

STUDY MAPS AS A TOOL FOR THE ADAPTIVE TESTS CONSTRUCTION

Chapter 1

Number of university students is very high in some courses. Verification of their knowledge cannot be made by an oral examination and similar time-consumption methods, hence computer aided testing is used very often. Questions that can be processed by the computer itself are used in the computer tests, these questions are composed from the taught curriculum selection by qualified teachers. But this composing process has not a systematic support and an appropriate balancing of the share of parts and difficulty. Thus, results obtained from these tests usually does not represent required explanatory power. The test result is often depicted as an amount of points and student obtains appropriate mark on this point basis. We detect some principal questions in this context: What knowledge, however, such a mark match? Have two students with the same score the same knowledge? Is it possible to get passed mark, even if a student's knowledge is insufficient? These problems were discussed in many articles with many points of view. For example, prof. Vaníček (1999) says: ...numbers (marks) do not reflect any operation of composition in the real world. There even is not any empirical justification between particular grades. The difference between 3 and 4 is a difference between success and fail, the difference between 1 and 2 is a different intensity of a complementary praise to the success in passing the exam. If we calculate an arithmetic mean from the grades, we gain a number that has no meaning in an empirical world. 1 This cite from a year 1999 sums up one of topics of our research and this paper – the topic is finding an answer and a possible solution on a question – is one number (grade) able to describe student's knowledge? Especially when the number was obtained from a computer assisted test with closed questions? In 2012 a preliminary pedagogical experiment of comparing a computer assisted test (with closed questions) with an oral exam was performed. The goal was to get an objective prove that the results of this kind of test doesn't provide an appropriate information about student's knowledge. Simultaneously we wanted to gain a proper foundation for a design of a computer assisted adaptive test competing with the drawbacks of currently used regular computer assisted tests. Hand in hand with this goal goes another need – a design of a representation of the results of the test depicting structure of student's knowledge verified by the test. Objective of this paper is to describe results of this pedagogical experiment and to introduce a proposal of a method for computer adaptive testing with a focus on knowledge of context.

MATERIALS AND METHODS

The experiment was performed within a university course Informatics for Economists II (Informatika pro ekonomy II, an abbreviation IPE2 is used in following text) taught at Faculty of Business and Economics (FBE) and within a course Informatics in Agribusiness (Informatika v agrobyznysu, INAGB) taught at Faculty of Agriculture (AF), both at Mendel University in Brno (MENDELU) during the summer semester 2011/2012. Goal of the experiment was to intercept the difference between "classic" computer based test (computer assisted test with closed

questions) and an oral exam. The experiment was performed on 90 tests and it consisted of a written computer test and of a following oral exam. Written computer assisted test was performed via an e-learning application in a University Information System of Mendel University (UIS). The test from the course IPE2 is a test with 25 closed questions of different types: dichotomous questions, ordering questions, multiple choice questions with one or with more correct answers. Main condition of this test construction was computer evaluation without an intervention of teachers, because these courses are studied by very large number of students. This condition excludes the use of more complex or open questions. The whole test is divided into five thematic modules; at least 50% from each module is necessary to pass the test (the reason for their implementation is described in details in Haluza, Talandová, 2009). Guessing correction is implemented in the test, i.e. wrong answers are penalized with negative amount of points. This amount is proportional to the probability of guessing the correct answer. Test from the course INAGB is a test with 17 closed questions of different types, no division into modules is implemented here, guessing correction is used, the minimum to pass the test is 50% (after recalculation due to the guessing correction).

Chapter 2

The form of the oral exam was following: an open argument with a student – few questions were selected from the test and then interdependent questions were asked. From each module generally one question was selected – with emphasize to wrong-answered in previous test, but randomly were used right-answered or skipped questions. Interdependent questions on related terms were asked in dependency on student's answer with emphasize on a broader context and relations between terms. All questions and answers were recorded by a record-keeper together with a personal evaluation of student knowledge. After the experiment was completed, all the data from the computer test was uploaded to a database which was created for this purpose and also the notes from the oral examining were rewritten into this database. Detailed description of the course of the experimentation, manner of posing the questions and the format of the database is in a paper from conference PefNet 2012 (Dlabolová, 2012). The primary method used to examine the results of the experimentation was chicanery – particular tests and following interdependent questions were chosen as representatives of common cases or observations during the experimentation and described in details to illustrate the results.

