

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Jaroslav MELEŠKO

FORMATIVE ASSESSMENT METHODS FOR INTELLIGENT LEARNING SYSTEMS

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
INFORMATICS ENGINEERING (T 007)

Vilnius, 2022

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2016–2022.

Supervisor

Assoc. Prof. Dr Jevgenij KURILOV (Vilnius Gediminas Technical University, Informatics Engineering – T 007), 2016–2021.

Assoc. Prof. Dr Simona RAMANAUSKAITĖ (Vilnius Gediminas Technical University, Informatics Engineering – T 007), 2021–2022.

The Dissertation Defense Council of Scientific Field of Informatics Engineering of Vilnius Gediminas Technical University:

Chairman

Prof. Dr Dalius MAŽEIKAS (Vilnius Gediminas Technical University, Informatics Engineering – T 007).

Members:

Dr Robertas DAMAŠEVIČIUS (Kaunas University of Technology, Informatics Engineering – T 007),

Prof. Dr Habil. Gintautas DZEMYDA (Vilnius University, Informatics Engineering – T 007),

Prof. Dr Arnas KAČENIAUSKAS (Vilnius Gediminas Technical University, Informatics Engineering – T 007),

Dr George Angelos PAPADOPOULOS (University of Cyprus, Informatics Engineering – T 007).

The dissertation will be defended at the public meeting of the Dissertation Defense Council of Informatics Engineering in the Senate Hall of Vilnius Gediminas Technical University at **2 p. m. on 29 August 2022**.

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.

Tel.: +370 5 274 4956; fax +370 5 270 0112; e-mail: doktor@vilniustech.lt

A notification of intention to defend the dissertation was sent on 28 July 2022.

A copy of the doctoral dissertation is available for review at Vilnius Gediminas Technical University repository <http://dspace.vgtu.lt>, at the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania) and at the library of Kaunas University of Technology (K. Donelaičio st. 20, LT-44239 Kaunas, Lithuania).

Vilnius Gediminas Technical University book No 2022-032-M

doi: 10.20334/2022-032-M

© Vilnius Gediminas Technical University, 2022

© Jaroslav Meleško, 2022

jaroslav.melesko@vilniustech.lt

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Jaroslav MELEŠKO

INTELEKTUALIOMS ELEKTRONINIO
MOKYMOSI SISTEMOMS SKIRTI
FORMUOJAMOJO VERTINIMO METODAI

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI,
INFORMATIKOS INŽINERIJA (T 007)

Vilnius, 2022

Disertacija rengta 2016–2022 metais Vilniaus Gedimino technikos universitete.

Vadovai

doc. dr. Jevgenij KURILOV (Vilniaus Gedimino technikos universitetas, informatikos inžinerija – T 007), 2016–2021.

doc. dr. Simona RAMANAUSKAITĖ (Vilniaus Gedimino technikos universitetas, informatikos inžinerija – T 007), 2021–2022.

Vilniaus Gedimino technikos universiteto Informatikos inžinerijos mokslo krypties disertacijos gynimo taryba:

Pirmininkas

prof. dr. Dalius MAŽEIKAS (Vilniaus Gedimino technikos universitetas, informatikos inžinerija – T 007).

Nariai:

dr. Robertas DAMAŠEVIČIUS (Kauno technologijos universitetas, informatikos inžinerija – T 007),

prof. habil. dr. Gintautas DZEMYDA (Vilniaus universitetas, informatikos inžinerija – T 007),

prof. dr. Arnas KAČENIAUSKAS (Vilniaus Gedimino technikos universitetas, informatikos inžinerija – T 007),

dr. George Angelos PAPADOPOULOS (Kipro universitetas, informatikos inžinerija – T 007).

Disertacija bus ginama viešame informatikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje **2022 m. rugpjūčio 29 d. 14 val.** Vilniaus Gedimino technikos universiteto senato posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.

Tel.: (8 5) 274 4956; faksas (8 5) 270 0112; el. paštas doktor@vilniustech.lt

Pranešimai apie numatomą ginti disertaciją išsiųsti 2022 m. liepos 28 d.

Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto talpykloje <http://dspace.vgtu.lt>, Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva) ir Kauno technologijos universiteto bibliotekoje (K. Donelaičio g. 20, LT-44239 Kaunas, Lietuva).

Abstract

The dissertation analyzes the development trends of intelligent multi-agent systems for personalized learning, proposes a new system's conceptual model, and explores the perspectives of individual intelligent agents, increasing the efficiency of individualized teaching. Based on the results, computer adaptive testing algorithms are proposed for assessing student achievement.

The dissertation consists of an introduction, three chapters, general conclusions, references, and lists of the author's publications on the dissertation's topic.

The introductory chapter discusses the research problem, the relevance of the thesis, describes the object of research, formulates the aim and objectives of the work, describes the research methodology, scientific novelty of the work, the practical significance of the results, and defended statements. The introduction closes by listing the author's publications and conference papers on the dissertation's topic and presenting the dissertation's structure.

The first chapter introduces the latest literature review. This chapter focuses on current trends in smart learning, including personalization and formative assessment methods. Several assessment models and algorithms are reviewed. This chapter concludes by clarifying the main objective and tasks of the thesis and a summary of the literature review findings.

The second chapter describes some review results, which were executed to get a view of students' knowledge testing media and learning style preferences. Additionally, it presents the smart multi-agent learning system's conceptual model. It is based on integrating existing learning style estimation methods, data mining and neural network solutions, and newly presented formative assessment agents.

The third chapter presents the developed computer-adaptive methodologies and algorithms based on Upper-Confidence Bound and Elo rating for formative assessment capable of suggesting a further course of study. The experiments and modeling results are presented to illustrate the benefits of the proposed algorithms compared to existing formative assessment methods.

Eleven articles were published on the dissertation's topic: two in journals with an impact factor included in the Clarivate Analytics Web of Science database, three in other peer-reviewed journals, and six in conference proceedings, five of which are included in the Clarivate Analytics Web of Science database. Six presentations were made on the dissertation's topic at national and international conferences.

Reziumė

Disertacijoje nagrinėjamos intelektualių daugiaagenčių sistemų, skirtų mokymui ir/ar mokymuisi personalizuoti, vystymosi tendencijos, siūlomas naujas sistemos konceptualus modelis, tiriamos atskirų intelektualųjų agentų perspektyvos didinant individualizuoto mokymo efektyvumą. Atsižvelgiant į rezultatus, pasiūlomi nauji kompiuteriniai adaptyvaus testavimo algoritmai, skirti studentų pasiekimų formuojamajam vertinimui.

Disertaciją sudaro įvadas, trys skyriai, bendrosios išvados, naudotos literatūros ir autoriaus publikacijų disertacijos tema sąrašai.

Įvadiniamе skyriuje aptariama tiriamoji problema, darbo aktualumas, aprašomas tyrimų objektas, formuluojamas darbo tikslas ir uždaviniai, aprašoma tyrimų metodika, darbo mokslinis naujumas, darbo rezultatų praktinė reikšmė, ginamieji teiginiai. Įvado pabaigoje pristatomos disertacijos tema autoriaus paskelbtos publikacijos ir pranešimai konferencijose, disertacijos struktūra.

Pirmasis skyrius skirtas literatūros apžvalgai. Šiame skyriuje dėmesys sutelkiamas į dabartinę intelektualaus elektroninio mokymosi tendencijas, įskaitant personalizavimo ir formuojamojo vertinimo metodus. Apžvelgti keli populiariausi studentų vertinimo modeliai ir algoritmai. Skyrius baigiamas disertacijos tikslo ir uždavinių paaiškinimu bei literatūros apžvalgos išvadų santrauka.

Antrajame skyriuje aprašomi kelių atliktų apklausų rezultatai, rodantys mokinių priklausymą konkrečiam mokymosi stiliui ar preferencijas testavimo tipui. Taip pat pasiūlytas individualizuotos kelių agentų el. mokymosi sistemos konceptualus modelis. Šis modelis remiasi esamais mokymosi stiliaus nustatymo, duomenų gavybos ir neuroninių tinklų tinkamu, naujai pasiūlytais agentais, skirtais formuojamojo vertinimo ir jo gautų rezultatų pritaikymu adaptyviam ir personalizuotam mokymuisi.

Trečiajame skyriuje pateikiami nauji adaptyvaus kompiuterizuoto mokinių žinių testavimo metodai, grįsti aukštesnio pasitikėjimo (angl. *Upper-Confidence Bound*) ir Elo reitingo algoritmais. Nurodomi eksperimentų ir modeliavimo metu gauti rezultatai, rodantys siūlomų metodų pranašumą, lyginant su tradiciniais žinių testavimo metodais.

Disertacijos tema paskelbta 11 straipsnių: 2 – mokslo žurnale, įtrauktame į *Clarivate Analytics Web of Science sąrašą*, turinčiame citavimo rodiklį, 3 – kituose mokslo žurnaluose, 6 – konferencijų straipsnių rinkiniuose, iš jų 5 – konferencijų straipsnių rinkiniuose, įtrauktuose į *Clarivate Analytics Web of Science Proceedings* duomenų bazę. Disertacijos tema perskaityti 6 pranešimai nacionalinėse ir tarptautinėse konferencijose.

Notations

Abbreviations

ANN	– Artificial Neural Network;
BYOD	– Bring Your Own Device;
CBLA	– Competency-Based Learning and Assessment;
CTT	– Classical Test Theory;
E-learning	– Electronic learning
E-portfolio	– Electronic portfolio
FSLSM	– Felder and Silverman Learning Styles Model;
ICT	– Information and communication technologies;
IoT	– Internet of Things;
IRT	– Item Response Theory;
KRS	– Knowledge Repository System;
LDA	– Latent Dirichlet allocation;
LTT	– Latent Trait Theory;
MAB	– Multi-Armed Bandit;
MIRT	– Multidimensional Item Response Theory;

MLE	– Maximum Likelihood Estimation;
M-learning	– Mobile learning;
MOOC	– Latent Dirichlet allocation;
OER	– Open Educational Resources;
PILS	– Personalized intelligent learning system;
RFID	– Radio frequency identification device;
SLE	– Smart learning environments;
STEM	– Science, technology, engineering, and mathematics;
WoS	– Web of Science;
UCB	– Upper-Confidence Bound.

Contents

INTRODUCTION	1
Problem Formulation	1
Relevance of the Thesis.....	2
Object of Research	2
Aim of the Thesis	2
Objectives of the Thesis	2
Research Methodology.....	3
Scientific Novelty of the Thesis	3
Practical Value of the Research Findings.....	3
Defended Statements.....	4
Approval of the Research Findings	4
Structure of the Dissertation.....	5
 1. LITERATURE REVIEW ON LEARNING PERSONALIZATION INTELLIGENT SYSTEMS.....	 7
1.1. Current Trends in Smart Learning.....	7
1.2 Intelligent Learning Systems.....	12
1.3. Learning Styles.....	13
1.4. Learning Analytics and Artificial Neural Networks.....	15
1.5. Formative Assessment and Feedback.....	17
1.5.1. Proportion Correct and Item Response Theory	18

1.5.2. Elo and Glicko Rating	21
1.5.3. Multi-Armed Bandit Algorithms	24
1.6. Conclusions and Formulation of the Thesis Objectives	26
2. DESIGN OF PERSONALIZED MULTI-AGENT E-LEARNING SYSTEM	29
2.1. Requirements for an Intelligent Personalized E-Learning System.....	29
2.2. Conceptual Model of Personalized Intelligent E-Learning System.....	32
2.3. Intelligent Agents	38
2.3.1. Learning Style Identification Agent	38
2.3.2. Learning Analytics Model	41
2.3.3. Learner Profile Control Agent.....	44
2.3.4. Formative Assessment Agent	46
2.3.5. Pedagogical Suitability Estimation Agent	49
2.4. Comparison of the Proposed system to Existing Solutions	52
2.5. Conclusions of Chapter 2	53
3. FORMATIVE ASSESSMENT BASED PERSONALIZED CONTENT FOR E-LEARNING ENVIRONMENT	55
3.1. Base for a Formative Assessment Agent.....	55
3.2. Upper-Confidence Bound Based Algorithm	56
3.2.1. Modeling Formative Assessment as a UCB Problem.....	57
3.2.2. Simulated Students' Experiments.....	59
3.2.3. Real Students' Assessment Methodology	59
3.2.4. Statistical Analysis	60
3.2.5. Results of UCB Formative Assessment.....	61
3.2.6. Discussion of UCB Assessment Results.....	67
3.3. Elo-Rating and Multi-Armed Bandit-Based Algorithm	70
3.3.1. Elo-Bandit Algorithm	70
3.3.2. Other Formative Assessment Algorithms.....	75
3.3.3. Simulated Students' Data	75
3.3.4. Real Students' Data	76
3.3.5. Validation Measurements and Metrics	77
3.3.6. Results of Elo-Bandit assessment experiments.....	79
3.4. Discussion of the Elo-Bandit algorithm.....	83
3.5. Conclusions of Chapter 3	85
GENERAL CONCLUSIONS	87
REFERENCES	89
LIST OF SCIENTIFIC PUBLICATIONS BY THE AUTHOR ON THE TOPIC OF THE DISSERTATION	103
SUMMARY IN LITHUANIAN.....	105

Introduction

Problem Formulation

The application of intelligent software agents and multi-agent systems in education has been actively evolving for the past two decades. Intelligent agents can adapt learning material to different learning styles of students and leverage innate predispositions for knowledge acquisition on intellectual, sensory, and emotional levels. The use of adaptive systems in education provides a multiparametric approach for improving the effectiveness of knowledge acquisition.

Some studies on personalized, intelligent, multi-agent systems were based on developing learning style-based learner profiles. However, the multidimensional nature of learner and e-learning environment requires more complex solutions rather than solely relying on students' learning styles. E-learning systems require multiple perspectives or agents for establishing a fundamental base for "optimal" (in terms of the suitability to a particular learner's profile) learning units. The use of multiple agents allows the integration of different personalization perspectives. While some outcomes have already been achieved in the field of learning style estimation and learning path generation, modern formative assessment methods have been lacking attention.

Formative assessment is specifically intended to generate feedback on performance to improve and accelerate competency acquisitions, but the problem of

choosing questions during a formative quiz is not a straightforward task and quite frequently defines the accuracy of the procedure. Accurate determination of gaps between current students' knowledge and their aims and where to go next to alleviate those deficiencies is required to foster the benefits of formative assessment and smart-education systems.

Relevance of the Thesis

Standardized education in schools and universities is a well-established model of learning. However, it barely accounts for individual students' traits and needs. Quite frequently, e-learning involves delivering information without feedback to the learner. Feedback in the form of a mark, as provided by traditional tests, has limited use to a student as it is not congruent with good feedback practices. Developing a computer-based educational environment with integrated multi-agent systems potentially allows tuning the learning experience. The investigated computer-based methods and algorithms solve this problem by quick identification of the lacking areas of knowledge and thorough data exploration to assist the teacher and the learner. The results of this dissertation and the developed algorithms directly contribute to the development of new smart-learning platforms.

Object of Research

The object of the research – formative assessment algorithms and their integration in multi-agent e-learning systems.

Aim of the Thesis

The thesis aims to improve the adaptability of e-learning systems by developing an algorithm for formative assessment, that shortens test length and does not require calibration, and to integrate it with dominant course personalization technologies.

Objectives of the Thesis

The following objectives were raised to solve the stated problem and reach the aim of the thesis:

1. To analyze the applications and effectiveness of learning personalization technologies such as learning styles and formative assessment in intelligent learning systems.

2. To design a conceptual model of a multi-agent learning system based on integrating different aspects for e-learning personalization.
3. To develop adaptive, student-oriented formative assessment algorithms for e-learning systems.
4. To compare the developed formative assessment algorithms to Multi-dimensional Item Response Theory (MIRT) based methodologies in the sense of students' workload and knowledge gap estimation efficiency.

Research Methodology

Numerical (simulations and optimization) and experimental methods were applied in the study. The multi-agent learning system and its agents were proposed, applying data mining, semantic clustering, and neural networking concepts. The formative assessment algorithms have been developed using Upper-Confidence Bound (UCB) and Elo rating methods and assessed in the context of MIRT and Maximum Likelihood Estimation (MLE) models.

Scientific Novelty of the Thesis

The aspects of scientific novelty in the field of Informatics Engineering are as follows:

1. Formative assessment was modelled as an exploration-exploitation problem, and a UCB algorithm was adapted for student assessment in the context of classical test theory, with positive results indicating a strong reduction in test length.
2. A novel algorithm inspired by Elo and UCB has been developed and applied for the formative assessment of students in the context of latent trait theory. This method combines the advantages of IRT and traditional testing approaches: it does not require item calibration and reduces test length.

Practical Value of the Research Findings

The proposed multi-agent systems and formative assessment algorithms can be used to establish new smart-learning platforms congruent with the established strategic visions of Competency-Based Learning and Assessment (CBLA). The combination of the developed methods with presently widespread mobile devices

and computer-assisted remote education has the potential to provide timely, personalized, and automatized feedback to the learner while addressing the issues such as test creation complexity and long test times.

Defended Statements

1. The proposed smart e-learning system has a wider range of e-learning adaptability and personalization level as it integrates novel, student-oriented formative assessment algorithms.
2. Proposed assessment algorithms prioritize feedback to students and comply with Hattie's three-step effective feedback model.
3. Modelling assessment as a UCB problem enables the identification of the weakest topics of knowledge for a student faster than random or linear traditional tests.
4. Elo-Based formative assessment can offer test length reduction and accuracy comparable to item response theory-based methods without the need for pretesting.

Approval of the Research Findings

Research results were published in eleven articles: two in journals with an impact factor included in the *Clarivate Analytics Web of Science* database (Meleško 2019; Meleško 2021), three in other peer-reviewed journals (Meleško 2017a; Kurilov 2017; Meleško 2018a), six in conference proceedings (Kurilov 2016a; Meleško 2016; Kurilov 2016b; Meleško 2017b; Meleško 2017c; Meleško 2018b), five of which were included in the *Clarivate Analytics Web of Science Proceedings* database.

Research results on the thesis subject have been presented at four scientific conferences in Lithuania and abroad:

- 9th annual international conference of education, research, and innovation “*ICERI 2016*,” 2016. Seville, Spain.
- 4th Workshop on Advances in information, electronic and electrical engineering “*AIEEE 2016*.” Vilnius, Lithuania.
- 10th annual international conference of education, research, and innovation “*ICERI 2017*,” 2017. Seville, Spain.
- 8th international conference on web intelligence, mining, and semantics “*WIMS 2018*.” Novi Sad, Serbia.

