

2. RESEARCH BACKGROUND

Continuous Monitoring of Students' Activities

The changes caused by the implementation of the Bologna Process on the European higher education can be viewed from many perspectives. One of the most significant changes brought about by the Bologna reform has been the new approach towards the ways of passing on knowledge to students. The basic assumption is that students are supposed to earn passing grades through continuous effort and studying during the entire semester, and not exclusively through final examinations. Therefore, individual activities and knowledge of each student have to be monitored and evaluated continuously for every course to determine whether a particular student has met the criteria required for obtaining the course ECTS credits. The fact that universities were free to design their own activity tracking resulted in a number of different tracking models: project-oriented; eliminatory (all activities are equally important – the student must exceed the elimination threshold for each activity); accumulative (points for each activity are added up and the final sum of points must exceed the elimination threshold), etc. In this paper we focus on a variant of the accumulative activity tracking model, which is one of several tracking models used in the authors' institution. A specific feature of this accumulative variant is that content units are not assessed only once (i.e. within a single formal mid-term exam), but multiple times, so that every subsequent mid-term exam includes re-assessment of earlier content units too, not only the first-time assessment of new units.

Continuous Knowledge Assessment

Reliance upon standardized final testing, i.e. predominant use of summative knowledge assessment ([4], [5], [11], etc.) does not stimulate efficient learning for all types of students ([1], [7]). Teachers should therefore also be able to continuously monitor students' progress and to adapt their teaching to the needs of different groups of students accordingly. It is those principles that have led to the development of continuous knowledge assessment.

Essentially, continuous knowledge assessment is a type of formative assessment because it enables the collection of various indicators of learning progress while students are still actively engaged in the learning process and not only at the end of the educational process. Those indicators can be used to perform corrective actions within the teaching process, e.g. to adapt teaching methods to the needs of individuals or groups of students. McAlpine [9] defines continuous assessment as "… the more modern form of modular assessment, where judgements are made at the end of each field of study". Generally speaking, continuous assessment is not focused solely on knowledge assessment activities (tests, assignments, etc.) but also on everyday activities which are the foundation for the assessment itself. Also, continuous assessment is not all about *constant* knowledge assessment, but it rather presupposes that various knowledge assessment techniques are blended with appropriate teaching techniques and learning stimulation techniques [10].

According to [9] and [10], the main advantages of continuous knowledge assessment arise from the fact that it:

- Supports the educational process in parallel with students' development – data which are gathered while monitoring students' development can be used to adapt the teaching process to further stimulate their development;
- Improves learning – it stimulates students to utilize deeper thinking and understanding of learning content, while timely feedback from the teachers facilitates necessary changes in the students' learning process;
- Facilitates teachers' own professional development – while trying to better understand their students and steering their learning and thinking, teachers are in a position to better analyze their own work and improve their teaching techniques;
- Gives all the necessary information to create formal reports of students' progress – reports for parents or institutions that supervise education, etc.

One of the biggest disadvantages of continuous assessment is increased teachers' workload, because they have to spend more time to prepare and carry out frequent, almost daily, activities to track their students [9].

3. CONTINUOUS KNOWLEDGE ASSESSMENT AND ADAPTIVITY

Most of the existing studies in the field of adaptive online knowledge assessment are concerned with various aspects of adaptability within a single test, usually within self-assessment and formative assessment ([2], [3], [6], [8], [12], etc.) and are primarily focused on systems that must provide adaptive capabilities within a single assessment.

However, none of those systems facilitates adaptive knowledge assessment that spans across a sequence of assessments (e.g. tests), in which the selection of questions for each subsequent test would depend upon the achievement in previous tests. It is obvious that in order for this type of assessment to be meaningful and feasible, particular units of learning content (learning objects) have to be assessed more than once, through multiple iterations, and not only within one test, i.e. one iteration. In one possible implementation of such knowledge assessment that we propose in this paper earlier content can be re-assessed within each subsequent test so that:

- There are N iterations of testing during the class cycle (e.g. series of N assessments during one semester);
- Within the first iteration only new learning content is assessed;
- In every subsequent iteration earlier content units are re-assessed, but new content units have to be introduced as well and assessed for the first time. Earlier learning content is re-assessed to a lesser extent, with a diminishing number of questions in every subsequent iteration (e.g. content from the 1st iteration could make only 40% of the total content assessed in the 2nd iteration, or only 20% of the total content assessed in the 3rd iteration, etc.).