RESULTS

To depict the difference between a computer test and an oral exam, particular tests were selected. Three main kinds of situations were observed:

- The test result is similar to the oral exam results.
- Deep disregard of the topic – wrong answer doesn't mean only a mistake in the specific question, but behind the wrong answer there is an absence of knowledge of main terms.
- The result of the written test doesn't correlate to the result of the oral exam – in both ways.

The detail description of specific chosen examples follows. The first case is an acceptable state showing only a beneficial difference in the results of both tests. In the tests from this category, students generally don't know incorrectly answered questions and knows correctly answered ones. Sometimes there are some small mistakes which are caused by wrong understanding of the question or similar reasons. There can be also found some amount of questions just presumed or learned by heart from the study materials. In this category there is in some cases a primary background knowledge behind incorrectly answered or skipped questions. Following two tests are examples of deep disregard of the topic, one item was selected to depict the problem from both tests. First example is from a test which is named A12, which is from a course INAGB, the conclusion of the test was 56.02%. One of the items argued during the oral exam was *The term relation in a relation database system means:* with a wrong answer: *a relationship between different entities from different sets*, the correct answer for this item was: *selected factors originated from a Cartesian product of sets, which represent data types*. Student was asked to describe different kinds of database systems – ranked, network and relational. He gave fundamental characteristics, which were exactly citing aspects of the database systems as they were stated in the study material, thus the terms from the question were decided one by one in following dialog:

Chapter 3

Q: What kind of relationships in the database system are they between?

A: The relationships in the database system are between computers and computer systems.

Q: What does an entity mean?

A: Services.

Q: What does a database system mean?

A: It assembles and supplies information.

Q: What kind of components of a database system are there?

A: Information.

Q: What does a database management system mean?

A: Data are controlled there.

It can be proved that the student doesn't have information about the database topic thanks to this dialog.

A course INAGB from a test A2 is another example. The result was 51.77%, one of the asked questions was *Compression is:* it was answered correctly (*process of verbosity which decreased in data*). Following interdependent questions were given with no or wrong answer. The student was given a term's short explanation from the previous question:

Is there a main principle of compression?

What kind of purpose of using compression is there?

Did you use compression practically?

What does compression rate depend on?

Do you know what kind of differences are there between BMP and JPG format?

Which file format is too close to an image with no compression? A: GIF.

Again, the absence of the information, now behind a correctly answered item, is obviously seen.

Deep difference between the written test result and the oral exam result follows with three examples from the third category. E1, which is the course IPE2, is the first example and the result was 86.98%, which was the best result for IPE2. Three of 25 questions were answered incorrectly, one of them was skipped and one was partly correct. First tested question was: *Which mathematical function can transform too many different realizations of a phenomenon on the quantity corresponding to the amount of knowledge?* it was answered correctly (logarithm of any base). Real very poor student's knowledge about the information theory and about logarithm function appeared during the consecutive dialog. Next asked item was related to recognizing of a file's content in a CSV format. The student knew the definition of textbook what CSV format is, but didn't know the usage. Then questions to a database module were asked (all the five items in the database module were quite similar, hence without any particular one was chosen, the result of the module was 65%). It was revealed that the student doesn't know what a database is thanks to the further questions – e.g. he does not have any idea about who or what poses the questions for a database.

The other example is named E2, the course IPE2 again. Its result is 54 points and 66.75%, just one module was missing with result 40%. The student couldn't answer any of interdependent questions, even they were about the items that the student could answer correctly. The last example for this category is named E3, course IPE2, it was a re-sit, the previous result was 56% (the oral exam's result was similar to the result of the written test during the first term, but in some cases, student couldn't answer question correctly). The result of E3 was 77.33 points, 83.33%, all modules were accomplished. And the result was the second best for IPE2. The student mostly answered no questions, some of his answers were totally wrong. Anyway, the student alleged that he could understand the course's subject better because of the previous oral exam.