A picture containing calendar

Description automatically generated

Bucharest Academy of Economic Studies

Faculty of Cybernetics, Statistics, and Economic Informatics

IT&C Security Master

Security of a Computerized Adaptive Testing Solution

Dissertation Thesis

**Dissertation Advisor:**

Conf. Univ. Dr. Toma Cristian Valeriu

**Graduate:**

Mareș D.S. Robert-Dorian

Bucharest 2023

Declaration of Originality

I hereby declare that the results presented in this paper are entirely of my own creation unless reference is made to the results of other authors. I confirm that any material used from other sources (journals, books, articles, and Internet websites) is clearly referenced in the paper and is indicated in the list of references.

Contents

[Abstract 4](#_Toc136705790)

[1. Introduction 5](#_Toc136705791)

[1.1. History of education and evaluation 5](#_Toc136705792)

[1.2. Multiple-choice examinations 5](#_Toc136705793)

[1.3. Education in Post-Covid Conditions 6](#_Toc136705794)

[2. Computerized Adaptive Testing with Security Related Aspects 7](#_Toc136705795)

[2.1. Introduction to C.A.T. 7](#_Toc136705796)

[2.2. Structure of a C.A.T. 7](#_Toc136705797)

[2.3. Process of a C.A.T. 8](#_Toc136705798)

[2.4. Correlation with Item Response Theory 8](#_Toc136705799)

[2.5. Advantages and Disadvantages of C.A.T. based on results. 9](#_Toc136705800)

[3. Technology Stack and Architecture 11](#_Toc136705801)

[3.1. Base of development 11](#_Toc136705802)

[3.2. Development technologies and architecture 11](#_Toc136705803)

[3.3. Security techniques and architecture 12](#_Toc136705804)

[3.4. Implemented Architectural Standards 12](#_Toc136705805)

[4. Implementation 14](#_Toc136705806)

[5. Conclusions 15](#_Toc136705807)

[Bibliography 16](#_Toc136705808)

# Abstract

Such a solution is needed because of the way in which education is evolving today. As you may already know, post-covid-19 pandemic education has taken a turn for the better adopting a mostly digital approach, in the way that humanity managed to realize that someone with internet access may have all the possibilities to become a great mind in this world, thus, any form of digitization of this concept is a big step towards the great minds of tomorrow.

It is understood why it is important to have computerized examinations, but a step forward is introducing the concept of adaptability. Adaptability is needed because of the nature of standardized testing, a concept that may lead to many academic fails for most students for ages now. Out of all standardized test methods, one kept constantly rising in popularity due to its simplicity, that being the multiple-choice question.

Multiple choice-based examinations give the least possible room for the student’s creativity, as well as being most prone to academic dishonesty. Their simple form makes good easy work of the evaluation by the teacher, yet circling the right answer on a paper, without having a word to say, may sometimes succumb to the factor known as luck, rather than knowledge. The adaptability part of computer-based adaptive examinations is trying to make exams more interactive and knowledge solicitating for the examinee.

Finally, the matter of security is an essential factor in every concept linked to computers. For as many features a machine has, a bigger number of security issues may be exploited in an attacker’s favour. Taking a digital test is just the same, we keep asking ourselves the question “How may I alleviate academic dishonesty?”. Fortunately, computers may help with this task, only if the security norms are respected.

# Introduction

## History of education and evaluation

From the beginning of time and until now, learning and education have paved the way for the society we live in today. Without this figurative hunger for knowledge that the human being has shown in its all so evolving and fast-adapting behavior, society would have not progressed thus far, to the point of its members being able to reflect on it and themselves. The act of learning led to the creation of a concept known today as education, which was common ground for the great evolution society has seen in recent history. Education, being this pillar of society, gave birth to an institution whose sole aspiration is to educate the population. This institution is best known as “school”, and it is such an important aspect for humans, that today it is part of every living human, regardless of their nationality, beliefs, or race.

