Computers & Education 92-93 (2016) 142–160

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Evaluation of knowledge in Object Oriented Programming course with computer adaptive tests

Sanja Maravić Čisar [a, *], Petar Čisar [b], Robert Pinter [a]

[a] Subotica Tech-College of Applied Sciences, Marka Oreškovića 16, 24 000 Subotica, Serbia
[b] The Academy of Criminalistic and Police Studies, 196 Cara Dušana Street, 11 080 Belgrade, Zemun, Serbia

ARTICLE INFO

Article history:
Received 1 February 2015
Received in revised form 13 October 2015
Accepted 14 October 2015
Available online 19 October 2015

Keywords:
Computer adaptive tests
E-learning
Knowledge tests
Marking of knowledge
Adaptability

ABSTRACT

This paper presents and analyzes the results of the application of computer-adaptive testing in knowledge evaluation of the programming language C++. The participants of the research were college students majoring in computer engineering from a polytechnic college (Republic of Serbia) with an average age of 20. The research was conducted over a two-year period. Overall 199 students took part in this research. Two groups of students, the experimental, and the control one were observed. The students in the experimental group took the computer adaptive test and the students from the control group did the paper-and-pencil test. For the purposes of determining the effects of applying the computer adaptive test for knowledge evaluation, the adaptive test was realized in the MATLAB software package. The test consists of multiple choice questions with five possible answers. The questions for the test are divided into three clusters according to the level of difficulty (easy, medium and hard). The score is calculated from the level of difficulty. The difficult questions earn more points than the easy ones. The examinees may answer the same percentage of questions correctly, but an examinee will gain a better score assuming they correctly answered more difficult items. The results of the research show that there is a statistically significant difference between the results of the control and experimental group. The students who worked on a computer-adaptive test achieved a higher average score than the students who did the traditional test.

1. Introduction

One of the most common ways of determining the measure of knowledge acquisition in a certain field of science is by testing. The main goal of testing is to determine the level of the students' knowledge from one or more subject areas in which knowledge is checked. Different methods of knowledge evaluation are in use, such as in class presentations, writing essays, projects, etc. However, the most common “tool” that is used to examine knowledge is the test and oral exam. For the examinee, it may be discouraging if they are faced with questions that are too difficult for them. However, at the other end of the spectrum are the too easy questions, which may cause the examinee to lose interest in the testing procedure, as they lack any challenge for their knowledge level. One of the methods that may offer a suitable solution to this problem is the

* Corresponding author.
E-mail addresses: sanjam@vts.su.ac.rs (S. Maravić Čisar), petar.cisar@kpa.edu.rs (P. Čisar), probi@vts.su.ac.rs (R. Pinter).

http://dx.doi.org/10.1016/j.compedu.2015.10.016
0360-1315/

implementation of computer adaptive tests (CAT), whereby the system will provide the examinees with varying levels of questions, adapting to the examinee's abilities.

Using the computer adaptive test makes it possible for the course teacher to generate a test that will match the knowledge level of the given examinee. During the process of examination the system relies on the level of the previously given answer by the examinee when determining the level of the subsequent answer. Assuming that the examinee gave a correct answer, the following question will be harder. Conversely, an incorrect answer will be followed by a question from the easier category. The adaptive process ensures that the test is closely matched with the examinee's capabilities. What final score the examinee is eventually given depends not only on the percentage of given correct answers, but also on the level of difficulty of the questions. In the early phases of psychological measurement Alfred Binet (1857–1911) was among the first to implement the concept of adapting a test to the examinee's level of knowledge. More than half a century later Binet's adaptive approach to how capacity is measured is still the only functioning adaptive test. In the middle of the 20th century the US military did some research work into the how adaptive testing could be applied in the traditional form of testing (using paper and pen) as well as in expanding into machine-based testing (Bayroff, 1964; Bayroff, Thomas, & Anderson, 1960). It must be stated, though, that neither of these approaches yielded satisfactory results.

Adaptive tests are used in many programs for gaining certificates for different professions, such as for accountants in the United States (AICPA, 2015), medical technicians from the American Society of Clinical Pathologists Board of Registry Certification Examinations (ASCP, 2015) and for determining the qualifications for service in the armed forces of the United States (ASVAB, 2015).

The research group for computer adaptive testing from the University of Madrid (Autonoma University of Madrid), developed an adaptive test of English for those examinees whose mother tongue is Spanish (Olea, Abad, Ponsoda, Aguado, & Díaz, 2011). Cito, the National Institute for Educational Measurement in the Netherlands applies computer adaptive tests for checking knowledge of mathematics, reading (Dutch as a first language), geography, and reading and comprehension tests of Dutch as a second language (Eggen, 2007). Adaptive tests have been developed also for children from age 4 to 6, which include knowledge of basic arithmetic and language skills as well as orientation in space and time (Cito, 2015). Civil servants working in Brussels or in municipalities with linguistic facilities have to prove their active proficiency of the second national language (either Dutch or French). The Belgian Civil Service Commission tests over 15,000 candidates a year, so they need a fast, objective and secure testing system. In order to satisfy these conditions a research group at the University of Antwerp has developed an adaptive system for evaluating second language proficiency (Schonenberg et al., 2013). Jacobsen et al. (2011) in their article described process of development of the CAT delivery system designed to test English literacy and numeracy skills for the Singapore Workforce Development Agency. Lilley, Barker, and Britton (2004) present their research at the University of Hertfordshire on the use of computer adaptive tests in higher education. The application was designed to estimate the level of proficiency in English for those students whose first language is not English.

The Japanese Computerized Adaptive Test (J-CAT, 2015) is a computer-adaptive test that can be done online, and is used to evaluate foreign students' proficiency in the Japanese language. It consists of four parts: text comprehension, grammar, reading and vocabulary (J-CAT, 2015).

Chatzopoulou and Economides (2010) describe their work on a web-based adaptive test for testing knowledge of basic elements of programming and the pseudo-language Glossa. It was applied in two Greek secondary schools. The results of the test were used to predict the scores of high school students in a series of exams known as the PanHellenic Examination which students need to pass in order to gain entrance into university and become university students.

This work presents the results of the research on the application of computer-adaptive testing in knowledge evaluation of the programming language C++. The participants of the research were college students of a polytechnic college (Republic of Serbia) majoring in computer engineering. The research was conducted over a two-year period. Overall 199 students participated in this research.

The paper is organized as follows. After the introduction, Section 2 presents a theoretical basis of computerized adaptive tests. In Section 3, the design of CAT is described, including item calibration, starting the test, selecting the next item to be administered and stopping the test. Background research, the examination procedure itself, as well as the methodology are given in Section 4. Section 5 presents the results of the study, which are then discussed in Section 6 before conclusions are drawn.

2. Theoretical basis of computerized adaptive tests

CAT is a type of test developed to increase the efficiency of estimating the examinee's knowledge. This is achieved by adjusting the questions to the examinee based on his previous answers (therefore often referred to as tailored testing) during the test duration. The degree of difficulty of the subsequent question is chosen in such a way that the new question is neither too hard, nor too easy for the examinee. More precisely, a question for which it is estimated, with a probability of 50%, that the examinee would answer correct is chosen. Of course, the first question cannot be selected in this way because at this point nothing is known about the examinee's capabilities (a question of medium difficulty is chosen), but the selection of the second question can be better adapted to the given examinee. With every following answered question, the computer is able to evaluate the examinee's knowledge increasingly better.

Some benefits of the CAT are as follows (Rudner, 1998):