Structure of the Dissertation

The dissertation includes an introduction, three chapters, general conclusions, a list of references with a separately presented list of scientific publications by the author, and a summary in Lithuanian.

The dissertation consists of 128 pages, 27 figures, 12 tables, 22 equations, and 182 references.

Literature Review on Learning Personalization Intelligent Systems

This chapter focuses on the review of the latest results in the application of intelligent multi-agent systems in education, including methods for personalization of learning and formative assessment. Various adaptive models and algorithms are reviewed. This chapter concludes with the clarification of the main objective and tasks of the thesis, as well as a summary of the literature review findings.

Four scientific publications were published on the topic of the section (Kurilovas, Krikun, & Melesko, 2016; Melesko & Kurilovas, 2016; Meleško & Kurilovas, 2018; Meleško, Kurilovas, & Krikun, 2017).

1.1. Current Trends in Smart Learning

With the ever-increasing technological complexity of human civilization, each new generation must learn more to successfully contribute to society. Thus, increasing learning efficiency is important and will only become more crucial with time. There are three principal criteria for learning (Schunk, 2012). The first criterion is that learning involves a change in behaviour or capacity for behaviour. It is also important to note that learning is not observed directly but through its outcomes. Learning and demonstration of newly gained knowledge or proficiencies are often separate. The second criterion of learning is that learning endures over

time. There is always a possibility of forgetting and loss of proficiency. However, the change in behaviour or capacity for behaviour must persist past the removal of the cause of that change. The third criterion is that learning occurs through experience, separating learning from hereditary changes in behaviour, such as crawling and standing up of children with maturation. This separation is not clear-cut, for example, the predisposition to learn languages in children. As new technology and knowledge about the natural world become available, it is put to use to improve learning efficiency. This is also true for information technology, computer science, and engineering. The annual number of publications on the subject in the Web of Science database and their distribution across research categories can be seen in Figs. 1.1 and 1.2.

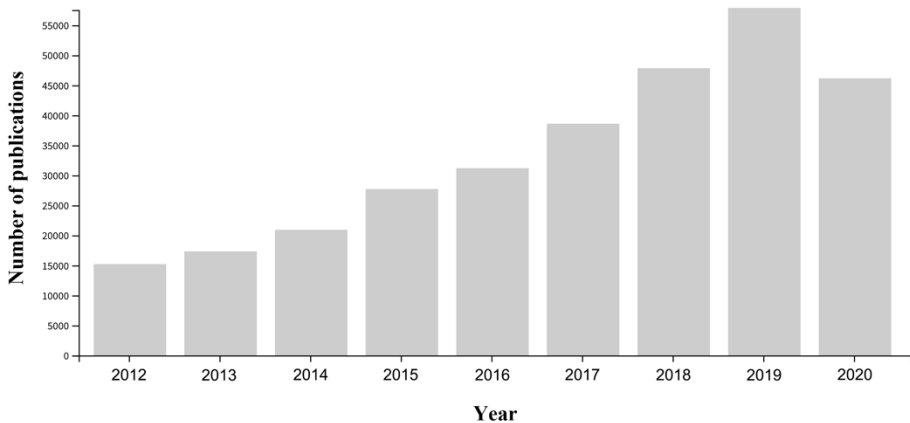


Fig. 1.1. Annual number of publications in the Web of Science database on education and learning in information technologies, computer science, and engineering categories

The annual trend of computer-assisted learning publications in the period from 2012 to 2020 shows a stable increase in the academic interest in the subject. Most research falls into computer science artificial intelligence, computer science theory methods, and computer science information system categories. We can see that education is highly influenced by information and communication technologies (ICT). Learning enhanced by these technologies is often called e-learning (electronic learning) or smart learning.

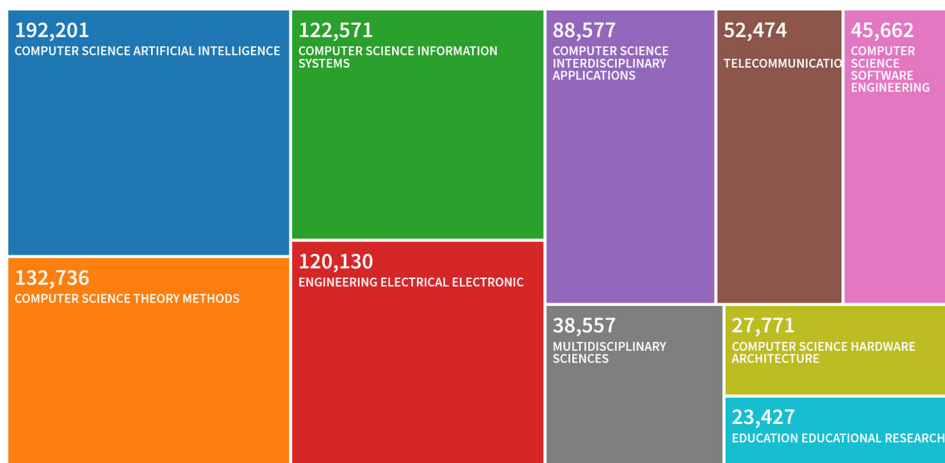


Fig. 1.2. Categories and numbers of publications on education and learning in the Web of Science database

A fine-grained look at the state of the field can be obtained from the topical clustering of publications in the Scopus database from 1989 to 2019 (Chen, Zou, Xie, & Wang, 2021). Chen et al. analyzed 555 publications on the subject of “smart learning” and “smart classroom.” Frequency analysis revealed the most used and important keyword in smart learning studies. “Content” was the most frequently used word which appeared in 97 publications (17.5%), overtaking even “learner,” which appeared in 96 publications (17.3%). Other frequently used keywords were “user” (16%), “university” (16%), “framework” (15%), “network” (15%), “service” (15%), and “mobile” (15%). Based on these statistics, most research efforts focus on the content that is served to students and learners.

Latent Dirichlet allocation (Campbell, Hindle, & Stroulia, 2015) was used to analyze abstracts of the 555 publications and organize them into 22 topics. The results of the analysis are given in Table 1.1. These topics were subjected to hierarchical cluster analysis using a complete-linkage agglomerative algorithm (McCabe, Sneath, & Sokal, 1975). Hierarchical clustering of smart learning topics reveals several cross-topic research vectors, such as course content management through feedback and assessment in massive open online courses (MOOC), collaborative inquiry-based learning with the help of context-aware devices, and attendance and attention recognition with the internet of things (IoT) cloud-connected devices.

Table 1.1. Frequency of the key words used in smart learning studies with Mann-Kendall trends of their popularity (Chen et al., 2021)

Topic label	Topic proportion, %	Trend
Interactive and multimedia learning	5.38	↑
STEM education	3.82	↑↑↑
Attendance and attention recognition	4.15	↑
Blended learning for smart learning	6.64	↑↑↑
Affective and biometric computing	3.54	↑↑↑
Collaborative inquiry-based learning	5.09	↑
Feedback and assessment	3.61	↑
Infrastructure for smart learning	3.55	↑
Smart learning analytics	3.66	↑↑↑↑
Semantics and ontology	3.94	↑
Software engineering for e-learning systems	4.5	↑↑↑↑
Apps for smart learning	3.01	↑↑↑
Augmented and virtual reality	3.95	↑
Sensor and wireless	3.17	↑
Context-aware services	3.42	↑
Mobile learning	8.26	↑
MOOCs and course content management	5.43	↑↑↑
Smart learning for special needs education	4.81	↓↓
IoT and cloud computing	6.28	↑↑↑↑
Ecosystem and ambient intelligence	5.7	↑↑↑
English and language learning	4.54	↑
RFID and laser techniques	3.55	↓

As argued in the following sections, feedback and formative assessment are core principles of smart and effective education. In the context of MOOC, the number of students might be too high to provide timely individualized tutoring

and feedback for instructors and course facilitators (Shatnawi, Gaber, & Cocca, 2014). Thus, it is unsurprising that developing technologies to imbue these learning environments with intelligent solutions to act in lieu of educators is an active research direction resulting in the close proximity to “feedback and assessment” and “MOOCs and course content management” topics.

“IoT and cloud computing” and “attendance and attention recognition” topics reveal another connection, which can be explained by the use of IoT devices in concert with face recognition software. Video streams from IoT devices can be used to identify learner attributes such as gender, age, and identity to facilitate learning (Qi, Liu, & Schuckers, 2018).

The distribution of annual proportions of smart learning topics in the last twenty years has its own trends. There is a decrease in the interest in radio frequency identification devices (RFID) and laser techniques, interactive and multimedia learning, and smart learning for special needs education. Also, collaborative inquiry-based learning, context-aware services, and mobile learning are exhibiting a downward trend. MOOCs and course content management, software for engineering for e-learning systems, IoT, and cloud computing are increasingly investigated in the scientific literature.

The use of learning technologies is not limited to academia; however, the adoption and use of computer-assisted smart learning are not uniform across different industries. The prevalent use of technologies and terminology in publications was researched by Giannakos and colleagues (Giannakos, Mikalef, & Pappas, 2021). Their research presents the analysis results of 47 papers from SpringerLink, Wiley, ACM Digital Library, IEEE Xplore, Science Direct, SAGE, ERIC, AIS eLibrary, and Taylor & Francis databases from 2010 and 2019 on organizational e-learning. The information technology sector is the leading other industry in terms of the adoption of smart learning technologies. Research suggests that this might be a purely historical precedent. A study by Yanson and Johnson failed to measure the impact of e-learning on the acquisition of different types of knowledge (hard skills vs. soft skills and procedural vs. declarative knowledge) (Yanson & Johnson, 2016). There is also a clear prevalence of online and web technologies with the emergent trend for social learning. Predominant objectives of studies appear to be user satisfaction, use, fostering learning, and performance. These objectives reveal two focal points of research. The first is engaging students, so they are satisfied and continue learning using the system. The second is fostering the learning and performance of systems’ users.

In conclusion, the body of research on computer-assisted smart learning is ample and is growing. However, there are several exceptions to this general trend, such as studies on topics of RFID, special needs, and collaborative education. The areas of the most active research are software engineering for e-learning systems and affective and biometric computing. Within those topics, the adoption of smart

learning in the industry is led by the information technology sector. Most industry-related research is focused on user engagement and fostering of learning. Research also indicates limited interest in large volume data mining for educational purposes and social collaborative learning.

1.2. Intelligent Learning Systems

Intelligent learning systems are a way to integrate educational technologies into a comprehensive environment that fosters education. They increase the functionality of conventional hypermedia by combining content with smart personalization. Transferring pedagogical and technological considerations into software design specifications is not straightforward. The adaptive techniques are usually based on the user's needs, goals, interests, learning styles, and knowledge level.

The research in this domain can be traced back to two ideas: hypertext and user modelling. In abstract terms, the main components of an adaptive educational system can be generalized as the source of adaptive learning, the target of adaptive learning, and the pathways of adaptive learning (Ennouamani & Mahani, 2018). If the goal is to adapt to a learner, the system must give more importance to their characteristics. Learners' characteristics are encapsulated in the learner's model, which is an essential component of the adaptive education system (Brusilovsky & Millán, 2007). Although such models are often only implicit, they can also contain such information as learning style, goals, preferences, skills, emotions, the level of knowledge, history, navigation, and test results (Verbert et al., 2012). The adaptation target is the learning material. The system can present courses in a different way to each student, provide unique learning supports, adapt media format to user's needs, and even alter the user interface and the amount of collaboration (Sensuse, Hasani, & Bagustari, 2020). Pathways of adaptive learning are methods used to break down courses into segments and then insert, remove, dim, or alter fragments to provide the learner with the part they need. These principles are also summarized as three models: the learner, the domain, and the adaptation.

In more practical terms, a recent review of intelligent tutoring systems identified a number of actually used learner characteristics and personalization techniques (Mousavinasab et al., 2021). The authors found that the learning performance, knowledge level, and behaviour in the learning path were most often used characteristics, followed by learning styles and learning preferences. Other factors, such as emotional states, culture, and intelligence level, were rarely used. As for techniques, condition-action rule-based reasoning, data mining, and Bayesian-based techniques are the most popular. Followed by fuzzy logic, natural language processing, intelligent agents, and ANN-based techniques.

Methods for evaluating learning systems can be organized into groups: based on system performance, learner performance, learner experience, and expert evaluation. A literature review of 53 papers showed that the student performance metric was used the most often, i.e., in 75% of studies (Mousavinasab et al., 2021). The system's performance metric was used in 43% of studies and student experience in 23% of cases. In 15% of studies, three metrics were used. Criteria used for evaluating the system performance included accuracy, precision, sensitivity, adaptiveness, reliability, recognition rate, usability, and mean square error (MSE).

In Lithuania, during the past ten years, several dissertations on adaptive teaching systems have been issued. Burbaitė's dissertation focuses on using advanced generative learning objects (AGLO). The author tries to improve generative learning objects (GLO), learning content that is produced and handled automatically. The pedagogical effectiveness of AGLO combined with a robot programming exercise was evaluated using the "engagement levels" methodology and showed an increase in learning improvement from 6 to 15 percent (Burbaitė, 2014). Žilinskienė's work examines the application of ant colony optimization method for choosing an optimal learning path based on learning styles. The author reports a shorter learning time for students who used the recommendations provided by the system compared to those who didn't use it or used it in less than 30% of cases (Žilinskienė, 2013). Juškevičienė's dissertation explores the use of web 2.0 tools in a learning style-based recommendation system. The author reports positive results supported by the expert evaluation of her system (Juškevičienė, 2014). Similarly, Urbonienė relies on expert evaluation to support learning styles based on personalization in her dissertation (Urbonienė, 2014). Žilinskienė's, Juškevičienė's, and Urbonienė's results are at odds with the recent critique of learning style-based personalization presented in Section 1.3. However, Juškevičienė's and Urbonienė's judgment is based on expert evaluation and is in line with the observation that learning styles are popular and well-regarded by educators despite shaky empirical support.

1.3. Learning Styles

Coffield's critical review of learning styles identifies 30 different dichotomous learning styles (Coffield, Moseley, Hall, & Ecclestone, 2004), and the list is still in use today (Kirschner, 2017). A review by Newton found that 89 percent of papers reviewed in his study endorsed learning styles in higher education (Newton, 2015), and Howard-Jones determined that 95 percent of teachers in Great Britain believed that people learn better in their preferred style, for example visual, auditory or kinesthetic (Dekker, Lee, Howard-Jones, & Jolles, 2012; Howard-Jones, 2014).

Well-established models exist for the estimation of a learning style. Felder and Silverman Learning Styles (FSLSM) (Felder Richard M. & Silverman Linda K., 1988) are usually identified with the help of the Soloman–Felder questionnaire (Soloman, Carolina, & Felder, 2012). Felder and Silverman Learning Styles Model (FSLSM) was developed to be the most suitable for STEM (Science, Technology, Engineering, and Mathematics) subjects and e-learning. The dedicated psychological questionnaire is used to explore students' learning styles according to FSLSM. FSLSM classifies students according to where they fit on several scales depending on the ways they receive and process information. Students are classified by information type. They are either sensory (SEN) – concrete, practical, oriented towards facts and procedures, or Intuitive (INT) – conceptual, innovative, oriented towards facts and meaning. Students are also separated by the preference for a sensory channel for knowledge acquisition. Visual (VIS) students prefer visual representations of presented material, e.g., pictures, diagrams, and flow charts. Students with Verbal (VRB) learning styles prefer written and spoken explanations. Students also process information differently. Active (ACT) students learn by trying things out and working with others, while Reflective (REF) students learn by thinking things through and working alone. Finally, students understand the material in two ways. Sequential (SEQ) students learn linearly, orderly, in small incremental steps, while Global (GLO) students are holistic systems thinkers and learn in large leaps.

There is, however, growing opposition to the learning styles approach of learning personalization. As early as 2003, some studies questioned the reliability and consistency of measured learning styles, finding them only a weak predictor of study results (Veenman, Prins, & Verheij, 2003). In particular, with relevance to Felder and Silverman's learning styles model, self-reported preference for visual or verbal learning style by students was found to be only weakly related to actual performance (Massa & Mayer, 2006; Rogowsky, Calhoun, & Tallal, 2015). However, it must be noted that students are not reliable in self-assessment, even when reporting time spent on homework (Rawson, Stahovich, & Mayer, 2017). The study by Rawson and colleagues has found no correlation ($r = -0.16$) between time spent on homework as reported by students and course grade, while there was a correlation between measured time and the grade ($r = 0.44$). It is unclear from these results if it is the learning styles model or the measurement methods that are at fault. Kirschner offers a compelling critique of the learning styles approach to learning personalization in agreement with Pashler, Westby, and Newton (Kirschner, 2017; Newton & Salvi, 2020; Pashler, McDaniel, Rohrer, & Bjork, 2008; Westby, 2019).

1.4. Learning Analytics and Artificial Neural Networks

Learning Analytics and educational data mining techniques are rising in popularity. Both terms are used to describe common technologies and techniques, although in recent years, “learning analytics” has been preferred (Cristobal Romero & Ventura, 2020). The use of ANN (artificial neural network) techniques in learning analytics was an uncommon approach until recently. In an early literature review, Jindal and Borah reported three papers using ANN (Jindal & Borah, 2013), and a review by Husain and Rashid two years later presented just four papers using ANN (Shahiri, Husain, & Rashid, 2015). Herrera-Flores and colleagues, in the 2019 review, however, identified 41 papers on the subject published between 2015 and 2019 (Hernández-Blanco, Herrera-Flores, Tomás, & Navarro-Colorado, 2019). Finally, Okewu et al. analyzed 190 papers on ANN-based educational data mining and found that although it is used for student retention, student progression, and cost-saving, several challenges (hardware, theoretical, and training) still remain (Okewu, Adewole, Misra, Maskeliunas, & Damasevicius, 2021).

Artificial neural networks are computer systems inspired by biological neural systems (Br, 2003). A brief description of the main ANN components and terminology is as follows. It is a connectionist intelligence model developed to realize such abilities as learning from prior information to produce and discover new knowledge. This is done through generalization based on existing data. Such a process is called training. ANN is composed of elements, or neurons, which are connected into a network by links or synapses. In the ANN model, each connection has a numeric weight attached to it, and the collection of these weights stores knowledge acquired by the network. The neurons act like processing components; they receive signals from other neurons or directly from input and compute a result by combining and normalizing these signals. This process roughly approximates the behaviour of biological neural networks.