Education, as a process, has many relevant aspects, yet this study aims to underline the one that exposes the results, namely, evaluation. The English Dictionary defines the word, evaluation, as the action of determining or fix the value of something, thus, in the context of education, evaluation follows the exact same principle, specifically, it is the action to determine the level of knowledge of the person being educated, commonly known as “the student”. It is common sense for the process of education to conclude with evaluation, thus, this step is necessary, so that they can bring their newly gathered knowledge into society. It is fairly known that each human being has different needs and wishes, no two people may want to receive the same treatment or cure, in this case no evaluation should be the same for every two individuals.

According to censuses conducted by the world governments, it is stated that population growth rate upsurged in the last century, the educational system, which was put in place by each country’s government for its population as a public service, grew accordingly to provide paths of performance for each citizen.

Considering the growing population, as well as each student’s needs regarding evaluation, standardized testing was adapted as the main procedure for evaluation. Standardized testing takes many forms, ranging from multiple-choice questions to interviews held face to face with an examiner, but the main aspect of it is that it applies the same set of questions and tasks for each examinee, without their specific needs and unique behavior being taken into consideration.

## Multiple-choice examinations

With standardized testing being the status quo for most educational facilities, the multiple-choice based tests proved to become the most popular among examinators, due to their distinguished aspects, among which we mention low effort on behalf of the examiner, since they are only required to validate a corresponding answer in the answer sheet, thus becoming a comfortable way of evaluating a large volume of students. On the other hand, for the examinee, a multiple-choice test drastically limits creativity and adaptability while providing an answer. By limiting one’s field of action, the results strain the truth even further.

Through the years, fundamental disadvantages of multiple-choice based tests were discovered, yet the academic system preferred the advantages and managed to find workarounds for what was missing. Naming these fundamental disadvantages goes as follows:

* Limiting one’s ability to express, a multiple-choice test drastically limits creativity and adaptability while providing an answer. By limiting one’s field of action, the results strain the truth even further.
* Academic dishonesty has been and still is to this day the main issue in all forms of evaluation. It is not specific to the multiple-choice question-based evaluation, but it is easier to perform in the case of these types of tests. The answers have a simple form, e.g.: 1.A, 2.B, 3.C and so on, thus examinees may very well discreetly communicate this information among each other through various communication channels, and the smaller the payload, the easier it would be to transmit to other examinees.
* Random choices exist as part of every human being. Unlike computers, which have a hard time creating something truly random, humans have this ability at their fingertips. For the examinee, to choose one of the answers to the multiple-choice question, and that answer being the right one there is a very high possibility, depending on the number of available answers to choose from. Even if the odds are not that high, they are still much higher than a random choice in another form of evaluation, for example written text. It is clear to say that if an examinee answered correctly to a question by using the random factor, that does not validate their abilities in the process of education.

## Education in Post-Covid Conditions

With the passing of time these problems persisted and even escalated with the evolution of technology, and the ease of transmitting messages through various channels. In the post-covid context, many educational institutions were forced to perform evaluations in an digitalized manner through the internet, thus giving the examinees enough room to commit academic dishonesty. [1]

To survive and not leave the students in a limbo state, educational facilities had to succumb to creating online environments for courses and laboratories to take place. Entering online teaching students were facing accommodation with many new tools that may be used for either sustaining a video conference between teacher and student, or for sharing resource and examinations. [2]

This study’s focus is aiming towards aiding the first two presented disadvantages. The stiffness of multiple-choice questions can be torn down by an adaptive component, meaning that the logic behind the test builds a model of the examinees capabilities and models the process according to their current performance. The academic dishonesty factor could never be deemed as fully eradicated, yet it could be threatened by implementing security solutions to the support platform.

All aspects for creating a better experience out of the multiple-choice exam are to be interlaced with each other into a single application designed specifically for sustaining such an exam.

# Computerized Adaptive Testing with Security Related Aspects

## Introduction to C.A.T.

Unlike the basic multiple-choice based test, an adaptive multiple-choice test is going to give better results, while using less items. From a technical point of view, computerized adaptive testing aims towards the successive selection of the next item out of the item pool. This selection is going to be based on the previous performance as the specific process indicates. [3]

The adaptive testing factor may be hidden from the user, letting the hardware process the examinee’s profile, yet the results will be present. If the examiner demands, the process of adaptability could be visible to the examinee, yet this kind of exam would be more appropriate for research studies, since seeing the process and the performance may bias the user and change their natural behavior.