When a certain learning object (LO) is assessed for the first time, it is assumed that this LO has entered the initial, non-adaptive phase of assessment. When that same LO is assessed for the Nth (N>=2) time, it is assumed that it has entered the adaptive phase of assessment. For more details about the differences between the adaptive and non-adaptive phase, see the explanations in Chapter 5.2.

This form of knowledge assessment represents a type of **continuous** (carried out through multiple iterations) and **cumulative** (iterations cannot be considered mutually independent, because subsequent iterations include earlier content alongside the newly introduced content) knowledge assessment. As such, it should be well suited for usage within the Bologna Process, especially in courses that embrace cumulative models of monitoring and evaluation of students' activities.

The basic idea behind such a form of knowledge assessment is to individually adapt the questions for each student, so that the levels of learning goals achievement for each student would improve as much as possible by the end of the continuous knowledge assessment cycle. The foundations for such an adaptive knowledge assessment system will be described in the following chapters.

4. ADAPTIVE RULES

The following adaptive rules will be used to select questions during the adaptive phase of assessment:

A) **Rules to select the difficulty of questions** – depending on the level of a student's achievement of a learning goal in a previous iteration (achievement of student S_i for learning goal $LG_{n,m}$ which is related to learning object LO_m), select difficulty in such a manner that the current achievement levels are at least maintained or improved:

1. **previous iteration = "Fail"**: for student S_i select high-difficulty questions (i.e. highest difficulty available) for $LG_{n,m}$ ("striving to improve the non-satisfactory level of achievement") – **rule r1**;

2. **previous iteration = “Sufficient” or “Good”**: for student S_i select medium- and high-difficulty questions available for $LG_{n,m}$ (“maintaining a decent level of achievement, with an incentive for improvement”) – **rule r2**;
3. **previous iteration = “Very good" or “Excellent"**: for student S_i select low- and medium-difficulty questions available for $LG_{n,m}$ (“do not forget this learning content”) – **rule** r**3**.

B) **Rule to decrease the number of questions used during the adaptive phase (rule r4)** – to avoid the inevitable question inflation caused by the repeated assessment of all learning goals from all previous iterations and first-time assessment of the new learning goals (which would also lead to assessment duration issues), for each learning goal $LG_{n,m}$ of learning object LO_m in the adaptive phase the system shall use at least one question and at most **$n(I_f)/I$** questions, where:

- **$n(I_f)$** is the number of questions that were used to assess learning goal $LG_{n,m}$ of learning object LO_m during the non-adaptive phase of the assessment, i.e. during the iteration in which learning goal $LG_{n,m}$ of learning object LO_m was assessed **for the FIRST (f) time**;
- **I** is the number which indicates how many times learning goal $LG_{n,m}$ of learning object LO_m has been assessed (for the first time in the adaptive phase – I=2, for the second time in the adaptive phase – I=3, etc.).

C) **Rule to increase the number of questions used for individuals with poor achievement (rule r5)** – this rule is a modification of rule r4. For students that achieved the “Fail” level for learning goal $LG_{n,m}$ in the previous iteration, the system shall individually increase the total number of questions used to re-assess learning goal $LG_{n,m}$ (compared to the number obtained by applying only rule r4), but making sure that the total number of questions used for the re-assessment of learning goal $LG_{n,m}$ does not exceed number **$n(I_f)$.** Because of the poor achievement of learning goal $LG_{n,m}$, its re-assessment in the adaptive phase should again be thorough.

5. ADAPTATION OF QUESTIONS

This section describes the procedure of adapting questions to a particular examinee based on the set of adaptive rules introduced in the previous section. For the sake of simplicity and clarity, the process of adaptation will be described as a small case which includes only one learning object (LO), several learning goals (LG) associated with that LO and several questions of varying difficulty per each LG. Also, this example includes only one examinee. The LO is then tracked as it progresses through multiple iterations of the continuous adaptive knowledge assessment. Although the given example is very simple (comprising one examinee and one LO), the adaptation procedure itself does not impose any limitations in terms of either the number of LOs used in assessment, or the number of examinees.

Example Set-Up

The following example consists of one learning object to be assessed entitled “Decision support and expert systems” (LO1). Learning goals associated with LO1 are listed in Table 1.

Table 1: Learning goals associated with learning object LO1

Code	Learning Goal
LG1	Define characteristics of decisions and of the decision making process.
LG2	Define decision support systems (DSS) and expert systems (ES).
LG3	Describe the structure and applications of DSS and ES.