The ANN operation of is primarily controlled by the number of neurons, the bias of neurons, the transfer function of neurons, the connections between neurons, and the weights associated with neuron connections.

ANN is composed of elements, or neurons, which are connected into a network by links or synapses. In the ANN model, each connection has a numeric weight attached to it, and the collection of these weights stores knowledge acquired by the network. The neurons act like processing components. They receive signals from other neurons or directly from input and compute a result by combining and normalizing these signals. This process roughly approximates the behaviour of biological neural networks.

The operation of ANN is primarily decided by the number of neurons, the bias of neurons, the transfer function of neurons, the connections between neurons, and the weights associated with neuron connections.

Vector $\mathbf{p}$ of R scalar inputs p are multiplied by the weight coefficients w associated with the connection used to pass the input to the neuron (Hagan M.T. & Beale, 1996). The weighted inputs are then summed together with bias b of the neuron by the summer function to form the net input n :

$$n = w_1p_1 + w_2p_2 + \dots + w_Rp_R + b. \quad (1.1)$$

In matrix form:

$$n = \mathbf{Wp} + b. \quad (1.2)$$

The net input is then fed into the transfer (or activation) function, which produces the scalar neuron output. The neuron output can be written as:

$$a = f(\mathbf{Wp} + b). \quad (1.3)$$

The transfer function can be either a linear or a nonlinear function taking the summer result in as in input a return a value to be passed to the next neuron layer. Transfer functions are chosen based on the task at hand, their mathematical properties, and to normalize neuron output. Log-sigmoid transfer function takes as input any number between plus and minus infinity and ‘squashes’ it into the range 0 to 1. The log-Sigmoid function can be expressed as:

$$a = (1 + e^{-n})^{-1}. \quad (1.4)$$

The Log-Sigmoid function is often chosen for use in multi-layer neuron networks because they are trained using a backpropagation algorithm, and the function is differentiable.

The individual neurons are first organized into layers, which are in turn organized in networks. Usually, each input is connected to each neuron in the layer. The weights for the layer are stored in the weight matrix, where each row holds input weights for one neuron in the layer. The layer also includes the bias vector b , the summers, the transfer function, and the output vector. The output vector is composed of outputs of the layer’s neurons. While not all neurons in the layer must share the same transfer function, it is common to use one function for simplicity. The layers are organized into a network. The layer that outputs the final result is called the output layer. The number of neurons in the output layer is usually determined by the problem that is being solved by the network and the desired number of outputs. Other layers are called hidden layers.

There are two proposed ways to classify the papers in the neural network-assisted learning analytics field. The first consists of eleven categories (Cristbal Romero & Ventura, 2010). The second, published in 2018, builds on previous work and expands the number of categories to thirteen (Bakhshinatagh, Zaiane, ElAtia, & Ipperciel, 2018). These categories are predicting student performance, detecting undesirable student behaviours, profiling and grouping students, social

network analysis, providing reports, creating alerts for stakeholders, planning, and scheduling, creating courseware, developing concept maps, generating recommendations, adaptive systems, evaluation, and scientific inquiry. Herrera-Flores review identifies that the largest number of publications, twenty-two papers, belong to predicting student performance category, with most papers reporting positive results (Hernández-Blanco et al., 2019). The second most popular subfield was detecting undesirable student behaviours with ten entries. Both these applications of neural networks are relevant to personalized adaptive learning. When compared to other machine learning techniques, the studies report superior performance, with review studies mostly supporting these claims (Coelho & Silveira, 2017; Hernández-Blanco et al., 2019). More rigorous replication studies usually report lesser advantages gained from deep learning without disproving their superiority (Doleck, Lemay, Basnet, & Bazelais, 2020).

1.5. Formative Assessment and Feedback

Formative assessments have been proposed to make education more accessible and more effective (Bennett, 2011; Black & Wiliam, 1998; Dunn & Mulvenon, 2009; T. Guskey, 2005). The distinction between summative and formative roles of assessment was first proposed by Scriven (Scriven, 1967) and then applied to students by Bloom (B.S. Bloom, 1971; Benjamin S Bloom, 1969). Formative assessment is specifically intended to generate feedback on performance to improve and accelerate competency acquisition as opposed to summarizing the achievement status of a student (Moss & Brookhart, 2019; Sadler, 1998). Any learning activity has potential value as a formative assessment, from oral discourse to conventional quizzes (Gareis, 2007). It should be stressed that seminal Bloom's paper "The 2 Sigma Problem" also reports one sigma improvement for the groups of students who were given formative assessments; that is a massive improvement upon results produced by learning style-based personalization (Benjamin S. Bloom, 1984).

Three core principles form the basis for formative assessment (Assessment Reform Group, 1999). Firstly, formative assessment should be viewed as an integral part of instruction, and it should be used in real-time to guide the learning process. The material provided to students should depend on their current state of knowledge and understanding. Secondly, formative assessment fosters student involvement. Students are not punished for their mistakes since assessment does not affect the final grade. Instead, they can use assessments for self-guidance and monitor their progress towards learning objectives. Thirdly, formative assessment requires constructive feedback. Feedback is the key element in formative assessment (Sadler, Royce, 1989); it is claimed to be the single most powerful factor in enhancing achievement (Robert J Marzano, Pickering, & Pollock, 2001). Meta-

analysis of the effects of feedback shows a weighted mean effect of 0.8 on immediate post-test results and a weighted mean of 0.35 for delayed post-tests (Azevedo & Bernard, 1995; Shute, 2008); however, a critique of these results exists (Bennett, 2011). One of the goals of the educational process is to identify the gaps between what is known and what is aimed to be known (Hattie & Timperley, 2007). Feedback, which is separated from competency demonstration by days or weeks, as it happens with paper-based tests that have to be manually graded, has diminished value (Gareis, 2007). Feedback in the form of a mark as provided by traditional tests has limited use to a student as it is not congruent with good feedback practices (Nicol & MacFarlane-Dick, 2006). A mark does not specify the deficiencies in the student's knowledge that they should address in the future (T. R. Guskey & Bailey, 2001). Another strength of formative assessment is that it can aid in using students' strengths and weaknesses to frame learning goals and monitor their progress towards them (Tomasik, Berger, & Moser, 2018).

Formative assessment is a powerful tool to explore students' knowledge and suggest further learning material in accordance with Hattie's feedback model (Hattie & Timperley, 2007). First, the instructor forms an item pool divided into topics or competencies – "Where am I going?" Second, students partake in formative assessment – "How am I doing?" Third, the results of formative assessment can be used to direct further instruction – "Where to next?"

To answer the "How am I doing?" question, one can employ several statistical models for ability estimation. The two most widely used models are Classical Test Theory (CTT) and Item Response Theory (IRT). Recently Elo Rating and Glicko models are gaining traction in educational systems.

1.5.1. Proportion Correct and Item Response Theory

A simple and widespread way to quantize a student's performance on a test or measure the difficulty of a test item is the proportion correct method. It is a ratio of correctly answered items to a total number of items seen by the student or a ratio of the number of correct answers to total answers to an item, respectively. This method is easy to use, and the scores can be computed even by hand. However, its limitation is that estimated this way, students' ability levels can be only compared to those who have taken the same test. The simple ratio does not account for the varying difficulty of the items or the varied student's proficiencies.

Item Response Theory (IRT) is an alternative to the classical proportion correct method, which involves more computation, and thus becomes practical only with the use of a computer to calculate the scores (Lord & Novick, 2008). However, this method accounts for item difficulty. It can measure the change in students' knowledge when the same student is exposed to different test items on separate occasions or to compare different students who attempted different test items

(Crocker & Algina, 2008). The increase in the availability of computational resources in testing leads to greater interest in IRT and its wider applications (Jones & Thissen, 2006). There are three notable IRT models: 1PL (the one-parameter logistic, also known as the Rasch model), 2PL (two-parameter logistic), and 3PL (three-parameter logistic).

Item Response Theory is also known as Latent Trait Theory. There is an underlying variable trait in many educational situations that is of interest. It can be the intelligence of a student or proficiency at some class of tasks. Both directly unobservable human attributes can be called latent traits or abilities. The goal of IRT is to quantify and measure such latent traits. We can assume that every examinee responding by undertaking a test has some amount of the measured ability represented by a numerical value or score S that places them somewhere on the ability scale. For each score S , there is a probability that the student will answer an item correctly $P(S)$, which takes the shape of the logistic curve. The probability of obtaining the correct answer nears zero for low scores, and for high scores, it approaches one. In IRT, the curve that describes the relationship between the score and the probability of the correct answer is sometimes also called the item characteristic curve.

Several properties can describe item characteristic curves: guessing or guess factor, item discrimination, and difficulty. The difficulty parameter indicates where the item will be most effective in gauging students' knowledge. If an easy item is served to competent students, we can expect all of them to answer it correctly in the process, providing us with very little information. If, on the other hand, a difficult item is presented to the same students, some of them can be expected to fail to answer, which allows separating the most knowledgeable students in the group.

The discrimination parameter shows how well the item can differentiate between examinees below and above its difficulty threshold. It is the slope of the logistic curve. If the curve is steep, then the difference in the probability of correct answers between a knowledgeable student and an ignorant one will be large. If the curve is flat, then students with relatively large ability differences will have a similar chance to answer the item correctly, and the item will have a weak discrimination parameter.

Finally, the guess factor shows the probability of answering the question correctly, which is independent of the student's knowledge. For example, a quiz item with four possible options will have a guess parameter $P=0.25$.

The equation that describes the 3-parameter item characteristic curve can be seen below.

$$P = g + \frac{1 - g}{1 + e^{-d(A - D)}}, \quad (1.5)$$

where g is the guessing parameter, d is the discrimination parameter and D is the difficulty of an item, and A is the student's ability level. The formula can be reduced to a two-parameter model if g is set to zero. The Rasch model or one-parameter expression can be acquired if d is a constant equal for every item.

At the moment of creating parameters of the item characteristic curve of each test, the item is unknown and thus has to be experimentally estimated with the help of real students. The process is called calibration. One way to calibrate item parameters is by using an iterative maximum likelihood estimation using the Birnbaum paradigm (Birnbaum, 1968). After several students attempt to answer the items and the results are recorded, it is possible to estimate the parameters of each item. For that, an ability level must be assigned to each student. A raw test score can be used as a starting point. With assumed student knowledge and observed answers as input, it is possible to estimate item parameters that would most likely result in the observed outcome under the model assumptions. After the most likely item parameters are found, student ability level can be re-estimated. With the newly re-estimated student abilities as input, item parameters can be estimated again with higher precision. In such an iterative manner, item parameters can be found. A detailed illustration of this process can be found in Baker's lectures (Magnus, 2018). A more analytic examination of IRT mathematical models is provided by Lord, Baker, and Kim (Baker & Kim, 2004; Lord, 2012).

In addition to traditional linear tests with fixed test items, adaptive tests based on Item Response Theory (IRT) have been proposed (Lord, 2012) and implemented (Y. M. Huang, Lin, & Cheng, 2009). In IRT computer-adaptive tests, each test item is dynamically chosen based on the student's answers. This can improve the precision of testing by adapting to the knowledge demonstrated by each student and shortening test times (McDonald, 2002). These advantages come at the cost of introducing several new issues. Most problematic is a greater test creation complexity because of the necessity to calibrate the question pool and the premise that different questions measure one common trait (Wainer, 2001). This premise is not always useful when assessing separate competencies, even in the context of Multidimensional Item Response Theory (MIRT) (Reckase, 1997), which can even further exacerbate the challenge of item calibration. In practice, the main drawbacks remain the complexity and time required to create a test, which may dissuade teachers from using it as a learning tool.

A systematic review of item response theory by Min and Aryadoust found that in terms of research, nearly every new issue of major measurement journals publishes at least one paper on IRT (Min & Aryadoust, 2021). Most applications are in the field of education for summative and formative assessment (Lee, Kim, Choi, & Kang, 2020; Mizumoto, Sasao, & Webb, 2019). Multidimensional IRT models do not necessarily offer a superior fit to unidimensional models (Hohensinn & Kubinger, 2011). But it is also often used in psychometric

healthcare assessments to estimate depressive and anxiety symptoms in refugees (Schlechter et al., 2021) or to assess the internet gaming disorder (Schivinski, Brzozowska-Woś, Buchanan, Griffiths, & Pontes, 2018).

1.5.2. Elo and Glicko Rating

The Elo rating system was originally developed for estimating the skill level of chess players and matchmaking (Elo, 1987). But ever since, it has been adapted for the use in education (Tyner, 1977; K. Wauters, Desmet, & Noortgate, 2011). Elo rating is a statistical model of ability where the skill of a player can go up and down based on wins and losses weighted by the strength of the opponent.

The Elo system used in chess includes matching opponents of comparable skill, competitive ranking, setting skill-based cut-off values for tournaments, and organizing player placement in the tournament tree so that strong and famous players do not play each other too early to make finals more appealing to the observers (Glickman, 1995). The Elo system is also employed in multiplayer computer games where similar challenges to chess exist. Players need to be matched by their skill and may wish to see their ranking as compared to other players. Elo rating can be combined with other metrics, such as ping and time spent in a matchmaking queue, to arrange a favourable match for participants (Véron, Marin, & Monnet, 2014). Elo rating can also be used for estimating students' proficiency level and task difficulty level and for selecting tasks most suitable for each student. Elo rating is being explored in the educational setting in theory and practice (Pelánek, 2016).

Within the system, each participant is assigned a numerical rating, and in the case of educational applications, so are the test items. The value of the initial skill estimate in chess can be calculated based on the average skill ratings of fellow players competing for the tournament and player placements (FIDE, 2020), or in a computer game, a player may be simply given a starting, average rating which is then adjusted based on performance (Véron et al., 2014). The skill level of the player is always increased if they win a match and decreases if they lose; however, the degree of change is dependent on the strength of the opponent or task difficulty in the educational setting. Losing to a strong opponent or failing at a difficult task will lessen the point loss. However, losing to a comparably weaker player will indicate that the current rating is wildly incorrect and trigger a large change in skill evaluation. Thus, the change in skill estimation post-match is a function of a probability of a win. The expected probability P that a player wins or a student successfully completes the task is then given by the logistic function:

$$P = \frac{1}{1 + 10^{\frac{Elo_o - Elo}{400}}}, \quad (1.6)$$

where Elo is the Elo rating of the player in question, and Elo_o is their opponent. If educational outcomes are modelled, Elo_o can be the difficulty rating of the task. The denominator in the exponent, 400, is the width of the Elo rating system. As follows from the formula, every 400-point advantage over the opponent signifies a tenfold increase in the probability of victory. Naturally, the width can be chosen freely. The value chosen for chess rankings helps the Elo rating to be human-readable. The same is true for the base of the exponent, which is sometimes changed to Euler's number e in the educational context.

After the outcome of the match is known, the Elo ratings can be adjusted. The winner's rating is increased while the loser loses points. The magnitude of change depends on the previously calculated probability. The new Elo rating of the player Elo' can be calculated by the following formula:

$$Elo' = Elo + k(O - P), \quad (1.7)$$

where O is the actual outcome of the match, 1 for victory, 0 for loss, and k is a constant that regulates how quick the system is to react to a change in skill. The value chosen in chess for k is often 32. As the second formula indicates, the change in rating estimation is larger for unexpected outcomes. Such outcomes imply incorrect estimations. Finally, the measure of the system's sensitivity can be made variable to reflect the uncertainty of rating estimation. As players compete in more games, the estimate can be expected to be more statistically accurate, as their ratings should be adjusted in smaller increments (Pelánek, 2016).

Glicko formula is a variation of the Elo system that adds a standard deviation to rating estimation to improve its accuracy (Glickman, 1995). New players get an agreed-upon starting value, then to calculate Rating Deviation, RD , for a player after downtime, Glicko suggests the following formula:

$$RD'_P = \sqrt{RD_P^2 + c^2 t}, \quad (1.8)$$

where RD' is the new rating deviation, RD previously held rating deviation, t is the time elapsed since the player last played, and c is a constant that controls the impact of time on rating.

The win probability for a player is calculated similarly to the Elo rating, where G and G_o are Glicko ratings of the player and their opponents:

$$P = \frac{1}{1 + 10^{\frac{G_o - G}{400g}}}. \quad (1.9)$$

The new parameter g is introduced that accounts for uncertainty. And is calculated with the following formula:

$$g = \frac{1}{\sqrt{\left(1 + 3q^2 \frac{RD_o^2}{\pi^2}\right)}}, \quad (1.10)$$

where q for traditionally employed in chess constant is:

$$q = \frac{\ln 10}{400} \approx 0.00576. \quad (1.11)$$

The post-match Glicko rating is updated using the same formula as the Elo rating, except for the k constant, which is now calculated based on rating deviation:

$$G' = G + qRD'_P g(O - P). \quad (1.12)$$

Rating deviation is updated after the match with the formula:

$$RD'_P = \frac{1}{\sqrt{\frac{1}{RD_P^2} + q^2 g^2 P(1 - P)}}. \quad (1.13)$$

Thus, Glicko is an extension of the Elo rating system. Elo itself estimates the probability of correct answers in the same way as one parameter IRT model (the Rasch model) (Pelánek, 2016). The difference is the parameter estimation procedure. Analogically, MIRT multivariate Elo extensions have been tested (Doebler, Alavash, & Giessing, 2015). Elo rating-based systems, like IRT systems, depend on knowledge of item difficulties to estimate students' proficiency.

Recently, Elo rating-based systems have been used for ranking purposes, such as tennis match predictions (Vaughan Williams, Liu, Dixon, & Gerrard, 2021), defect detection in fabrics (Kang & Zhang, 2019), and rehabilitation for stroke patients (Garcia-Rudolph et al., 2021). Applications in education are also being explored: Elo rating is used to estimate the linguistic proficiency of students and the difficulty of the concepts they are learning (Hou et al., 2019), or, similarly, to match the difficulty of coding exercises to students' skill (Mangaroska, Vesin, & Giannakos, 2019). Studies indicate positive results and suggest further research in adapting Elo rating for material selection and student assessment.