## Structure of a C.A.T.

As stated in [4] and explored furtherly in [5], to make sure the computerized adaptive test is properly providing the results required, there are multiple components required.

* The item pool: It should be as vast as possible, while still maintaining the quality of the item since low quality items may only drive the algorithm in an uncalibrated state. As it is the saying in all popular machine learning solutions, garbage in, will only result in garbage out.
* The response model: It refers to the psychometric model that the algorithm will compute in each iteration to retrieve the optimal question for the examinee.
* The test entry procedure: It requires an item of average difficulty to leave equal room for the algorithm to evolve in both directions, ascending or descending.
* The item selection procedure: Goes as instructed by the adaptive algorithm design.
* The scoring procedure: Grading is based on the evolutive stagnation and crystallization of the psychometric model since the lack of modification is the true form of the examinee’s capabilities.
* The test termination procedure: May be based on multiple criterions such as running out of time or the previously explained evolutive stagnation. Another one could the R.O.O.T exams (run out of time), in which in a given amount of time the adaptive test is going to take place and the examinee is required to pass a specific threshold in order for the test to be marked as PASSED. [6]

Other components may be considered as optional, used to provide a better experience for the user, but the ones named above are utterly essential for the process. Many examination solutions provide quality-of-life features in the digital view of the exam, such as a timer, the progression or even multimedia features to construct a more appealing item for the examinee to tackle on.

## Process of a C.A.T.

As stated in [7] and reflected on the same idea in [5], process of adaptive evaluation may very well divide the into four elementary steps, that state the base for an efficient testing process following the means of adaptability to the examinee’s capabilities and behavior.

* The first step requires selecting the optimal element from the item pool. The quality of the item is going to be deduced based on estimations from previous iterations. Even as a principal selection criterion, the difficulty of the previous item may not serve as the only one, other criteria, like the category in which the item is placed, will provide valuable information for the adaptive algorithm.
* Secondly, the optimal item should be presented to the examinee, following the user input, regardless of the results, the digital model will be updated with the gathered information during the previous iteration.
* Thirdly, the digital model of the examinee which embodies the user’s abilities is going to be rebuilt and serve as the new train model for the algorithm.
* Finally, steps 1 to 3 get repeated until the exit criterion is triggered. The exit criterion may vary based on the examiner’s demands, usually it is based on a stagnation in the change of the digital model. When iterations become more similar from one to another, the algorithm may stop. [8]

## Correlation with Item Response Theory

For the first iteration the algorithm may not have a proper model to build on, thus the items with an average difficulty will be presented, to have a better approximation during the next iterations. Each time the building of the digital model is going to be different since the human behavior is never the same on two individuals. [9]

Item response theory is yet another component present in an adaptive algorithm created for computerized evaluation. Presented in [10], inside adaptive testing, IRT makes the difference between a low and a high difficulty item. Based on this methodology, the psychometric model will be able to mirror the examinee’s capabilities [11], as well as making a clear comparison between the item’s difficulty and category compared to the examinee and his status. IRT compares under the same metric all these aspects to provide optimal calibrated results. [12]

As stated in [13], from a mathematical standpoint we can refer to the psychometric model as *θ*. In the context of IRT, the probability of obtaining the exact answer could be calculated through iterative functions of *θ*. Based on the item score vector, certain decisions must be made regarding the probabilities obtained by computing *θ*. The probability may pe specified using 1 to 3 parameter logistic models [14]. A model example for three parameter logistic models was provided in [13].

ag – item discrimination parameter (based on category)

bg – item difficulty

cg – the probability of a correct response for every low *θ* examinees

## Advantages and Disadvantages of C.A.T. based on results.

Some observed advantages of computerized adaptive testing are as follows:

* Compared to static testing, adaptive testing offers better precision of results [7], while static tests tend to give better results only for medium trained. As a result, the outliers would face a bad examination in the case of static tests.
* In a computerized adaptive test, the difficulty of a question wears the specific weight, thus a more difficult item would give more score to an examinee, unlike static tests where usually all questions have the same value.
* Since an adaptive test may end based on a finish criterion, they tend to take 50% less time [3]. A static test always has a predefined time length since an examinee must iterate through all items to finish the test.
* Overall gives the user a better experience, since it avoids stressful situations where an examinee may be faced with consecutive difficult questions. The adaptive factor manages to help the user to correct their mistakes and improve their final score.
* Adaptive tests may find their way not only in academic environments, but also in the recruitment sector. Recruiters may setup recruitment tests in an adaptive manner to make sure they find what they are really looking for in a candidate. [15]
* All the security measures that can be applied on traditional multiple-choice test can find their way in computerized adaptive test, per se:
  + Scrambling the order of the answers to decrease the easiness of performing academic dishonesty by transmitting only one indicator. If each examinee has a different answer order, they may not be able to just express the position of the answer in the answer list, they are going to have to express the whole answer, which drastically complicates the process of academic dishonesty.
  + Presenting different starting questions, and progressive questions, if there are enough items in the item pool to choose from.
  + Encapsulating the test in a secure application that is built specifically to work against cheating by limiting the application that can run on the computer.

On the other hand, some disadvantages have been observed as it follows:

* The first few iterations of the process are used by the adaptive algorithm to calibrate the item pool. Such algorithm, in the case that it is not optimized, requires a larger item bank, thus more iterations spent in the calibration process. [3]
* In the case of multiple examinees showing identical behaviour in the interaction with the algorithm, these users may commit academic dishonesty by experiencing a similar exam and being able to consult each other during the evaluation process. [16]
* Processing of the psychological model requires hardware resources, thus the algorithm should either be optimized properly to run smoothly or be hosted by powerful hardware that can process the model fast enough.

# Technology Stack and Architecture

## Base of development

To conduct proper research on this topic and assure that the studies before are indeed still relevant in the context of post-covid education, we may construct an informatic solution, designed to englobe both computerized adaptability for item selection, as well as security aspects to mitigate academic dishonesty.

In the sense of security, such solutions may only be constructed using new technologies that benefit from long-term support from the developers. Open-source technology developers choose to adopt a certain method of security, by revealing the whole source code to the community, they build an entire environment around it, thus every contributor may come with an opinion regarding security and useful features. By gathering many contributors and reviewers, the open-source framework benefits from multiple diverse opinions. There are multiple frameworks currently developed and continuing to receive long-term support, each of them taking a different part of the market share, thus at the end of the day, it comes to generic preference when choosing a framework to develop a solution.

## Development technologies and architecture

In the scope of the research conducted in this thesis, a solution was built using the microservice architecture of Node.js development framework. The main component of this solution is a central service meant to connect the others through HTTP. This server would be responsible for transiting most of the data and forwarding it to other microservices that may process or aggregate data. The microservice architecture, even being more complex and requiring more resources, is superior to the monolith architecture, since a more distributed and scalable solution has better adaptability in case of any hazard, being it hardware or software, may occur. For instance, triggering an uncovered exception is going to cease the activity of the monolith solution, since the sole server was stopped, on the other hand this case in a microservice architecture is going to stop just one of the microservices, not affecting the other and leaving them to run properly in their own sandbox.

To be used, a solution requires an attractive user interface to facilitate interaction and user experience. In this scope, Angular v16 was used alongside its many features to facilitate an easy-to-use user interface that manages to establish communication with the main server. This interface may provide the examiner with the ability to create question items that may be assembled into a test, the test itself may later be activated and become accessible by the examinees. On the other hand, the examinee may use the solution to take a computerized adaptive test. They would first access the test by a unique identifier, afterwards the user is going to take the test, items will be displayed to them through the Angular user interface, after providing an answer, the current gathered information, information like the correctness of the answer provided, question category, answer time, is going to be processed by one, or multiple microservices in order to provide an accurate next question. This is an iterative process that is going to end when a finish criterion has been achieved.