A short summary of the questions pool available to assess these LGs, including question type, amount and difficulty, is presented in Table 2. The proposed adaptive system supports the following types of questions:

- SC and MC – multiple choice questions, with only one (SC) or more than one correct answers (MC);
- FILL – questions that require filling in the missing words in sentences;
- MATCH – questions that require items matching;
- ESSAY – questions that require free-form and possibly long textual answers (essay-type questions, manual grading).

One qualitative indicator, which expresses the difficulty of a question, must be assigned (by the teacher) to each question in the pool. The system supports three levels of difficulty: “Easy” (DL1), “Medium” (DL2) and “Hard” (DL3).

Table 2: Questions pool associated with learning goals

Learning Goal	Question Type (Number of Questions)	Difficulty Levels
LG1	SC (2), MC (1)	DL1
	MC (1), MATCH (2)	DL2
LG2	SC (2), MC (1)	DL1
	FILL (1), ESSAY (2)	DL2
LG3	MC (2)	DL1
	ESSAY (2)	DL2
	ESSAY (2)	DL3

The levels of learning goals achievement are calculated for each LG during assessment evaluation. Those are quantitative indicators which show an examinee’s average score (converted to percentage) for all questions that were used to assess a particular learning goal, expressed in a typical grading scale:

- 0-49.99% – Fail (1, F)
- 50-62.49% – Sufficient (2, D)
- 62.5-74.99% – Good (3, C)
- 75-87.49% – Very good (4, B)
- 87.5-100% – Excellent (5, A).

5.2. Discussion of an Adaptation Scenario

The total number of iterations within continuous assessment is not limited by the proposed algorithm. In this scenario, 3 iterations will be included. In the first, non-adaptive iteration (I1) of the assessment, both the number and the difficulty of questions used to assess all learning goals (LG) are pre-defined by the teacher during the initial assessment structure definition. Each examinee receives an equal number of questions and equal distribution of question difficulty per learning goal. The questions themselves do not necessarily have to be identical for all examinees – in the first iteration of the assessment, the system can also select random questions (within the required difficulty level) for each examinee. In our example (see Table 3), the teacher has allowed all available question difficulty levels for all LGs in iteration I1. During answers evaluation, the system assigns an appropriate level of achievement to each LG included in the assessment. According to the last row in Table 3, our examinee has acquired the following achievement levels in iteration I1: “Fail” for LG1, “Good” for LG2 and “Very good” for LG3.

Following a full evaluation of an iteration, each student also receives personalized feedback which includes his/her achievement levels per LG. For every LG, feedback also contains an announcement of the difficulty of questions to be used in next iteration to re-assess that LG. These announcements are based on the current LG achievement levels and the internal logic of adaptive rules r1, r2 and r3 (i.e. if the current achievement level for LG1 is “Fail”, according to rule r1, difficult questions for this LG should be announced in next iteration, etc.). The next iteration of the assessment does not begin immediately – the teacher decides when it will be scheduled. The interval between two iterations should be long enough to allow students to accept suggested difficulty announcements and to adapt their learning strategies so that they are able to improve their achievement of earlier LGs during re-assessment.

In the next iteration I2 (Table 4), LO1 moves on to the adaptive phase. The number and the difficulty of questions per LG are therefore determined according to the levels of achievement from the previous iteration, using adaptive rules. For LG1, with the previous achievement level marked as “Fail”, the system will apply adaptive rules during re-assessment of LG1 in iteration 2 (I2) in the following way:

- Difficulty – the system selects the questions of highest available difficulty, according to adaptive rule r1 (normally, rule r1 assumes only questions with difficulty DL3 (hard), but since there are no DL3 questions for LG1, the system uses the fallback mechanism and selects the highest available difficulty, DL2 (medium)).
- Number of questions – is determined according to rules r4 and r5. Global rule r4 is always activated first (irrespective of achievement levels) in order to reduce the number of questions used during re-assessment of earlier LGs (to suppress otherwise inevitable question inflation). Since this is the 2nd iteration, rule r4 reduces the number of questions used to ½ of the original number of questions for LG1, i.e. $n(I_f)/2$. Rule r5 is then activated, based on the level of achievement marked as “Fail”; due to poor achievement, the number of questions determined by r4 is increased up to $n(I_f)$.