When compared to IRT, Elo rating approaches are generally slightly less accurate but incomparably simpler to use (Antal, 2016; K. Wauters et al., 2011; Kelly Wauters, Desmet, & Van Den Noortgate, 2012). Thus, it is a great alternative for settings where sample size and the lack of calibration data do not allow for the IRT approach. In the study by Antal, the correlation coefficient for Elo

results and IRT ratings was 0.94 (Antal, 2016). Apart from IRT and Elo rating, there are other approaches to model learning (Desmarais & Baker, 2012), such as the performance factors analysis (Pavlik, Cen, & Koedinger, 2009) and Bayesian knowledge tracing (Corbett & Anderson, 1994). However, these models can be equally hard to implement and use due to non-trivial calibration and set-up procedures.

1.5.3. Multi-Armed Bandit Algorithms

The multi-Armed Bandit (MAB) family of algorithms is named after a problem for a gambler who has to decide which arm of a K -slot machine to pull to maximize their total reward in a series of trials (Vermorel & Mohri, 2005). The problem was first introduced by Robbins (Herbert, 1952). These algorithms are capable of negotiating exploration-exploitation trade-offs and are applied in several modern applications, such as advertisement placement, website optimization, network packet routing, and clinical trials (Bubeck & Cesa-Bianchi, 2012; Hardwick & Stout, 1992).

An opaque stochastic K -armed bandit can be viewed as a set of distributions $B = \{R_1, \dots, R_K\}$ (Vermorel & Mohri, 2005). Each distribution is associated with rewards from a particular lever: $\mu_1, \dots, \mu_K$. The gambler plays one level at a time and receives a reward from the chosen lever. The gambler's goal is to maximize the sum of the collected rewards. The number of rounds that are still to be played is called the horizon, H . There is a variety of variations of the problem, such as when the distributions for rewards are known. The exploration-exploitation dilemma does not exist in that family of problems. However, in an educational context, this is rarely the case, especially in terms of student assessment. Regret is the difference between the obtained rewards by using a particular strategy and the possible reward from an optimal strategy. The regret R_n after several rounds N is:

$$R_n = N \max_{k \in K} \mu_k - \sum_{n=1}^N X_n, \quad (1.14)$$

where, $\max_{k \in K} \mu_k$ is the mean of the largest reward a, and X_n is the reward for round n . A strategy whose average regret tends to zero as the horizon tends to infinity is called the zero-regret strategy. The problem of finding the best strategy for the gambler is called the multi-armed bandit problem.

Perhaps the simplest multi-armed bandit problem solution algorithm is ε -greedy. This strategy usually chooses a level with the highest estimated mean based on the rewards observed so far, and with frequency ε chooses a random lever. The choice of ε is left open and must be in the open interval $(0, 1)$. The first

component of strategy can be thought of as an exploitation move and the second as exploration. Such methods that have a clear distinction between exploitation (greedy choice) and exploration (uniform, random choice) are known as semi-uniform methods.

ϵ -first strategy is a variant of the ϵ -greedy strategy. This method consists of a purely explorative phase, where for an ϵN number of rounds, in the beginning, only random levers are pulled. After which, an exploitative phase follows, where for the rest of the $(1 - \epsilon) N$ rounds, the lever with the highest estimated mean is played. Analysis of this strategy by Even-Dar and colleagues shows that

$O\left(\frac{K}{\alpha^2} \log\left(\frac{K}{\delta}\right)\right)$ exploration pulls are required to find the α -optimal arm with the probability of at least $1 - \delta$ (Even-Dar, Mannor, & Mansour, 2002).

The simplest ϵ -greedy strategy is sub-optimal because the constant factor ϵ prevents the method from playing the optimal lever even as the number of rounds approaches infinity. A variant of ϵ -greedy that attempts to address the issue is called the ϵ -decreasing strategy. In this method, ϵ is a function of rounds played,

and its value is calculated by the equation $\epsilon_n = \min\left\{1, \frac{\epsilon_0}{n}\right\}$, where $\epsilon_0 > 0$, and its

choice is left to be fine-tuned for the problem at hand. The analysis of this strategy was first undertaken by Cesa-Bianchi and Fisher (Cesa-Bianchi & Fischer, 1998).

For this dissertation, the Upper Confidence Bounds (UCB) family of algorithms is particularly relevant as they operate on the idea of optimism in the face of uncertainty. This approach was first proposed by Lai and Robbins (Lai & Robbins, 1985), and the algorithm was invented by Auer and collaborators (Auer, Cesa-Bianchi, & Fischer, 2002). One of the variants, the UCB1 algorithm, is simple to implement and can achieve the $O(\log n)$ bound of regret. It calculates the number of times each arm has been played $t_k(n)$ in addition to their empirically observed means μ_k . The algorithm greedily picks an arm for which the sum of the mean and uncertainty is the greatest. For round n algorithm picks arm $j(n)$:

$$j(n) = \arg \max_{k=1 \dots K} \left(\mu_k + \sqrt{\frac{2 \ln n}{t_k}} \right). \quad (1.15)$$

There is a variety of other algorithms which are either more complex, have larger regret values like in the case of ϵ -greedy, are less explored, or have a combination of these factors. However, there are empirical studies that compare these algorithms in more detail on a variety of problems (Kuleshov, 2000; Vermorel & Mohri, 2005).

There are emerging applications of MAB algorithms in education for optimal learning material selection (Clement, 2014; Clement et al., 2014; Clement, Roy,

Oudeyer, & Lopes, 2013; Lan & Baraniuk, 2016; Mui, Lin, & Dewan, 2021). MAB family of algorithms are also used to recommend the best personalized actions to students (Clement et al., 2013; Lan & Baraniuk, 2016), sequencing content (Segal, Ben David, Williams, Gal, & Shalom, 2018), what teaching materials to present to learners (Y. Liu, Mandel, Brunskill, & Popovi, 2014), and what explanations to show (Williams et al., 2016). All these applications are conceptually similar and straightforwardly apply MAB algorithms as they would be applied in the selection of the best advertisement element on a website, for example.

MAB algorithms were also explored for experimental design (Rafferty, Ying, & Williams, 2018, 2019). Randomized experiments are a cornerstone for the improvement of educational processes. However, they may often result in many students' inferior learning outcomes. Researchers explored the possibilities of improving average benefits for a student in a research context with mixed results: benefits were increased but required at least twice as many students.

The application of MAB algorithms in formative assessment strategies was scarcely explored prior to the studies presented in this dissertation. In response to the 2019 study (Melesko & Novickij, 2019), Fuhua Lin suggested using Thompson sampling for quiz question topic selection; however, no simulation or empirical results were provided to back up this study direction (Lin, 2020).

1.6. Conclusions and Formulation of the Thesis Objectives

1. Personalized tutoring can have a strong impact on the learning efficiency of students; consequently, there is a strong academic push to provide smart learning using computers.
2. Learning style-based personalization is widely popular but is controversial and lacks empirical support. Therefore, it must not be the only basis for a learning system.
3. Students are bad at identifying the most efficient learning styles themselves and, consequently, should be assisted by tools.
4. The main tool for diagnosis and personalization of student competencies should be a formative assessment, as its feedback has a strong positive effect on learning.
5. The key to effective formative assessment is feedback through a three-step process: "Where am I going?", "How am I doing?" and "Where to next?"

The refined objectives of the thesis:

1. To design a multi-agent learning system based on existing smart learning technology to integrate with formative assessment methods.
2. To develop computer-adaptive formative assessment algorithms that do not require pretesting for use in the context of classical tests and latent trait theories to supply learning feedback.
3. To test the developed formative assessment algorithms in real-life conditions and compare the efficiency with established Multidimensional Item Response Theory (MIRT) based methodologies.

Design of Personalized Multi-Agent E-Learning System

This chapter focuses on the components and design of a smart learning system that could adapt to the specific profile of each student by incorporating different aspects of study material personalization. Also, this chapter presents two surveys of Vilnius University and Vilnius Gediminas Technical University students to support the viability of proposed educational techniques. This chapter concludes with a summary of the potential smart learning system design and survey results.

Four scientific publications were published on the section's topic (Kurilovas, Kurilova, Kurilova, & Melesko, 2016; Kurilovas, Meleşko, & Krikun, 2017; Melesko & Kurilovas, 2017, 2018).

2.1. Requirements for an Intelligent Personalized E-Learning System

The agent-based modular nature of the system compartmentalizes the requirements. Each agent has its own requirements, and the system's requirements are the sum of these needs. However, an instructor can choose not to use every personalization option. For example, as indicated by a literature review, the efficacy of the personalization method based on popular learning styles can be questioned

and may be excluded by a pedagogue, removing the necessity for some of the requirements.

Requirements for an intelligent personalized e-learning system:

1. Learning style personalization requires that each subject or competency is taught in multiple learning objects suitable for students with various learning styles.
2. Learning analytics model used in a profile control agent requires ANN training data, particulars of which are discussed in the corresponding section.
3. Pedagogical Suitability Estimation Agent requires that the learning objects are supplied with metadata according to the LOM standard.
4. Formative assessment agent requires a pool of items assigned to topics on competences they are designed to evaluate.
5. The topics evaluated using the item pool for the formative assessment agent must be the same as those assigned in LOM.

From the pedagogical point of view, learning software and all learning processes should be personalized according to the main characteristics and needs of students. While not all student features are viewed as equally useful for pedagogical content personalization, their efficacy may vary depending on the quality of data available about the students. One of the most important components of a personalized e-learning system is formative assessment, which enhances the learning path estimation based on the knowledge level of the student. Thus, the model of the system focused on content personalization, and formative assessment is presented in this chapter with several agents that supplement it.

Students have different needs and characteristics, i.e., prior knowledge, intellectual level, interests, goals, cognitive traits (working memory capacity, inductive reasoning ability, and associative learning skills), learning behavioural type (according to their self-regulation level), and, finally, learning styles. Learning personalization may be improved by implementing personalized learning packages (scenarios, units) based on a recommender system suitable for particular learners according to their personal needs.

Learning styles-based profiles can be used to support e-learning system personalization and adaptivity. Literature review revealed that learning styles, while not as unequivocally proven to be effective, are overwhelmingly supported by educators. Thus, the proposed system should include learning styles to make it more attractive to instructors in addition to providing learner personalization. It can be based on, e.g., Felder and Silverman Learning Styles Model (Felder Richard M. & Silverman Linda K., 1988). Dedicated psychological questionnaires should be applied, such as the Soloman and Felder Index of Learning Styles (Soloman et al., 2012). Then, one should integrate other features into the learner's

profile: knowledge, interests, goals, cognitive traits, learning behavioural type, etc.

Educational data mining methods and techniques should be used to analyze the learners' behaviour in e-learning systems. The data on this learning behaviour could differ from the data obtained by using dedicated psychological questionnaires. Literature review showed that students are remarkably inept at analyzing and reporting their learning habits. Since this behaviour data shows the learners' motivation while using different learning components, this data should be kept in mind when creating revised versions of integrated learner profiles. Then, a personalized recommender system should be created to suggest learning components (LOs, learning activities and methods, platforms, environments, tools, apps, etc.) suitable to learners according to their profiles.

A recommender system should form the preference lists of the learning components according to the expert evaluation results. Probabilistic suitability indexes can be identified for all learning components in terms of their suitability level to learners. Probabilistic suitability indexes could be easily calculated for all learning components and all students if one should multiply learning components' suitability ratings (obtained while the expert is evaluating the suitability of the learning components to certain learning styles by probabilities of particular students' learning styles). These suitability indexes should be included in the recommender system, and all learning components should be linked to particular students according to those suitability indexes. The higher the suitability indexes, the better the learning components fit the needs of particular learners.

Thus, the recommender system should prioritize highly supported personalization methods, such as formative assessment, with supplementation of other techniques with context-dependent efficacy. Several intelligent technologies should be applied to implement this approach, e.g., ontologies, recommender systems, intelligent software agents, decision support systems to evaluate the quality and suitability of the learning components, personal learning environments, etc. The most suitable learning components for a student based on these parameters should be prioritized with inclusion in personalized learning packages (scenarios, units) that could be created based on learning goals and the current knowledge level of a student.

The learning process is evolving with the development of new technologies. The didactic teaching style with a paper-and-pencil summative assessment format was once a dominant form of instruction. Current studies indicate different preferences and more effective approaches, as discussed in the review section. A survey of current medium preference was held to gauge students' attitudes towards computer-assisted testing, bring your device (BYOD) mobile assessment, and traditional paper-and-pencil assessment. In total, 104 undergraduates, sophomore, and junior (third year) students in Vilnius Gediminas Technical University from

seven different subject groups were tested. The survey was executed on 25 April 2019. The survey included one question with three choice options: paper, phone, tablet, or laptop. The results of the media preference survey can be seen in Fig. 2.1.

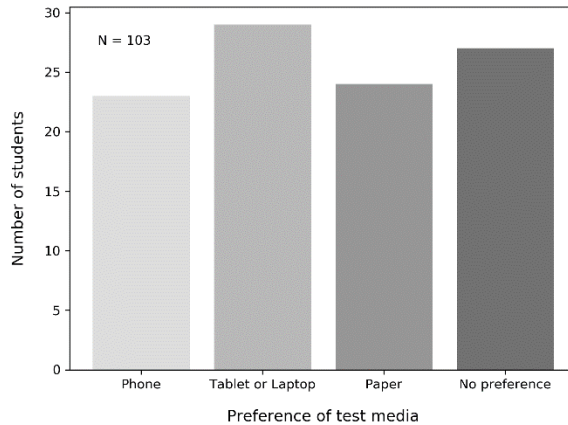


Fig. 2.1. Quiz media preference among students ($N = 104$). No preference bar indicates students who chose not to express their preference for a media type

Students appear to be largely ambivalent towards quiz media types. However, only 22% (23 out of 104) of students preferred the traditional paper-based quiz format. 26% (27 out of 104) students had no preferences, which left 52% (54 out of 104) students preferring computerized quizzes and electronic media. Thus, given the numerous advantages of computer-assisted testing, such as reduction in quiz length, monetary cost, and workload for a teacher, computer-assisted formative assessment should be a valuable pedagogical tool going forward.

2.2. Conceptual model of Personalized Intelligent E-Learning System

The chief novelty of the proposed system would mainly lie in its use of formative assessment to customize learning content. Such systems do not presently exist because of the lack of suitable algorithms.

Based on the approach outlined in the previous section, a personalized intelligent learning system could be created using intelligent software agents. Intelligent pedagogical agents could help to personalize learning. The abilities of software agents vary significantly, depending on their roles, capabilities, and environments. To describe these abilities, different notions of agents have been

introduced. According to Green et al., intelligent agents have three major concepts defining their behaviour (Green, Hurst, Nangle, & Cunningham, 1997). First, they can act on behalf of others. Second, they have a degree of autonomy. Third, they display reactivity and proactivity. There are also various taxonomies created for agents, e.g., collaborative, interface, mobile, information, reactive, hybrid, and smart agents. In this context, intelligent agents have been associated with various functions, e.g., personal assistants, information managers, information seekers, planning agents, coordination agents or collaborative schedules, user representatives, etc. Their application in the educational field is mostly as personal assistants, user guides, alternative help systems, human-system mediators, etc.

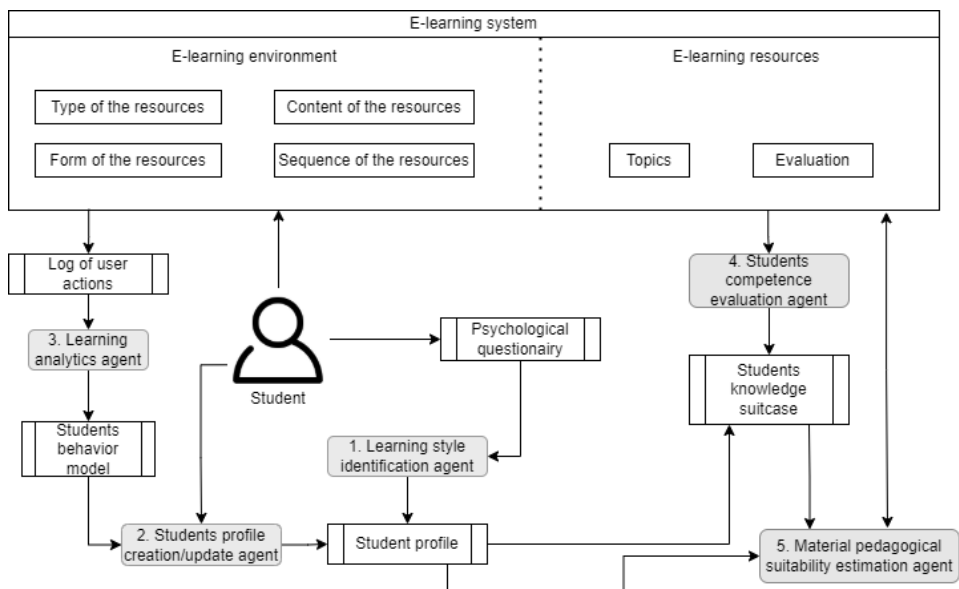


Fig 2.2. Conceptual model of a personalized intelligent learning system

Because pedagogical agents are autonomous, they inherit many of the same concerns that autonomous agents must address in general. In general, it has been argued that practical autonomous agents must manage complexity (Elmarzaki, 2018). They must exhibit robust behaviour in rich, unpredictable environments; they must coordinate their behaviour with that of other agents and must coherently manage their behaviour, arbitrating between alternative actions and responding to a multitude of environmental stimuli. In the case of pedagogical agents, their environment includes both the students and the learning context in which the agents are situated. Student behaviour is by nature unpredictable since students may exhibit a variety of aptitudes, levels of proficiency, and learning styles.

To create a conceptual model of a Personalized Intelligent Learning System (PILS), a multi-agent system pattern can be used. The conceptual model of a personalized intelligent learning system is presented in Fig. 2.2. The system should consider knowledge about students and pedagogical material to form the recommendations. Therefore type, form, sequence, and the content of resources should be adapted to students' personal needs. To enact personalization across different pedagogical elements, it's best to use several agents, where each has a clearly defined function. Such a modular approach can be valuable in the practical implementation and application of the proposed system, as well as initial planning and iterative improvement stages.

The formative assessment agent (4) is principal in following good feedback practices. It is the first step towards personalization, as it can ensure that each student can directly see their weak points and a path towards greater mastery of the subject. Of course, time and effort spent making and taking a formative assessment test can be otherwise spent directly studying. A formative assessment agent can mitigate prohibitive costs to a teacher by assessing students on their behalf. And the algorithms proposed in Chapter 3 of the dissertation allow mitigating the student costs, mainly by reducing test length.