The center of data processing and aggregation of the solution is the server which hosts the adaptive algorithm. This function dictates the entire flow of the application in the form that it is responsible for analyzing the psychological model of the user, so that it can apply custom filtering on the item bank, to retrieve the optimal item to be presented to the examinee. Custom filtering is going to be made based on the difficulty of the items, as well as the examinee’s status, which may be on an ascending or a descending trend. A secondary filtering is going to take place to not present items like the ones before that the examinee didn’t manage to provide the correct answer. This iterative workflow would be stopped by a finish criterion. Such criterion may constitute the stagnation of the psychological model in the last couple of iterations, thus deeming the evolution of the psychological model as finished. After the end of the exam, the results may be displayed to both the examiner and the examinee.

## Security techniques and architecture

In the context of security, the solution uses access and refresh token architecture so it can manage the secure transmission of payloads and avoid possible scripting attacks such as Cross-Site Scripting or Cross-Site Request Forgery. The token architecture is provided by the JSON Web Token Library which manages the in-code implementation.

Another security aspect is the integration with the open-source chromium-based software named Safe Exam Browser, which drastically limits the examinees’ available actions.

Yet another security reason at authentication is the implementation of a two-factor authentication system provided using a simple mobile application developed in parallel. On the action of user login, the mobile application is going push a notification that requires specific user sensitive data, needed for assuring the person that initiated the login is indeed the owner of the account.

Lastly, to persist the identity of the user and avoid impersonation, in parallel with the test running, video feed of the examinee is always gathered and forwarded to the examiner in the scope of surveillance. [17]

## Implemented Architectural Standards

The proper implementation of the system described previously would certainly be best created using an informatic system. Considering informatic architecture, such a solution would require specific architectural modules to offer enhanced user experience. The architectural standards identified in [18], [19], [20] and [21], are prime tools in building a sustainable and safe environment to facilitate computerized adaptive tests.

* The “State” design pattern is used in this type of solution to provide specific experience for the user, depending on their status, as an examiner or examinee, they may be presented with different guided user interfaces.
* The “Memento” design pattern serves as a core module in both storage of the user’s psychometric model in a digital form, as well as providing fail-safe techniques to restore the session in case of various hardware malfunctions.
* All these aspects come together under the rules of Clean Code, by which the source code can be easily altered and updated depending on the owner’s demand.
* The “rfc7231” enables proper usage of the HTTP and its functionalities. This protocol allows remote connections through browsers and mass aggregation of data on one server.

# Implementation

# Conclusions

To conclude what has been stated, such web solution is conceived with the scope of facilitating computerized adaptive tests in a secure manner, while benefiting from multiple quality-of-life features present to enhance user experience. The two main interfaces of the application, the examiner and examinee user interfaces, tend to achieve a comfortable and user-friendly experience.

During the development process, the technology stack proved to be the right choice for this type of application since the framework provided enough built-in features to make the process fast and reliable.

While being in a functional state, the adaptive testing application is not running optimally, such application would require a prolonged period and a committed development team to sit in an optimal state, yet as a proof of concept, it fits the role in the current state.

Unlike other online testing platforms, the solution brings the adaptive factor to the table as well as the security aspects used for academic dishonesty mitigation.