For LG2 (previous achievement level “Good”), difficulty is determined by rule r2 (mixture of DL2 (medium-difficulty) and DL3 (hard) questions), while the number of questions is determined only by global rule r4. Rule r5 does not apply here, since the achievement level of LG2

Table 3: First iteration of the continuous adaptive knowledge assessment (I1) for LO1

	LG1	**LG2**	**LG3**
Previous LG achievement level:	-	-	-
Adaptive rules used:	-	-	-
Number of questions: *	0 < n <= max(n)	0 < n <= max(n)	0 < n <= max(n)
Difficulty of questions: *	DL1, DL2	DL1, DL2	DL1, DL2, DL3
Level of LG achievement:	Fail	Good	Very good

* Both the number and difficulty of the questions used in the non-adaptive phase are determined manually by the teacher during the specification of the initial structure of an assessment, i.e. adaptive rules are not used at all in this phase.

Table 4: Second iteration of the continuous adaptive knowledge assessment (I2) for LO1

	LG1	**LG2**	**LG3**
Previous LG achievement level:	Fail	Good	Very good
Adaptive rules used:	r1, r4, r5	r2, r4	r3, r4
Number of questions: *	1. 1 < n <= n(I1)/2 (r4) 2. n(I1)/2 < n <= n(I1) (r5) => 1 <= n <= n(I1)	1 <= n <= n(I1)/2 (r4)	1 <= n <= n(I1)/2 (r4)
Difficulty of questions: *	DL2 (r1)	DL2 (r2)	DL1, DL2 (r3)
Level of LG achievement:	Very good	Good	Sufficient

* Both the number and difficulty of the questions used in the adaptive phase are determined solely by the adaptive rules.

is above “Fail”. Similarly, for LG3 (previous achievement level “Very good”), difficulty is determined by rule r3 (mixture of DL1 (easy) and DL2 (medium-difficulty) questions) and the number of questions is again determined only by global rule r4. Rule r5 does not apply here either, since the achievement level of LG3 is also above “Fail”. After the evaluation of answers in I2 is finished, it is evident that our hypothetical examinee has significantly improved his/her achievement level for LG1 (“Fail” -> “Very good”) and maintained the achievement level for LG2 (“Good”), while the achievement level for LG3 (“Very good” -> “Sufficient”) has significantly deteriorated.

testing within blended education in institutions that have adopted the Bologna Process. In the environments that either practice full online education and/or have not implemented the Bologna Process this model of assessment should therefore be applied with caution.

The proposed assessment model is designed to facilitate positive characteristics of continuous assessment. ICT-supported implementation of such an assessment model should also partially alleviate one of the major disadvantages of continuous assessment – heavier teachers’ workload. The model also incorporates theoretical findings concerning the possible impact of timely feedback and announcements on the development of students’ learning strategies in reasonably short periods of time, which in turn can stimulate the more desirable deep learning and better levels of learning goals achievement [13].

Table 5: Third iteration of the continuous adaptive knowledge assessment (I3) for LO1

	Learning goals		
	LG1	**LG2**	**LG3**
Previous LG achievement level:	Very good	Good	Sufficient
Adaptive rules used:	r3, r4	r2, r4	r1, r4
Number of questions: *	1 <= n <= n(I1)/3 (r4)	1 <= n <= n(I1)/3 (r4)	1 <= n <= n(I1)/3 (r4)
Difficulty of questions: *	DL1 (r3)	DL2 (r2)	DL2, DL3 (r2)
Level of LG achievement:	Very good	Very good	Good

* Both the number and difficulty of the questions used in the adaptive phase are determined solely by the adaptive rules.

Rule r5 does not apply to re-assessment of any LGs in iteration I3 (Table 5) since none of the achievement levels in I2 were marked as “Fail”. According to global rule r4, the initial number of questions is further reduced for all LGs – since the associated LO is now in its 3rd iteration, the initial number of questions per LG is $n(I_f)/3$. To re-assess LG1 (previous achievement level “Very good”), a mixture of questions of DL1 (easy) and DL2 (medium) difficulty will be used, according to rule r3 (although in this particular example only DL1 questions will be used; since there are no DL3 (hard) questions for LG1, the system is already using DL2 questions as “difficult” ones, therefore only DL1 questions remain in the pool of medium-difficulty and easy questions required by rule r3 – another example of the fallback mechanism). For re-assessment of LG2 (previous achievement level “Good”), according to rule r2, only DL2 questions will be used (rule r2 would require a mixture of DL2 and DL3 questions, but since there are no DL3 questions for LG2, the system will be forced to use only the DL2 level). To re-assess LG3 (previous achievement level “Sufficient”), again according to rule r2, a mixture of DL2 and DL3 questions will be used (no fallback this time, since both DL2 and DL3 questions are available for LG3). Like two previous iterations, I3 also ends with answers evaluation, wherein another set of achievement levels is assigned to all re-assessed LGs and the assessment can be continued through additional iterations.