Several additional agents can improve the degree of personalization of the system and supplement formative assessment. A learning style identification agent with a dedicated psychological questionnaire should be applied to obtain the probabilistic learning styles of students (Soloman et al., 2012). Learning styles identification software agent (1) should be created to obtain these values for student learning styles. Student's preference along the Active-Reflective axis indicates what type of resources should be prioritized: tasks that require proactive participation or passive comprehension. Depending on whether a student belongs to a Sensing or Intuitive learner group, the resources that are served to them should vary in abstraction level. Sensing students prefer concrete, practical procedures, while intuitive students prefer theories and generalizations. Visual and Verbal learners should be served resources that reflect their preferred modality. Finally, Sequential and Global learners differ in their thinking: global learners are holistic thinkers, while sequential learners should be provided content in a serial, orderly fashion.

Psychological questionnaire results are not always accurate and can be difficult to dynamically update (Graf, Liu, & Kinshuk, 2010). Possible causes of this are long questionnaire lengths and imprecise answers. The data on real learners' behaviour in the learning environment obtained using educational analytics should be used to correct students' profiles in conformity with the obtained data. Thus, a learning analytics software agent (2) should be created to diagnose learners' profiles by observing their real behaviour in the learning environment.

The data provided by learning style identification and learning analytics agents have to be reconciled with the help of a learner profile control agent (3). This agent will create and update learner profiles based on the data provided by the students as well as the data gathered from their activity within the learning environment. This data can also be supplemented with additional parameters and features, such as knowledge, interests, goals, cognitive traits, learning behavioural type, etc. For the educational system to function, it does not require the knowledge of every feature of the student; however, in some contexts, educators may wish to lean on one or another pedagogical strategy to supplement an effective feedback model. And the function of secondary agents is to enable them to do so.

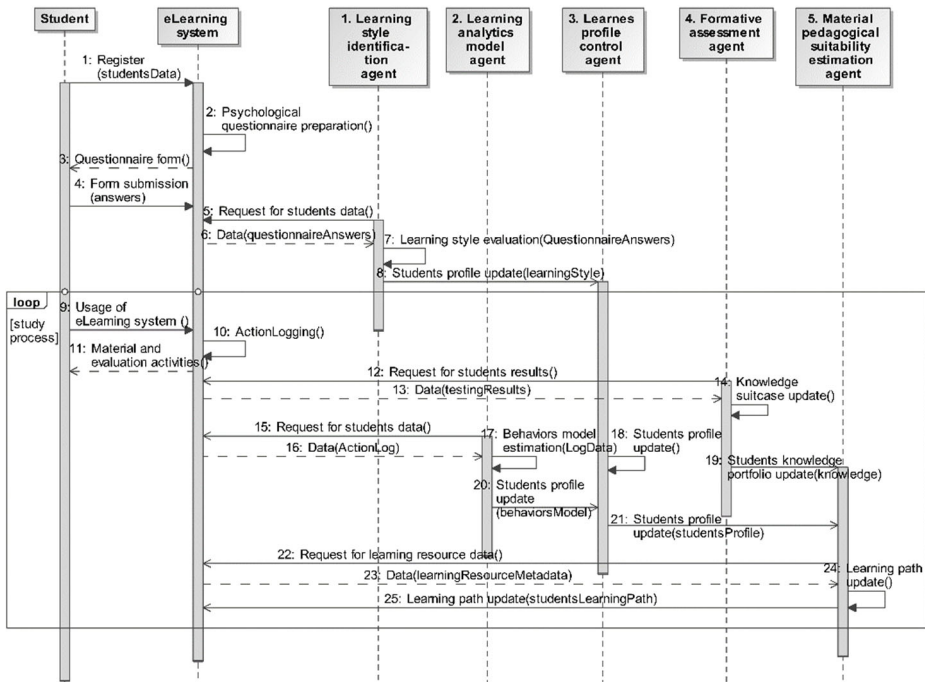


Fig. 2.3. UML sequence diagram of a personalized intelligent learning system behaviour and its agents

Next, a pedagogical suitability software agent (5) should be created to implement a recommender system. This pedagogical suitability software agent should link optimal learning components to particular students according to probabilistic suitability indexes. All learning components in the recommender system should be linked to particular students according to their probabilistic suitability indexes. The higher the probabilistic suitability index, the better the learning component

fits a particular student's needs. Considering the students' knowledge suite (evaluated by formative assessment agent (4)) and material suitability, the optimal learning scenario (i.e., the learning scenario of the highest quality) for a particular student can be generated. The levels of students' competencies, understanding, skills, etc., directly depend on the level of application of optimal learning scenarios in real pedagogical practice. Quickly identifying lacking knowledge in students (with a formative assessment agent (4)) and serving them the best-suited material that would improve their understanding of the learning environment will increase learning efficiency and should be implemented by the learning scenario agent. The UML Sequence Diagram of Personalized Intelligent Learning System Behaviour and its Agents is presented in Fig. 2.3.

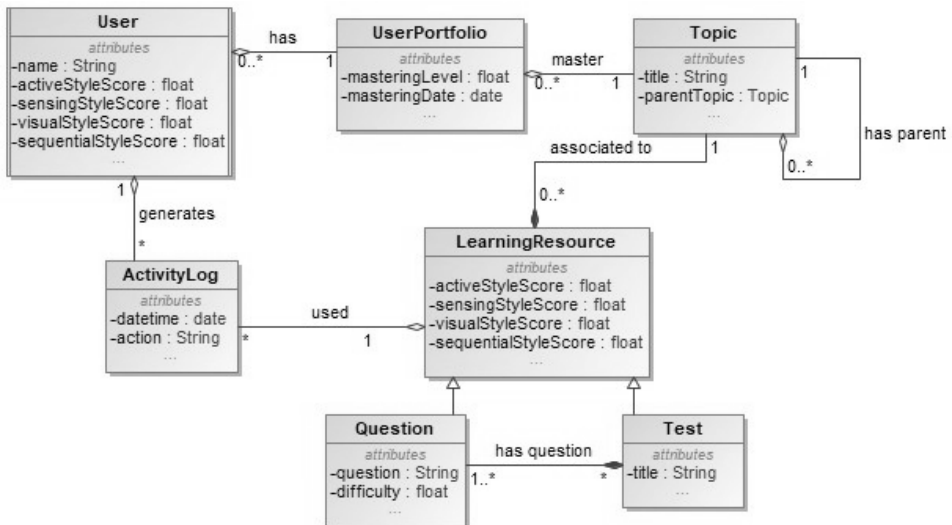


Fig. 2.4. UML class diagram of multi-agent e-learning system general data model

The general data model is presented in Fig. 2.4. The new student will be presented as a “User” object, but no (or default ones) learning style attribute values will be assigned. The learning style identification agent will estimate the potential scores for each learning style. This will be done by using a questionnaire, as in the beginning, there will be no “ActivityLog” records for student behaviour in the e-learning environment. The “ActivityLog” will be collected by recording student activity by using different “LearningResource” instances (viewing learning material, solving tasks or tests, etc.). The log data will be used to update the user’s learning style by a learning analytics agent. The user’s profile update regarding the learning style (related to the form of learning resources) will be done in the “User” object. While the “UserPortfolio” object (related to the content of learning

recourses) presents the set of “Topic” objects (specifying the topic or competence the course develops), the student has already mastered or needs to improve. The mastering level of each topic should be estimated with the help of the student’s competence evaluation agent. Each learning resource (formative assessment test and questions as one case of the learning recourse) should have an associated topic it teaches or evaluates. This will allow mapping between the test competencies and the student’s portfolio or mastered topics/competencies.

All learning resources should be specified based on their suitability for different learning styles. This can be done manually by the course owner, or it can be at least partly automated, considering the type of the learning resource, the format of the content, etc. Whether it is assigned manually or not, the learning material suitability estimation agent will evaluate all the “LearningResource” objects for a specific student based both on the learning style matching as well as topic mastering level (by eliminating already fully mastered topic resources and prioritizing the topics based on the mastering level).

Possible deployment principles of the high-level multi-agent e-learning system are presented in Fig. 2.5. The multiple agents would be integrated with the e-learning system, which is running on the webserver and has a user interface in the clients’ web browser. Only learning style identification and student competence evaluation agents have a user interface dedicated to the students. Those two agents intercept e-learning system calls or are integrated into the system. The rest of the agents have no direct user interfaces and run as services in the background. The integration of these agents is implemented via the system database, i.e., the agents get data from the e-learning system database, process it, and store the result in an appropriate database location. Such a solution allows easier integration of the agents into the e-learning system.

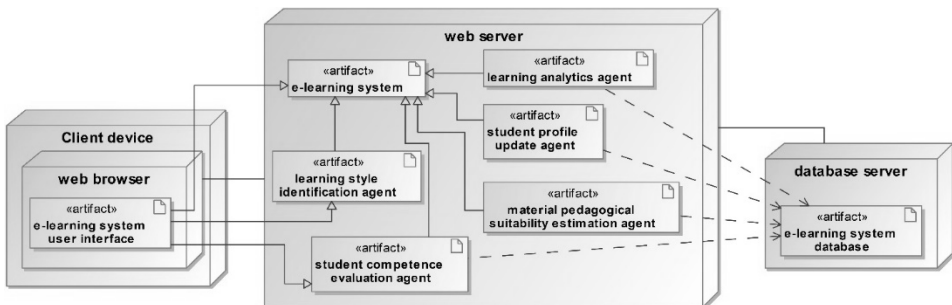


Fig. 2.5. UML deployment diagram of the high-level multi-agent e-learning system

The student competence evaluation agent, dedicated to formative assessment, is the most interconnected, as it requires data from the database on course topics

and questions for each of the topics and uses the data interactively based on user input (answers). The result of the formative assessment is stored back in the database for further use to view the user competence portfolio, generation of personalized learning path, etc.

2.3. Intelligent Agents

2.3.1. Learning Style Identification Agent

Well-established models exist for the estimation of a learning style, e.g., Felder and Silverman Learning Styles (FSLSM) (Felder Richard M. & Silverman Linda K., 1988) that are usually identified with the help of the Soloman–Felder questionnaire (Soloman et al., 2012). This model is designed to accommodate Science, Technology, Engineering, and Mathematics (STEM) subjects as well as other fields of study.

An experiment was carried out to investigate the application of the learning style estimation model. The target group of the Felder and Silverman Learning Styles Model case study was 22 final year students (two full groups) of the Faculty of Medicine of Vilnius University. Division by gender was 17 (77.27%) females and 5 (22.73%) males. The age of the students was from 23 to 25.

Students' learning styles according to four different FSLSM dimensions were: (1) by information processing: Active versus Reflective; (2) by information type: Sensory versus Intuitive; (3) by sensory channel: Visual versus Verbal; and (4) by understanding: Sequential versus Global. The analysis for each dimension is presented in the next four tables. The results indicate the split between two possible styles and students' distribution based on achieved values.

The Soloman–Felder questionnaire was presented to the students. Values of students' preferences for learning styles were established from 1 (the mildest) to 11 (the strongest) depending on the number of corresponding questions answered by students. Thus, students' preference for different learning styles was evaluated according to the survey results. Research results concerning the identification of students' learning styles according to FSLSM at Vilnius University are presented in the tables below: individual data is provided in Tables 2.1 and 2.2, and the overall results are given in Table 2.3.

Analysis of the survey results showed that:

1. 50% of students had a medium preference for learning styles, whereas even 45% had a strong preference for a certain learning style. Most tested students preferred Active (59.1%), Sensory (72.7%), Visual (86.4%) and Sequential (54.5%) learning styles. Most students had medium to a strong preference for learning styles:

- Visual (13 – 59.1%), Active (6 – 27.3%), Sensory (6 – 27.3%), and Sequential (4 – 12.2%), as seen in Table 2.3.
2. Multimodal (students having preferences for more than one mode) included 14 (63.6%) students. Nine (40.1%) preferred two learning styles, and five (22.7%) preferred three learning styles at the same time. Two (9.1%) students had a strong preference for multimodal learning styles.
 3. Only one student had no significant preferences for any of the learning styles (Table 2.2).
 4. Learning styles located on the left side of the scales were dominant: 59.1% of students were evaluated as Active, 86.4% – Visual, 72.7% – Sensory, and 54.5% – Sequential (Table 2.3).
 5. There was no significant difference in preference for learning styles between genders.

Table 2.1. Individual results of identifying student learning styles in group one

Learning style		Student no.										
		1	2	3	4	5	6	7	8	9	10	11
Active vs. Reflective	Active	0	0	1	9	0	5	5	1	3	0	0
	Reflective	7	5	0	0	1	0	0	0	0	5	1
Sensing vs. Intuitive	Sensing	1	0	1	1	0	3	3	3	3	7	1
	Intuitive	0	5	0	0	1	0	0	0	0	0	0
Visual vs. Verbal	Visual	5	5	1	0	5	7	9	7	7	0	9
	Verbal	0	0	0	9	0	0	0	0	0	5	0
Sequential vs. Global	Sequential	5	0	9	1	0	0	0	3	0	0	0
	Global	0	1	0	0	9	5	1	0	3	3	9

Table 2.2. Individual results of identifying student learning styles in group two

Learning style		Student no.										
		1	2	3	4	5	6	7	8	9	10	11
Active vs. Reflective	Active	1	0	5	0	1	0	0	3	1	9	7
	Reflective	0	3	0	1	0	3	9	0	0	0	0

End of Table 2.2

Learning style		Student no.										
		1	2	3	4	5	6	7	8	9	10	11
Sensing vs. Intuitive	Sensing	0	5	0	3	5	9	0	1	0	7	9
	Intuitive	3	0	3	0	0	0	3	0	5	0	0
Visual vs. Verbal	Visual	11	7	1	7	1	1	0	1	5	1	7
	Verbal	0	0	0	0	0	0	3	0	0	0	0
Sequential vs. Global	Sequential	0	0	1	1	7	5	0	3	3	3	1
	Global	3	3	0	0	0	0	1	0	0	0	0

Table 2.3. Overall results of identifying student learning styles

Learning style		Number of students		Students with medium to a strong preference	
		Absolute	Percentage	Absolute	Percentage
Active vs. Reflective	Active	13	59.1	6	27.3
	Reflective	9	40.9	4	12.2
Sensing vs. Intuitive	Sensing	16	72.7	6	27.3
	Intuitive	6	27.3	2	6.1
Visual vs. Verbal	Visual	19	86.4	13	59.1
	Verbal	3	13.6	2	6.1
Sequential vs. Global	Sequential	12	54.5	4	12.2
	Global	10	45.4	3	13.6

Additionally, an analysis was executed to determine how many learning styles can be assigned to one student. The summary in percentages is presented in Fig. 2.6.

Students having a strong preference for a learning style have more difficulties in traditional curriculum-based or program-based learning than learners with a mild preference for a learning style. Thus, learners with a strong preference for a learning style can especially benefit from applying intelligent technologies and adapting their learning packages (scenarios) to their preferred learning style. Research results on identifying students' learning styles show that 50% of students who participated in the case study survey had a medium preference for learning

styles, whereas even 45% had a strong preference for a certain learning style, according to the Felder and Silverman Learning Styles Model. These students could benefit from applying smart (intelligent) technologies, e.g., ontologies, recommender systems, and intelligent software agents, to create personalized learning packages matching their learning styles.

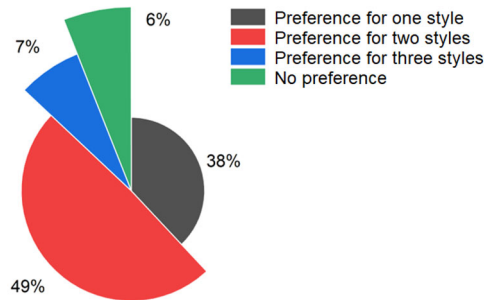


Fig. 2.6. Number of learning style preferences among students

The results of the research indicate that the existing learning styles model is not complex from the point of view of implementation, and it is possible to place learners on the learning style axes of the model based solely on their answers.

The agent's interaction with the system can be summarized in terms of percepts and actions. Its percepts are the answers to the Solomon–Felder questionnaire. Its action is to output the learning styles of students who took the questionnaire. This places the agent into the simple reflex agent class according to Russel and Norvig's classification.

2.3.2. Learning Analytics Model

The learning analytics agent should be implemented using an artificial neural network. As the literature review indicates, ANN can be successfully employed to tie students' psychological profiles to various behavioural and performance characteristics. In the case of the Felder–Silverman psychological model, the characteristics describe how a student approaches learning and performs during the assessment. If the learning analytics ANN agent is taught to correlate these factors to the Felder–Silverman learning styles of a student, it will be able to identify, re-evaluate and adapt to students who exhibit anomalous behaviour for their model. This could happen if a student made a mistake while answering a Solomon–Felder Index of Learning Styles Questionnaire or was dishonest in psychological self-diagnosis. In the long-term, this may lead to failure to meet academic goals by the

student, as the system would serve them with maladapted learning content. The psychological profile of a student may also change because of academic growth and exposure to other social factors. Thus, the ANN learning analytics agent is necessary for continuous adaptation to the students' needs.

The learning analytics agent proposed in this section should be trained based on multiple inputs and four outputs. Traditionally, the inputs should be chosen to reflect the Felder–Silverman learning styles. The six predictors are: a semantic cluster of independently visited learning content, the order in which learning objects are first studied, the total number of material revisits, amount of communication within the system, time spent on problems during the assessment, and performance during the assessment. The outputs are four numbers between 0 and 1, placing the student on four psychological axes: Sensing–Intuitive, Visual–Verbal, Active–Reflective, and Sequential–Global.

The first factor chosen as a predictor is the student's belonging to the semantic cluster. Students should be given not only personalized learning courses but also the freedom to explore and independently study the material they deem valuable. By extracting semantic data from the searches student performed in the learning system and learning objects they visited, the system can assign a student a cluster. This cluster category should be fed as one of the inputs of the ANN that shall learn to associate the student type with his psychological profile. This predictor is expected to directly categorize students by the type of information that they seek out on their own and thus give insight into their real learning patterns, as opposed to the self-reported learning style (Mecca, Raunich, & Pappalardo, 2007; Zhang & Dong, 2004). Mecca and collaborators introduced a clustering algorithm based on singular value decomposition (SVD) derived dynamic SVD clustering (DSC), which uses latent semantic indexing. It is based on an incremental computation of singular values and does not require computing the whole SVD of the original matrix (Mecca et al., 2007). The algorithm does not assume a fixed number k used to represent concepts in the document space. This separates it from snippet-based engines. Snippets are short document abstracts returned by syntactic Web search engines. These abstracts usually contain words around search query term occurrences. While snippets are usually 40 words and shorter, allowing for fast processing, they do not offer the best representation of the whole document content. The discussed algorithm considers the whole document content; this will allow a finer distinction between the learning objects visited by students. Based on these distinctions, clustering can be accomplished and then later used in an artificial neural network agent to diagnose incorrect Felder–Silverman profiles of students.