# Bibliography

|  |  |
| --- | --- |
| [1] | T. Gonzalez, M. A. de la Rubia, K. P. Hincz, M. Comas-Lopez, L. Subirats and S. Fort, “Influence of COVID-19 confinement on students’ performance in higher education,” *PLoS ONE,* vol. 15, no. 10, 2020. |
| [2] | K. Gamage, D. Wijesuriya, S. Ekanayake, A. Rennie, C. Lambert and N. Gunawardhana, “Online Delivery of Teaching and Laboratory Practices: Continuity of University Programmes during COVID-19 Pandemic,” *Education Sciences,* vol. 10, no. 10, p. 5, 2020. |
| [3] | D. J. Weiss and G. G. Kingsbury, “Application of computerized adaptive testing to educational problems,” in *Journal of Educational Measurement*, Philadelphia, NCME National Council for Measurement in Education, 1984, pp. 361-375. |
| [4] | G. G. Kingsbury and A. R. Zara, “Procedures for Selecting Items for Computerized Adaptive Tests,” *Applied Measurement in Education,* vol. 2, no. 4, pp. 359-375, 1989. |
| [5] | N. A. Thompson and D. A. Weiss, “A Framework for the Development of Computerized Adaptive Tests,” *Practical Assessment, Research and Evaluation,* vol. 16, no. 1, 2011. |
| [6] | “Computerized Adaptive Testing,” NCSBN Leading Regulatory Excellence, [Online]. Available: https://www.nclex.com/computerized-adaptive-testing.page. [Accessed 3 June 2023]. |
| [7] | H. Wainer, D. Thissen and R. Mislevy, “Testing Algorithms,” in *Computerized Adaptive Testing: A Primer, Second Edition*, New York City, Routledge Taylor & Francis Group, 2014, pp. 101-134. |
| [8] | C.-J. Lin and J. Spray, “Effects of Item-Selection Criteria on Classification Testing with the Sequential Probability Ratio Test,” ACT Inc., Iowa City, 2000. |
| [9] | H. Wainer and B. Green, “System design and operation,” in *Computerized Adaptive Testing: A Primer*, 2nd ed., New York City, Routledge Taylor & Francis Group, 2014, pp. 23-36. |
| [10] | H. Wainer and R. Mislevy, “Item response theory, Item Calibration, and Proficiency Estimation,” in *Computerized Adaptive Testing: A Primer*, 2nd ed., New York City, Routledge Taylor & Francis Group, 2014, pp. 61-100. |
| [11] | E. FitzGerald, N. Kucirkova, A. Jones, S. Cross, R. Ferguson, C. Herodotou, G. Hillaire and E. Scanlon, “Dimensions of personalisation in technology-enhanced learning: A framework and implications for design,” *British Journal of Educational Technology,* vol. 49, no. 1, pp. 165-181, 2018. |
| [12] | B. F. Green, R. D. Bock, L. G. Humphreys, R. L. Linn and M. D. Reckase, “Technical Guidlines for Assessing Computerized Adaptive Tests,” *Journal of Educational Measurement,* vol. 21, no. 4, pp. 347-360, 1984. |
| [13] | R. R. Meijer and M. L. Nering, “Computerized Adaptive Testing: Overview and Introduction,” *Applied Psychological Measurement,* vol. 23, no. 3, pp. 187-192, 1999. |
| [14] | H. Chang and Y. Z., “A Global Information Approach to Computerized Adaptive Testing,” *Applied Psychological Measurement,* vol. 20, no. 3, pp. 213-229, 1996. |
| [15] | D. O. Segall, “Computerized Adaptive Testing,” in *Encyclopedia of Social Measurement*, Arlington, Academic Press, 2005. |
| [16] | S. Dendir and M. R. Stockton, “Cheating in online courses: Evidence from online proctoring,” *Computers in Human Behavior Reports,* vol. 2, 2020. |
| [17] | Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu and X. Liu, “Automated Online Exam Proctoring,” *IEEE Transactions on Multimedia,* vol. 19, no. 7, 2018. |
| [18] | E. Gamma, R. Helm, R. Johnson and J. Vlissides, “State,” in *Design Patterns Elements of Reusable Object-Oriented Software*, Portland, Addison-Wesley, 1994, pp. 338-348. |
| [19] | E. Gamma, R. Helm, R. Johnson and J. Vlissides, “Memento,” in *Design Patterns: Elements of Reusable Object-Oriented Software*, Portland, Addison-Wesley, 1994, pp. 316-326. |
| [20] | R. C. Martin, “Chapter 2: Meaningful Names,” in *Clean Code A Handbook of Agile*, Boston, Pearson Education Inc., 2009, pp. 17-30. |
| [21] | R. Fielding, “rfc7231 Section#4,” IETF (Internet Engineering Task Force), June 2014. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc7231#section-4. [Accessed 26 March 2023]. |