6. CONCLUSION

In this paper we presented the foundations for an adaptive assessment system that supports continuous and cumulative assessment of students’ knowledge, which could be used in higher education. The current version of the system has been designed for implementation and

Further steps in our research involve (1) designing and developing the practical implementation of an online assessment system for continuous adaptivity that will use the question personalization algorithm described in this paper, and (2) testing the efficiency of such a system in real-world classes, e.g. to compare it with the efficiency of traditional, non-adaptive assessment.

LITERATURE

[1] Black, P., Wiliam, D., *Assessment for learning in the classroom*, Assessment and learning, Gardner, J. (Ed.), Sage, London, pp. 9-26, 2006.

[2] Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez de la Cruz, J. L., Ríos, A., *SIETTE: A Web–Based Tool for Adaptive Testing*, International Journal of Artificial Intelligence in Education, IOS Press, Vol. 14, pp. 1-33, 2004.

[3] Guzmán, E., Conejo, R., *Self-assessment in a feasible, adaptive web-based testing system*, IEEE Transactions on Education, Vol. 48, No. 4, pp. 688-695, 2005.

[4] Harlen, W., Deakin-Crick, R., *A systematic review of the impact of summative assessment and tests on students' motivation for learning*, EPPI-Centre Review, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, 2002., [Online], Available: https://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=

Pbyl1CdsDJU%3D&tabid=108&mid=1003 [20.08.2015]

[5] House of Commons, *Testing and Assessment*, Third Report of Session 2007-08, Volume I, House of Commons - Children, Schools and Families Committee, London, 2008., [Online] http://www.parliament.the-stationery-office.co.uk/pa/cm200708/cmselect/cmchilsch/169/169.pdf [20.08.2015]

[6] Huang, Y.-M., Lin, Y.-T., Cheng, S.-C., *An adaptive testing system for supporting versatile educational assessment*, Computers & Education, Vol. 52, No. 1, pp. 53-67, 2009.

[7] Kapukaya, K., *Assessment: A help or Hindrance to Educational Purposes*, International Journal of Humanities and Social Science, Vol. 3, No. 6, pp. 84-93, 2013.

[8] Marinagi, C. C., Kaburlasos, V. G., Tsoukalas, V. T., *An architecture for an adaptive assessment tool*, in 37th Annual Frontiers in Education Conference - Global Engineering: Knowledge without Borders, Opportunities without Passports, IEEE, Milwaukee, USA, pp. 11-16, 2007.

[9] McAlpine, M., *Principles of Assessment*, Robert Clark Centre for Technological Education, University of Glasgow, Ed. by Computer Assisted Assessment Centre, University of Luton, 2002., [Online], Available: http://www.caacentre.ac.uk/dldocs/Bluepaper1.pdf [20.08.2015]

[10] O'Brien-Carlson, M., Humphrey, G. E., Reinhardt, K., *Weaving science inquiry and continuous assessment: using formative assessment to improve learning*, Corwin Press, 2003.

[11] Russell, T., McGuigan, L., *Making Formative Use of a National Summative Assessment Regime*, chapter in *Research in science education - Past, present and future*, Behrendt, H., Dahncke, H., Duit, R., Gräber W., Komorek M., Kross A. and Reiska P. (Eds), Springer Netherlands, pp. 71-76, 2001.

[12] Wen, D, Graf, S., Lan, C. H., Anderson, T., Kinshuk, C., Dickson, K., *Adaptive Assessment in Web-Based Learning*, in 2007 IEEE International Conference on Multimedia and Expo, IEEE, Beijing, pp. 1846-1849, 2007.

[13] Zlatović, M., *Utjecaj adaptabilne online provjere znanja na poticanje strategija učenja i realizaciju ciljeva učenja/Influence of Adaptive Online Knowledge Assessment on Stimulation of Learning Strategies and Achievement of Learning Goals*, unpublished doctoral thesis, University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia, p. 296, 2011.