The next predictor should be the order in which the student studies presented learning objects. This should be directly reflective of their placement on the Sequential–Global axis. If the student finishes all presented learning material in the

order it was presented, they are probably a Sequential learner type, and the ANN training should reflect that. Otherwise, if the sequence is out of order, it may be indicative of Global learner tendencies.

The order in which material is studied may also reveal indirect clues to other personality traits, such as Intuitiveness, in other words, a need to know the reasons and deeper causes behind what is taught, and thus curiosity about what is taught next.

The Visual–Verbal axis may also have an impact on the order of study if some learning objects are richer in visual aids than others. Active–Reflective traits may manifest themselves as student leaves exercises for later or start applying knowledge right away. For these reasons, the order in which a student completes learning objects from a learning scenario may serve as a powerful predictor of their psychological type.

For similar reasons, the number of materials and revisits may serve as a predictor of students' position on Felder–Silverman axes. A direct connection is expected between the number of revisits and the Sequential–Global type of a student, as global learners are expected to study the presented material at the same time in small portions. Active students may also have a greater number of revisits as they seek greater interactivity during their study time. The other reason for including the number of revisits as a predictor was the means to distinguish between students who meticulously study the material once and those who visit it first and study it more thoroughly later.

The amount of communication is expected to be directly indicative of the Active learning type, as it is typical for people who score highly on the Active scale to prefer to try things out and discuss new concepts while working with others in groups. Increased communication may also indicate verbal learners. When content is rich in visual information, students may seek a more verbal didactic environment by participating in a conversation with their peers and instructors. Analogical effects may manifest themselves with content that is unfit for the student according to other axes, as students seek aid from others.

The final two predictors chosen for the ANN learning analytics agent were time spent on problems and performance during assessment. These factors should be indicative of general difficulties the student may be having due to mismatched content or because of the way the assessment problems are presented. It may provide insights other factors do not offer because of the universal nature of the assessment content. It is the content that is sure to be common to every student regardless of their psychological type, while learning objects will differ from person to person.

Such a network with six input groups and four outputs can be trained using backpropagation gradient descent to identify the learning style of students based

on their behaviour in the learning environment. To summarize, the six predictors are:

1. A semantic cluster of independently visited learning content
2. Order in which learning objects are first studied
3. Total number of material revisits
4. Amount of communication within the system
5. Time spent on problems during the assessment
6. Performance during the assessment.

After the ANN learning analytics agent is trained to classify the psychological types of students based on the six predictors, it will monitor the activity of new students and predict their learning styles. These predictions are then used by the learner profile control agent to troubleshoot possible digressions from learning styles as identified by the learning style identification agent.

However, later research papers on automated learning style estimation are based on data, which can be accessed in log files of existing massive open online courses (MOOC). Therefore there are 26 student behaviour data retrieved from nine learning object types (Gomede, de Barros, & de Souza Mendes, 2020). Those 26 features are used as input to the ANN. Such an increase in features and structure of newly used ANN allows to achieve up to 97% of classification accuracy (Alshmrany, 2022). Sami Alshmrany was able to achieve 98% accuracy for the Sequential/Global learning style, 95% accuracy for the Sensing/Intuitive learning style, 89% accuracy for the Active/Reflective learning style, and 79% accuracy for the Visual/Verbal learning style. These results indicate that the most difficult to evaluate is the Visual–Verbal learning style. However, it is mostly related to the fact that in the context of MOOC, it was not clear which resources were presenting material visually and which ones were dependent on the verbal instruction form. Nonetheless, the achieved classification accuracy was relatively high and didn't have a big influence on personalization: in most cases, the learning resources contained a mixture of visual and verbal information. The existence of a learning style classification model based on users' behaviour shows that such a solution is possible, and working implementations exist that rely on current ANN models.

Thus, the agent that utilizes this model would fall into model-based reflex agent classification and must accept percepts presented in Fig. 2.4 as input, and students' learning style would be its output.

2.3.3. Learner Profile Control Agent

Learner profile control agent takes inputs from learning style identification and learning analytics agents and uses them to create and update the user profile. The

initial learners' profiles should be created based on the Soloman–Felder questionnaire by a learning style identification agent; then, periodically learner profile control agent should make use of the learning analytics agent to diagnose students' behaviour in the system. If the prediction of the learning style by the analytics agent significantly differs from the results of the Soloman–Felder questionnaire, the student will be flagged for re-examination. Alternative learning scenarios may be presented to such a student, one based on data from the questionnaire and the other based on the predictions made by the agent. The path is chosen by the student, and the results he/she demonstrates during the assessment can be used to decide between two competing psychological models.

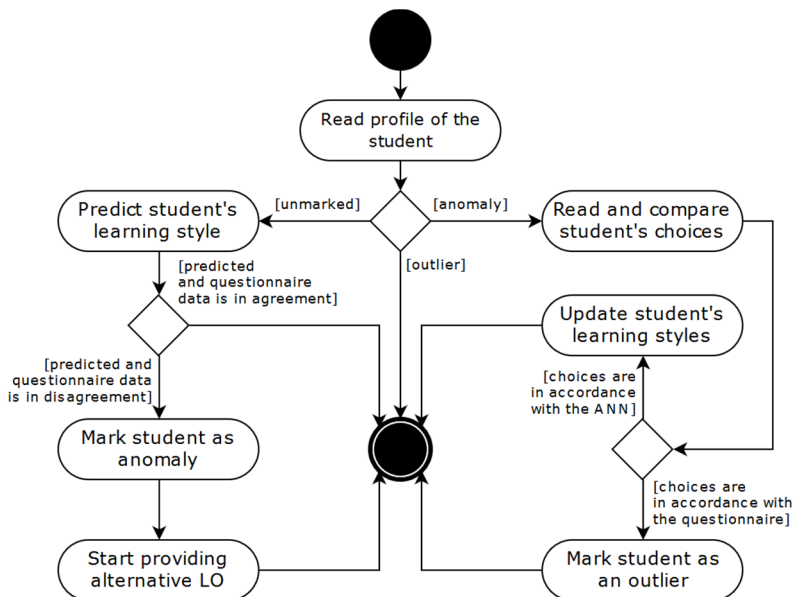


Fig. 2.7. UML 2 activity diagram of learner profile update procedure with the help of learning analytics agent

The learning analytics agent's interaction with the learner profile control agent is shown in Fig. 2.7. Anomalous students are students whose behaviour suggests a different learning style than the one assigned to them by the questionnaire. Outlier students are students who behave as if their learning style was misdiagnosed yet prefer to learn in a way than was originally assigned to them by the questionnaire. Unmarked students are all students who were not marked as an anomaly or an outlier.

By comparing the predicted student psychological profile with the one determined by the questionnaire, the agent will be able to identify the students who

have changed, irresponsibly filled out the questionnaire, or are outliers by their nature. Once these students are identified, the system can adapt to them by giving them an option to choose two alternative learning paths and by asking them to revisit the Soloman–Felder questionnaire.

The interaction of the agent with the system can be summarized in its percepts, i.e., the student's profile data. Namely, the student's currently assigned learning style, their behaviours within the system, and status. The agent's actions are to mark a student's status as either an anomaly or an outlier.

2.3.4. Formative Assessment Agent

The principal focus of the proposed learning system is to satisfy good feedback practices, which are widely supported as an effective educational technique. The formative assessment agent's purpose is to answer "Where am I now?" and "Where to next?" questions. These questions must be answered before any additional customization of the learning process can be of practical value: there is little point in providing the material that suits the students' learning style if it is what they already know or do not wish to learn. Therefore, each learning resource should be attached to a topic and its relevance to a student's estimate. The formative assessment agent should work in tandem with the pedagogical suitability estimation agent that uses Learning Object Metadata (LOM) to join formative assessment results with that of secondary agents. The mentioned LOM standard provides the possibility by presenting data in the classification category.

The formative assessment agent proposed in the dissertation aims to provide the following advantages to the personalized intelligent learning system:

1. Increased motivation through demonstration of acquired competencies
2. Assessment enhanced learning through identification of lacking competencies
3. No judgment and comparison of students by means of grades
4. Increase in quality of education through the requirement that all students acquire all necessary competencies
5. Personalized and individuality-sensitive assessment.

The listed advantages of a formative assessment agent are achieved by assigning each learning object and assessment activity with a competency which it is supposed to train or test in a student and requirement proficiencies needed to efficiently study it. Assessment milestones are lists of required competencies; the milestone is achieved when a student demonstrates all required competencies. Competence can be awarded to a student in two ways: by a teacher after reviewing the work in the student's portfolio or by the formative assessment agent upon successful completion of a learning object or assessment activity. This means that

successful completion of the course can be viewed as the acquisition of all necessary competencies and not just a minimum average mark. Of course, when the PILS is required to fit into a larger teaching context, the acquired competencies together with the portfolio can be translated into marks by the teacher.

The formative assessment agent works in tandem with the learning scenario composition agent to account for the sometimes-complex relationship between competencies. Some competencies can only be acquired after several other requirements are met.

When the student meets the requirements for one of the learning objects from the list of goal proficiencies, the formative assessment agent informs the learning scenario composition agent to include new learning objects in the student's learning scenario. In addition, the assessment agent records this milestone and its acquisition date in the student's portfolio, which is then presented to the student as a motivational aid. This way, the student is constantly receiving positive feedback from the system regarding their progress.

In addition to receiving positive feedback from a formative assessment agent, the student may access their portfolio and compare it to their learning goals. The learning goal is a list of proficiencies to be acquired during studies, which can be contrasted to the proficiencies already acquired and stored in the student's portfolio together with the demonstrating work. This way, the student has access to a clear roadmap of their past and future work and is empowered to engage in self-directed learning. As the list of acquired proficiencies grows, the student may derive motivation from seeing them gradually becoming competent in the subject. Also, by making the direct connection between the work done and the proficiencies acquired, the portfolio documents not only contain what was learned but also how it was learned. This is valuable to students and teachers because it can be used not only to assess performance more accurately but also to support future learning.

Testing students requires fine-grained connections between tasks and questions and the proficiencies that they are supposed to assess in the subjects. Specifying what each task seeks to establish in the recipient of the test enables the formative assessment agent to provide constructive feedback to students in the form of proficiencies that need further improvement. When a student makes a mistake, it is immediately tied to lacking proficiency. This way, the student receives personalized learning guidance instead of judgment in the form of a grade. Tests are included in students learning scenarios by the learning scenario composition agent when the formative assessment agent detects that a student completed all learning objects that were provided to them to teach the specific competencies the test is supposed to assess in students.

The agent's percepts during a test are the students' answers to test items; its action is to supply the next question. After the test, its action is to suggest a list of

topics for the student to study. Because the agent would have to gather and retain data about students' previous answers to serve the next question, it can be classified as a model-based reflex agent.

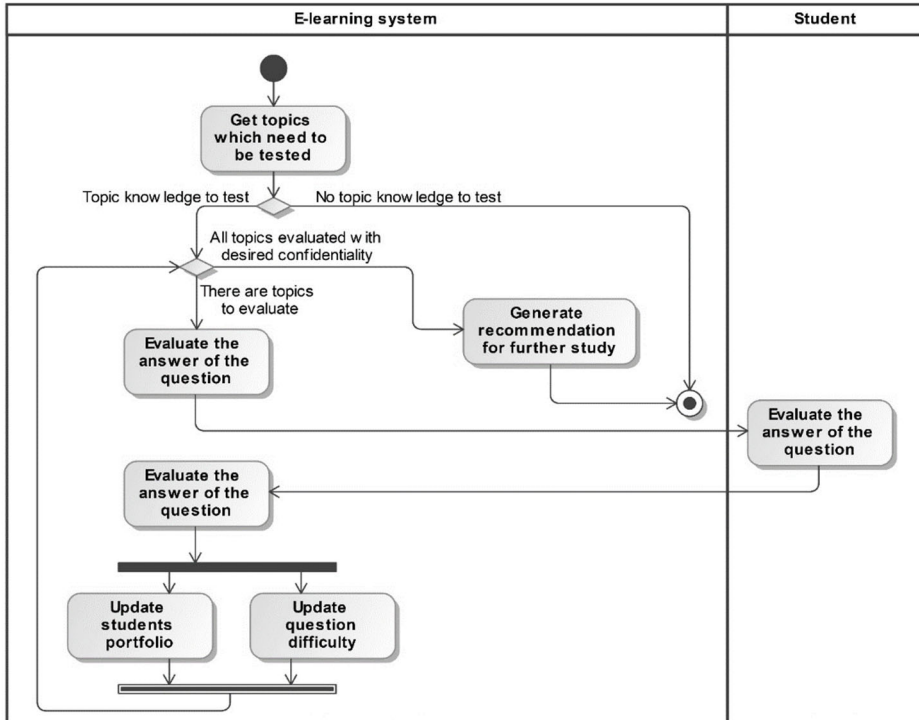


Fig. 2.8. UML activity diagram of the Elo-Bandit method application in an e-learning system

In an e-learning system, the Elo-Bandit algorithm should be used by the students' competency evaluation agent. The application method can vary, but the main principle is presented in Fig. 2.8. The agent should be executed on some event or time interval and start by estimating the need for formative assessment. Based on a course, the need for topic knowledge estimation can be based on the date when the topic mastery level was estimated the last time, i.e., the achieved topic mastery level, the schedule of the course, etc. If the topic for formative evaluation exists, the Elo-Bandit method will be used multiple times, while the confidentiality of the formative assessment for each topic will be achieved. For each of the questions selected to be used in the formative evaluation process, after the students provide answers and the correctness of the answers is estimated, the student rating and the question difficulty are updated.

When the formative assessment is done (the mastering level of all needed topics is estimated), further learning recommendations are generated. This could include a summary of formative assessment results, the highlight of topics that need improvement, an update of the student's learning path, etc.

While formative assessment can be beneficial to the learning process, it comes at a cost to teachers and students. These challenges did not have a clear solution at the time of compilation of this dissertation, and they deserve separate attention. They are addressed in detail in the third chapter of the dissertation, where novel algorithms are presented and tested. The formative assessment agent should make use of these algorithms to provide feedback to students saving their time and energy for correctional instruction. This is its principal novelty and advantage over currently existing personalization methods: the novel algorithms provided in the third chapter of the dissertation.

2.3.5. Pedagogical Suitability Estimation Agent

To estimate pedagogical learning resource suitability for a specific user is a comparison of learning resource metadata with the user profile. To do that, the learning resource must have metadata describing the properties of a specific learning type (Conlan, Hockemeyer, Lefrere, Wade, & Albert, 2001) or has a clear mapping between different types of learning resources to specific students' learning profiles (Milosevic, Brkovic, Debevc, & Krneta, 2007). The latter approach is not fully suitable as it does not consider the content of the learning resource, its difficulty, granulation, etc. Therefore, most existing learning object standards for learning object description, structure, and interactivity are based on learning object metadata.

The Learning Object Metadata standard (LOM) is a widely-accepted standard for the description of learning objects (Consortium & others, 2006). It has nine categories for learning object description: general, life cycle, meta-metadata, technical, educational, rights, relation, annotation, and classification. Each category has several components (Fig. 2.6). The description components of a learning object and its values might be used as base data for learning object pedagogical suitability estimation.

To utilize the LOM standard, its components and their values, such as were possible to obtain, are mapped to e-learning elements. Then, these values can be used to match learning objects with individual students in accordance with their learning styles (Table 2.4). The intersection between the learning object description's possible value and the dominant learning style illustrates the achieved resource r score for the component C_i , where i is one of five learning object description components.

E-learning environment element properties are closely related to e-learning styles; therefore, the type of the resource is associated with active and reflective learning types (type coefficients in Table 2.4 are associated with the active learning style), and the form of the resource affects additivity to visual or verbal learning style (form coefficients in Table 2.4 are associated with the visual learning style), content affects students of sensing and intuitive learning style (content coefficients in Table 2.4. are associated with the sensing learning style), sequence of the learning material forms the learning experience of sequential and global learning style students (sequence coefficients in Table 2.4. are associated with the sequential learning style). The non-presented values, in comparison to Fig. 2.8, have a coefficient of 0 for all analyzed properties. Therefore, it is not included in the table.

Table 2.4. Mapping of learning object description values to e-learning environment components and student learning styles

Learning object description		E-learning environment element property			
Component	Value	Type	Form	Content	Sequence
Interactivity type, C_1	Active	1			
	Expositive	0			
	Mixed	0.5			
Interaction level, C_2	Very low	0			
	Low	0.25			
	Medium	0.5			
	High	0.75			
	Very high	1			
Learning resource type, C_3	Exercise	1			
	Simulation	1			
	Questionnaire	1			
	Diagram	0.25	1		
	Figure	0.25	1		
	Graph	0.25	1		
	Slide	0	0.25		
	Table	0	0.5		
	Exam	1			

End of Table 2.4

Learning object description		E-learning environment element property			
Component	Value	Type	Form	Content	Sequence
	Experiment	1			
	Problem statement	1			
	Self-assessment	1			
	Lecture	0.25	0.25		
Semantic density, C_4	Very low			0	
	Low			0.25	
	Medium			0.5	
	High			0.75	
	Very high			1	
Relation, C_5	Required resources				1

The pedagogical match M between the learning resource r and the student s is estimated by the following formula (16).

$$\begin{aligned}
 M(r, s) = & 1 - \left(w_{active} \times \left| \frac{\sum_{k=1}^3 C_k(r) - LS(s, active)}{3} \right| + \right. \\
 & w_{visual} \times |C_3(r) - LS(s, visual)| + w_{semantic} \times |C_4(r) - LS(s, semantic)| + \\
 & \left. w_{sequential} \times \left| \frac{\sum prr(r, s)}{\sum ttr(r)} - LS(s, sequential) \right| \right). \quad (2.1)
 \end{aligned}$$

In the equation, w_i is the weight that sets the importance of a learning style, it may vary for each subject or group from 0 to 1, and the sum of all the weights should be equal to 1; $C_k(r)$ is the learning object r description component k value score (estimated by Table 2.9); $LS(s, style)$ is learning *style* score (probability to fit this learning style) for students s ; $ttr(r)$ is the number of learning object r relationships to required learning objects (if there are no such relationships, the $ttr(r)$ value is assigned to 1) and $prr(r, s)$ is the number of learning object r relationships to required learning objects, which are already taken by student s .

The higher the $M(r, s)$ value, the better the match between the resource r and the student's profile. The $M(r, s)$ value ranges from 0 to 1 and should be used to rank the student-resource match quality.

The estimation and filtering of the learning resource's pedagogical suitability based on students' already gained knowledge reveal a set of resources that should be presented to the student. However, it is important to arrange all the resources into a logical and student-oriented learning path. Therefore, the pedagogically suitable resources should be filtered additionally based on students' current knowledge and competence level.

Finally, the agent's interaction can be summarized in terms of percepts: LOM and student's profiles (see Table 2.8 and equation 18 for specific parameters). The agent's action is to output a measure of the learning object's suitability for a particular student: the value of pedagogical match M .

There are several promising approaches to the learning path design: traditional solutions, where each learning material is finished with a question-and-answer session or knowledge evaluation test to the sequence of learning activities (Yang, Li, & Lau, 2010); competence tree transformation based design that allows students to choose different forms of the learning path (Margienè & Ramanauskaitė, 2021); reinforcement learning models for the learning path selection (Li, Xu, Zhang, & Chang, 2021). In each e-learning system, the type of learning path design can be adjusted based on the student's needs. The agent-based learning system model lends itself to be used in combination with the learning design approach by setting which one of the multiple learning scenarios should be assigned to a specific course.

2.4. Comparison of the Proposed System to Existing Solutions

The concepts of personalized intelligent e-learning systems and agent-based architectures are not novel (Dorri, Kanhere, & Jurdak, 2018). However, the existing personalized e-learning systems are limited in their capabilities or personalization level. This is partly affected by the lack of different perspectives or agents. The review of research papers (published in 2018 or later) on personalized e-learning architecture (Table 2.5) shows that usually two (at most four) programming agents are used for the e-learning system personalization.

The newly proposed multi-agent personalized intelligent e-learning system has a unique combination of agents, allowing multi-view personalization of students learning experience. Agents for students' learning style and behaviours estimation are oriented on the learners' experience personalization in the sense of environment, forms of learning resources, and activities. Also, the combination of learning style and behaviour agents leads to more accurate, multi-criteria user profile estimation. At the same time, close to the resource form, learning material is personalized based on the formative assessment and methodological suitability estimation. These agents ensure that students get learning material in the form that

is the most appropriate for them, and at the same time, content is optimized for students' current level of knowledge.

Table 2.5. Summary of research papers proposing the use of different programming agents for the personalized e-learning system implementation

Authors	Learning style identification agent	Learning analytics model	Learner profile control agent	Formative assessment agent	Pedagogical suitability estimation agent
(Kausar, Huahu, Hussain, Wenhao, & Zahid, 2018)	—	Learning behaviour agent	—	—	—
(Ciloglugil & Inceoglu, 2018)	—	—	Learner modelling agent	—	Learning object repository agents
(C. Agbonifo & A. Obolo, 2018)	—	—	User account agent	Testing item agent	—
(Filatov, Zolotukhin, Yerokhin, & Kudryavtseva, 2019)	—	Learning behaviours agent	—	Testing agent	—
(Assami, Daoudi, & Ajhoun, 2019)	Learning style agent	—	Students profile agent	Knowledge level agent	—
(Xu & Wang, 2006)	Modelling agent	Activity agent	Learner agent	—	Planning agent

2.5. Conclusions of Chapter 2

1. The proposed smart e-learning system will be supplemented by popular learning personalization methods, Felder and Silverman Learning styles, supported by learning analytics to mitigate its drawbacks and increase the personalization granularity.
2. Surveyed students ($N = 104$) did not show attachment to pen-and-paper assessment (24 people preferred this medium). There was a tendency towards the use of laptops or tablets expressed by 29 peo-

ple, 23 people preferred to use phones, and 27 people had no preference. Such results are encouraging for the implementation of computer-assisted testing in classrooms today.

3. Surveyed students expressed clear preferences for learning styles within the Felder and Silverman Learning Styles Model, with 45% of students expressing a strong preference for a certain learning style. This result partly explains the observed widespread use of learning styles for personalization in the literature review, despite its controversial efficacy.
4. The proposed e-learning system is built around formative assessment, leveraging highly effective, according to pedagogical studies, but currently underused constructive feedback principles. However, it will require novel algorithms to achieve this vision of superior adaptability.

Formative Assessment Based Personalized Content for E-Learning Environment

This chapter presents two algorithms for formative assessment, i.e., one for the use with the classical test theory and based on the Upper Confidence bound algorithm, and the other employing the Elo Rating and the latent trait theory to explore students' knowledge and suggest the best path for knowledge correction.

The proposed method and results were published in two scientific papers and presented at an international conference (Melesko, 2017; Melesko & Novickij, 2019; Melesko & Ramanauskaite, 2021).

3.1. Base for a Formative Assessment Agent

The literature review reveals the value of formative assessment in the educational process and the need for formative assessment agents based on good feedback principles. The Personalized Intelligent Learning System described in the dissertation is built around such a formative assessment agent. This agent is designed to provide the following advantages to the learning system: motivate students by demonstrating their newly acquired competencies, support their learning by indicating the competencies in need of improvement, and provide a perspective to

students and teachers by eschewing grades in favour of corrective learning opportunities. It would also encourage all students to acquire all competencies rather than a passing grade.

The principal challenge for creating a formal assessment agent is negotiating the time and effort cost associated with evaluating students' competencies. However, most advancements in assessment methodologies are in the realm of summative assessment, such as the Item Response Theory and other computer-assisted techniques. Thus, to address this need, two novel solutions are offered in this chapter, one is based on Upper-Confidence Bound, and another that adds difficulty measure to items based on the Elo rating. The suggested algorithms for computer-assisted formative assessment have a common purpose: to reduce the cost of test creation by eschewing the need for pretesting (item difficulty calibration) and to reduce student workload by making the assessment shorter. These algorithms select the most suitable quiz composition for a student and ultimately return a list of topics or even items that the student should study to improve. Such results are immediately useful for corrective action on their own and can be further supplemented by secondary personalization agents.

3.2. Upper-Confidence Bound Based Algorithm

The problem of choosing the next question during a formative quiz can be modelled as a MAB problem (Melesko & Novickij, 2019). To the best of the author's knowledge, no applications of MAB algorithms for formative assessment item selection have been proposed. Research supports that the key elements of formative assessment are feedback, awareness of gaps between current students' knowledge and their aims, and where to go next to alleviate those deficiencies. The proposed assessment method addresses this need by quickly identifying the lacking areas of knowledge and thoroughly exploring them to assist further learning. The process can be viewed through the lens of Hattie's three-question feedback model (Hattie & Timperley, 2007). To utilize the algorithm, the teacher must first form topics or competencies ("Where am I going?"). The algorithm quickly identifies the lacking areas of knowledge ("How am I going?") and explores the topic in detail, helping further instruction ("Where to next?"). This approach, in combination with presently widespread mobile devices, has the potential to mitigate the aforementioned issues, such as test creation complexity and long test times, while providing accurate formative assessment data compatible with Hattie's three-question feedback, Competency-Based Learning, and Assessment methodologies.

3.2.1. Modelling Formative Assessment

When tutoring, a teacher will often engage in a dialogue with a student. The teacher may ask the student a series of formative questions to diagnose the gaps in the student's knowledge. Assume the material consists of two topics, and the teacher asked five questions on each topic. The first topic has two incorrect answers, so the knowledge appears to be in a worse shape than the knowledge on the second topic with one incorrect answer (Table 3.1).

Table 3.1. Rudimentary two topic quiz model

Question	1	2	3	4	5	6	7	8	9	10
Topic 1	+		+	−		+				−
Topic 2		+			+		+	−	+	

If there is time for five more questions before proceeding with didactic instruction, the teacher must face the dilemma of which topic they should explore with further questions? If the two incorrect answers to questions on the first topic are attributable to bad luck due to the small sample size, should the teacher explore the first or second topic? What if there are more than two topics?

The family of bandit algorithms is designed to cope with uncertainty by balancing exploration and exploitation (Girisgin et al., 2015). However, when applied to formative assessment, the exploitation component is non-obvious, as ultimately, the goal is to explore the knowledge of the student. The algorithm should probe and explore the different topics and engage in focused questioning, exploiting those which are possible in most need of instruction. This presents an opaque bandit problem where a unique answer, reward, is observed at each round, in contrast with the transparent one where all rewards are observed (Vermorel & Mohri, 2005). Thus, in the context of assessment, a sequential allocation problem is obtained when the assessor must choose from many questions on multiple topics or bandits and must repeatedly choose a topic to explore or which bandit's arm to pull. When choosing the next question to ask, the decision should depend on the history of already known answers. Then, a policy is the mapping from the individual history of the student to actions (questions to be asked of the student).

Suppose a student's knowledge of several topics $T = \{1, 2, \dots, k\}$. The reward in the case of a multiple-choice quiz with either a correct or incorrect answer to each question $X_r \in \{0, 1\}$ is binary-valued. Each topic corresponds to an unknown probability distribution. There exists a vector $\mu \in [0, 1]^k$ such that the probability that $X_r = 0$ given the algorithm chose topic $T_r = t$ is μ_t . This kind of environment is called a stochastic Bernoulli bandit. If the mean vector associated with the environment was known, the optimal policy is to always choose a question on one

topic $t^* = \operatorname{argmin}_{t \in T} \mu_t$. This will result in the exploration of the weakest area of students' knowledge to aid in further instruction (Table 3.2). In the pseudo-code, U_s denotes a set of all test items which were already used and have been served to the student s ; w_{it} is a weight that indicates how much the item i measures knowledge in topic T ; t is a set of all items that can be used to measure knowledge on the topic; finally, b_t is the lower confidence bound for a given topic. The explanation for the equation follows next.

Table 3.2. Pseudo-code for test item selection with the application of a UCB algorithm

Algorithm UCB Formative Assessment	
Input item-topic weights w_{it} , number of previously served items n , number of previously served items for each topic N_t , student's answer to the selected item X_{si} . Output updated number of previously served items n , updated number of previously served items for each topic N_t	
1	// Find unused items i for each topic T with non-zero item-topic weights
2	For each <i>topic</i> do
3	For each <i>item</i> do
4	If $w_{it} \neq 0$ and $i \notin U_s$ then
5	$t \cup \{i\}$
6	// Calculate lower confidence bound for the student's topic knowledge b_t
7	For each $t \neq \emptyset$ do
8	$b_t = \mu_t - C \sqrt{\frac{\log n}{N_t}}$
9	Choose a topic with the smallest b_t value $T_c = \operatorname{argmin}_{t \in T} (b_t)$
10	Add the item to the list of seen items by the student, $U_s \cup \{i_c\}$

The main challenge of the task is finding the weakest topic of a student. To do so, the algorithm must explore different topics and exploit particular topics to obtain a more accurate estimation of the student's level of knowledge on that subject. This basic exploration–exploitation dilemma is the key to obtaining a good strategy. A heuristic principle for dealing with this issue chosen in this dissertation is optimism in the face of uncertainty, and an algorithm that operates on this principle Upper-Confidence Bound (UCB). The UCB algorithm is one of the simplest algorithms that offer sub-linear regret. The algorithm suggests choosing the action with the largest upper confidence bound, or in the case of this model, a topic with

the smaller lower confidence bound. Then the question number n chosen on a topic t will be

$$t_n = \underset{t \in T}{\operatorname{argmin}} \left(\mu_t - C \sqrt{\frac{\log n}{N_t}} \right), \quad (3.1)$$

where C is a constant that can be chosen to regulate the impact the second exploration component has on the choice of the topic, and N_t is the number of questions on the topic that has been asked so far. As the number of questions on the topic increases, so the uncertainty and the exploration term of the formula decrease (Girisgin et al., 2015). Thus, the algorithm will seek out the weakest topics of knowledge for a student; once identified, it will thoroughly question the student on said topics. This is pedagogically valuable because once identified, the lacking topic knowledge can be corrected. In addition, the algorithm will gather more fine-grained information on the weakest topic by “exploiting it,” which will be useful in post-assessment knowledge correction. In the case when the item pool of the topic is exhausted, the algorithm chooses the topic with the second smallest value as estimated by equation (17).

3.2.2. Simulated Students' Experiments

Several experiments with simulated students were carried out before proceeding with testing using empirical data. The number of simulated students for each experiment was 1000 unless stated otherwise. Each simulated quiz had several questions on two or more topics. Then, each topic was represented as a vector of weights for each question in the quiz. Each weight would represent the relevancy of the question to the topic. In this dissertation, only experiments with binary weights, 0 or 1, were carried out. Moreover, each question was assumed to belong only to one topic. The number of question items on each topic was set to be equal. Each simulated student had a vector equal in length to the number of the questions in the quiz, where each element represented the knowledge on one question, either yes or no. For all simulation experiments, unless stated otherwise, the inter-topic correlation of answers was random. The probability that the student would know the answer on a topic p_t had uniformly distributed random bias between 0 or 1.

3.2.3. Real Students' Assessment Methodology

In this study, the feasibility of the application of the UCB algorithm to formative assessment is explored using the data collected from a 60-question quiz covering 15 subtopics (i.e., network topologies, networking devices, Internet Protocol version 4, Ethernet, cloud computing services). The assessment was held at Vilnius

Gediminas Technical University, Lithuania (on 25 April 2019). The test length and item pool size were set to 60 questions to keep the quiz length close to an hour. The number of topics was set to 15 because it is the number of lectures in the course. The quiz was designed to assess students' knowledge of basic computer networking and cloud computing technologies. In total, 104 undergraduates, sophomore, and junior (third year) students from seven different groups were tested. All students took the test at the same place and time (Saulėtekio al. 11, Vilnius, from 12:30 to 13:30). The question pool contained questions of varying difficulty (the hardest question was answered by 16 students, and the easiest — by 103). This was done to test the robustness of the algorithm, as it is meant as an alternative to methods that require item calibration. Therefore, the algorithm must be able to work with items of varying and priority unknown difficulty.

All questions in the 60-question quiz were multiple-choice questions with four options, only one of which was correct. Questions on the same topic were made to never appear in consequence to lessen the impact of deductive reasoning over the knowledge of the subject. Students were not allowed to assist each other during the assessment. All students were also informed that if they so desired, the test would have no summative impact on their grade and would serve an exclusively formative function. Every student gave written permission allowing their anonymized data to be used for scientific research. The answers to questions were aggregated in a comma-separated file (CSV), anonymized, and later processed using Python software written for this experiment.

The true knowledge of a student on each topic, the ground truth, was calculated by dividing the number of correct answers within a topic by the total number of questions within that topic. In the experiment with the participation of real students, the number of correct answers was known because all students were required to answer all questions in the item pool. After complete knowledge of the test material for each student was known, the algorithms would question the database. The accuracy was then measured as a relationship between an estimated student knowledge from the incomplete information accessed by the algorithm and the complete information in the database. Formulas used for accuracy calculation are provided in the following statistical analysis section.

3.2.4. Statistical Analysis

The accuracy (performance) of the test was established based on Positive Predictive Value (PPV), which defines the probability of supplying the correct learning material to a student after the formative assessment and evaluated according to the formula, $PPV = TP / (TP + FP)$ for one student, where TP is True Positive or the number of correctly identified weakest topics for a student. The number of weakest topics is not always one, because topic proficiency is assumed to be equal

to an expected value of an answer to a topic's question, a Bernoulli variable $E(T) = p_t$. This value can be the same for several topics; in that case, it is assumed that the student would equally benefit from instruction on any of the topics. The *FP*, False Positive, is a number of incorrectly identified topics for which the topic mean μ_t is larger than the smallest mean, μ_m . For the group of students, the average of individual accuracies was taken.

Where applicable experimental results were expressed as mean $\pm$ Standard Error of the Mean (SEM). It is calculated by dividing the standard deviation by the square root of the sample size. The correlation matrix of questions for heatmap visualization was computed using the Pearson correlation coefficient using Pandas Python data analysis library.

The variance of answers to questions on one topic in simulation experiments was calculated using $\text{Var}[T] = p_t(1 - p_t)$ formula for Bernoulli distribution. The probability of a correct answer on the topic p_t was known and controlled for each topic T to observe its effect on assessment accuracy. In the experiments where real students' variance was computed using the same formula, p_t was estimated using the formula

$$p_t = \frac{\sum_{s=1}^{n_s} \sum_{q=1}^{n_{qt}} a_{sq}}{n_s n_{qt}}, \quad (3.2)$$

where s is a student, q is a question within a topic, and a_{sq} is an answer of a particular student on a particular question within a topic, n_s and n_{st} are the number of students and questions within a topic, respectively.

3.2.5. Results of Formative Assessment

It starts by presenting a set of simulations to systematically explore different properties of formative assessment using the UCB algorithm. The UCB algorithm efficiency is dependent on the constant C , which regulates the impact of exploration term on the topic choice, as can be seen in Formula (3.1). To analyze this impact and to choose the most suitable C for assessing real students, several experiments with synthetic students were carried out.

A cursory result for the C impact on the assessment accuracy can be seen in Fig. 3.1. The UCB algorithm shows better performance for every plotted constantly over randomly asked questions. Note that the algorithm serving random questions never asks the same question twice of the same student; thus, it achieves 100% accuracy after serving all 64 question items. The exploration can have significant impacts on the accuracy, both positive and negative, as seen from better performance of $C = 0.45$ over $C = 0$ and $C = 1$.

Fig. 3.1 shows that the UCB algorithm, when applied to formative assessment, has the potential to significantly shorten test length. With a larger constant, the algorithm displayed relatively bad accuracy at quiz lengths from about 15 to 30 questions. This can be explained by the failure to exploit known bad topics to further explore topics about which little data is known. Finally, as seen from the best performance of $C = 0.45$ exploration component does have a positive impact on assessment accuracy.

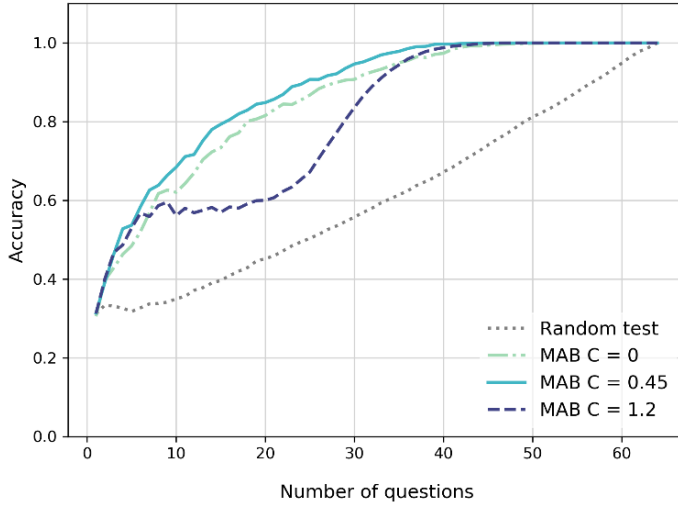


Fig. 3.1. Assessment accuracy is a function of the number of questions in a quiz. A quiz contained a pool of 64 questions on 16 different topics and shows the average accuracy for a class of 1000 synthetic students

The following experiment with synthetic students was carried out to choose the appropriate exploration constant for the real quiz, which had 60 questions (to set its duration at about 60 min). The number of questions was set to 64 to observe the importance of exploration in realistic scenarios: 4, 8, and 16 topics (Fig. 3.2). In the experiment, we measured the minimal quiz length (the number of questions) required to achieve accuracy greater than 95% in the class of 1000 synthetic students. The experiment was performed for every constant value from 0 to 2, with a step of 0.1, and the results were plotted in Fig. 3.2. A conclusion we can draw from Fig. 3.2 is that for every practical topic number in a 64-question test constant can be set to 0.5 for optimal results.

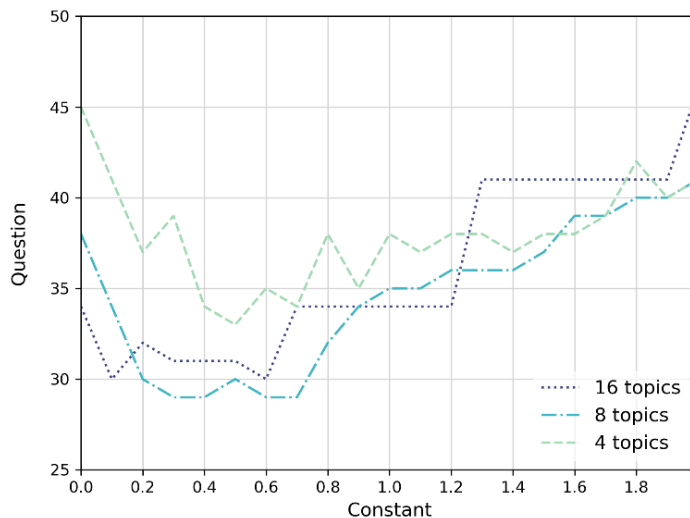


Fig. 3.2. Number of questions needed to achieve 95% accuracy as a function of exploration constant C for 4, 8, and 16 topics and 64 questions

A more thorough exploration of quiz space has been performed to gauge the practicality of implementing the UCB algorithm in different classroom situations. Table 3.3 demonstrates that the UCB testing method becomes more valuable as the pool of questions and the number of topics increase. The number of simulated students for this experiment was 500.

Fig. 3.3 presents the results from the experiment with real students ($n = 104$). A formative assessment quiz included items from 15 topics, with four question items in each topic, for 60 total questions. Results show that in this scenario, using traditional testing methods, the quiz length could be shortened from 60 questions to 55 if the goal was 95% accuracy in the weakest topic identification. With the use of UCB adaptive assessment, however, quiz length could be almost halved to 32 questions. Empirical results (orange and blue plots) are in line with projections drawn from simulations (grey plot). For this particular quiz and the group of students, UCB would offer a reduction in quiz length by 23 questions if the aim was the same ($>95\%$) accuracy.

Table 3.3. Reduction in quiz length necessary to achieve 95% accuracy conferred by using UCB algorithm with the optimal exploration constant over traditional random question approach. All numbers are given for the optimal C value, where T – number of topics; Q – total number of questions. The green colour indicates a large reduction in quiz length, and red indicates a lesser benefit

$\begin{matrix} T \\ Q \end{matrix}$	4	8	16	32	64	128	256	512
2	33%	0%	8%	0%	24%	-2%	22%	24%
4	0%	29%	31%	35%	36%	44%	43%	52%
8		0%	40%	45%	50%	54%	58%	60%
16			0%	57%	50%	59%	67%	71%
32				0%	65%	52%	64%	73%
64					0%	62%	56%	66%

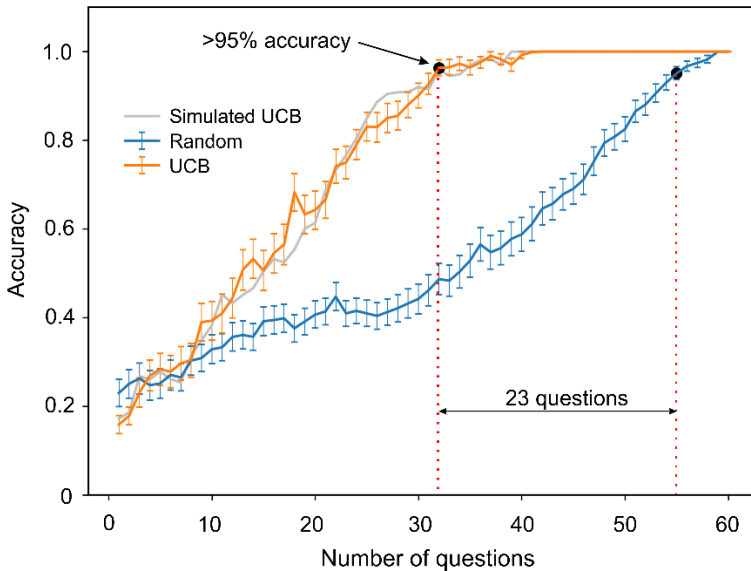


Fig. 3.3. Average accuracy of formative assessment for 104 students

The simulated Upper-Confidence Bound (UCB) line is for 104 synthetic students to show that synthetic experiments are congruent with reality. The error bars represent the Standard Error of the Mean (SEM), and the line plot shows the mean for the group. The SEM shows how closely a set of data represents that actual population. To quantify the similarity between simulation and obtained empirical result, the experiment was repeated 1000 times, each time computing the correlation between the newly generated and the tested class of 104 synthetic students

and real data. The average Pearson's correlation coefficient for the UCB test between the simulation and reality was 0.989 with a standard deviation of 0.006, as for the random test, the mean correlation was 0.989 with a deviation of 0.004. These numbers suggest that simulated results are indicative of real-life performance.

The ease at which the weakest topic of a student can be identified is dependent on how strongly the answers on the same topic are correlated. At an answer variance equal to zero, it is sufficient to ask only one question to know the student's knowledge of the rest of the items within the topic.

However, when answers are uncorrelated, estimation must be harder and might defeat the entire premise of the proposed UCB testing model. Thus, an experiment to measure the effect of knowledge correlation on items within one topic on method effectiveness was carried out. The impact of answer variance on testing accuracy can be seen in Fig. 3.4.

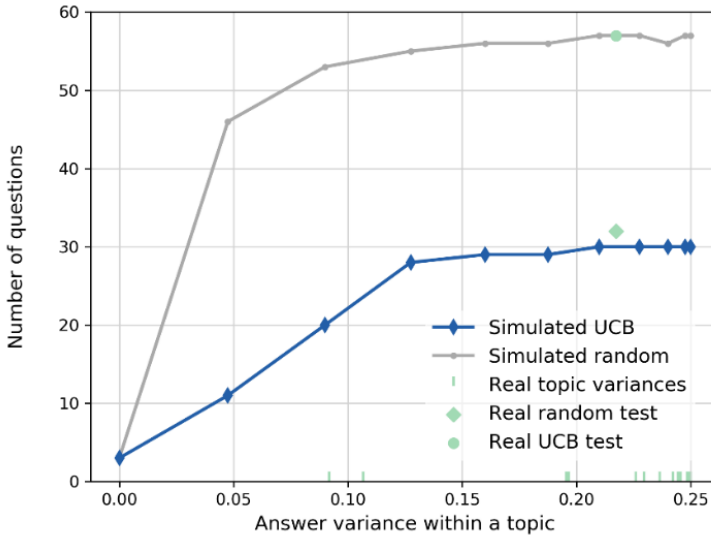


Fig. 3.4. Impact of answer variance within topics on testing accuracy (the number of questions needed to achieve 95% accuracy in a 60-question 15-topic quiz). The green markers indicate variance calculated using data from real tests; square and circle position along the x-axis represents the average variance

As anticipated, identifying weak topics is trivial for unrealistically strongly correlated answers. However, even for uncorrelated answers, UCB performs twice as good as random questioning. Also, plotted in Fig. 3.4 are variances calculated from answers of real students, $N = 104$ (the 15-topics 60-question quiz).

A correlation matrix for answers between different questions was constructed and is shown in Fig. 3.5. It indicates relatively low answer correlation even for inter-topic questions. Answers on each topic were grouped for this illustration (i.e., answers 1, 2, 3, and 4 all belong to the same topic). The matrix, in accordance with Fig. 3.5, shows a low general correlation of answers within one topic, especially for items 20–24 (questions about the physical layer of the OSI model).

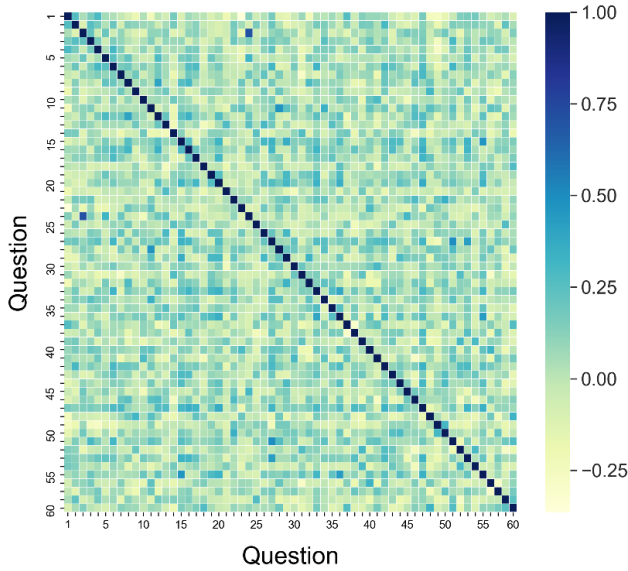


Fig. 3.5. Correlation matrix (Pearson coefficient) for quiz answers

To assess shorter quiz lengths and algorithm behaviour at a 95% accuracy value, a more detailed look at accuracy distribution is provided in Fig. 3.6. The figure shows that for the 60-question 15-topic quiz, the algorithm rarely displays accuracies between 50% and 99% for individual students.

At quiz lengths between 10 and 32 questions, a substantial portion (4–38%) of students were assessed very poorly, with less than 1% accuracy. Similarly, even at >95% accuracies, there can be a minority of students with incorrect weakest topic estimates. The root causes may be a small correlation of answers (Fig. 3.4 and 3.5) in the sample used for testing and a small number of items within a topic (only 4). There was an increase in the fraction of students with wildly wrong estimates (accuracy <1%) up to question 7. This is because the initially estimated knowledge on all topics is set to 0.5 in our implementation of the algorithm. The algorithm shows a steady increase in accurately (>99%) diagnosed students with no abnormalities.

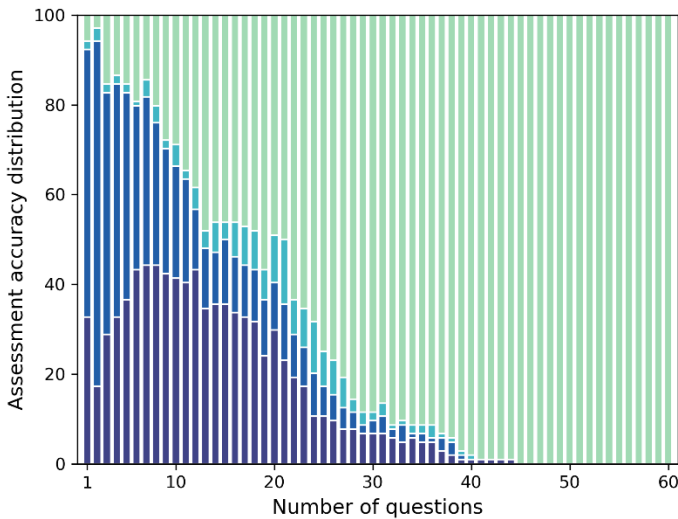


Fig. 3.6. Distribution of assessment accuracies for students, where dark blue <1%, blue 1–50%, cyan 50–99%, and green >99% accuracy

3.2.6. Discussion of UCB Assessment Results

The UCB algorithm has the potential to significantly reduce assessment length without the loss of accuracy. Even for very short tests with few topics, the algorithm offers a significant reduction in test length (Table 3.2). However, there is no advantage to quizzes with one question on each topic, at which point the notion of topic loses its pedagogical meaning. There is evidence that time allocated to study positively affects student performance (Bratti & Staffolani, 2013); thus, reducing the time spent on assessment is desired. The experiments with real students support the conclusions drawn from the simulations (Fig. 3). Such a big change in quiz length has the potential to change the dynamics in the classroom because the instructor would not need to spend an entire lesson just for formative assessment. This, in turn, may increase opportunities to provide personalized feedback to students linked to better performance (R.J. Marzano, Gaddy, & Dean, 2000; Shute, 2008). This advantage comes at no cost in quiz creation complexity, unlike IRT and Elo rating-based systems, where quiz creation can be prohibitively complex in some situations due to item calibration problems (Pelánek, 2016; Wainer, 2001). Because of this fundamental difference, the performance of UCB to IRT and Elo-based algorithms is not compared. Such a comparison would not discredit either approach. If UCB performed worse (as is safe to assume), its simplicity of use and independence from question item calibration would make it an interesting alternative to traditional assessment.

The exploration component of the algorithm becomes more important with the increase in the number of items within a topic. This can be observed in the bad performance of the algorithm in the 4-topic quiz when C was set to 0 in Fig. 2. It took 45 questions to reach the >95% assessment accuracy. Also, unintuitively, it takes fewer questions to identify the weakest topics for students when there are more topics and the size of the item pool is kept constant.

Compared to IRT and Elo algorithms, UCB will obtain less information about the student in a mathematical sense (C. J. Huang, Chen, & Chen, 2009a; Pelánek, 2016), and this can be seen as a disadvantage. However, not all information is equally pedagogically valuable (Nicol & MacFarlane-Dick, 2006; Sadler, Royce, 1989). For example, assume we are assessing students' knowledge on two topics. We are nearing the end of the quiz and have determined that knowledge on the first topic is adequate but lacking on the second topic. Because of the nature of the UCB algorithm topic, topic two is explored more than topic one; therefore, the expectation is to gain less information by exploring the second topic. However, from a pedagogical point of view, the information on topic two can be more valuable in addressing the assessment needs according to the three-question feedback model (Hattie & Timperley, 2007). According to the assessment, topic two requires instruction. While planning to teach it, the extra diagnostic data can be used to save time and effort by not teaching what the student already knows. Meanwhile, there is no immediate use for the more precise data about the first topic.

Simulation data presented (Fig. 3.1) and empirical results (Fig. 3.3) suggest that the UCB assessment approach can offer a significant reduction in quiz length for any practical item pool size and inter-topic variance of answers. This is an important result because it indicates the suitability of the method for any grouping of questions regardless of how correlated the answers are within one topic. Method effectiveness stays almost constant for any observed answer variance, which makes it easy to predict quiz length. This allows a teacher to group questions on each topic as they see fit according to the syllabus and the learning material at hand, regardless of the existence or lack of a common latent trait underlying the items within a topic. This separates the UCB method from IRT and Elo-based alternatives, which depend on the assumption of a common latent trait (Doebler et al., 2015; Wainer, 2001).

The exploration constant C has a significant impact on quiz length. For example, for $C = 0.6$ quiz length was 30 items, for $C = 2 - 46$ in a 16-topic 64-item quiz (Fig. 3.2). If C is further increased, the UCB algorithm degenerates to simple topic enumeration. It will sequentially choose each topic, disregarding the negligible impact of the μ_t term in Equation 3.1, as it will be dominated by the $C (\log n / N_t)^{1/2}$ term. Finally, the optimal C constant depends on the number of items in the quiz pool, a lesser number of topics, and even lesser targeted accuracy. When assessment accuracy is measured against the answers to every single item

in the pool, the most important factor is item pool size, and C should be chosen accordingly. In a classroom, the number of items in the pool is usually determined by the timeframe available for the quiz. As indicated by Fig. 3.2, the number of topics can be chosen freely as it has a lesser impact, and one value can satisfy the quiz length minimum condition for a wildly different topic number 4, 8, and 16 with 64 total items.

The most important factor that decides the practical value of an assessment method is its dependence on items used and their parameters, such as difficulty, student numbers, and competence. These factors are highly dynamic and hard to estimate. This is what is holding back the leading assessment methods based on the latent trait model, also known as IRT. These parameters have no impact on the exploration constant C and, as a consequence, on the proposed assessment method in general.

At quiz lengths between 10 and 32 questions, a substantial portion (4–38%) of students were assessed very poorly. This implies that using shorter assessments is morally questionable, as most students receive very accurate guidance while the rest are tutored on topics that they already know. This presents a problem for more important formative assessments (i.e., entire semester assessment). This property of the algorithm can be offset by a small increase in quiz length, as seen in Fig. 3.6.

A presented novel approach to formative assessment based on the UCB algorithm shows promising results when compared to traditional assessment methods. This approach can significantly reduce quiz length without a reduction in accuracy. For quizzes with an item pool equal to eight questions, the reduction is 29%, and for quizzes with an item pool of 512 questions, it is 73%. The variance of answers to questions within the same topic has little impact on assessment accuracy for empirically observed values (0.1 to 0.25); thus, the algorithm is suited for situations where items do not necessarily measure the same latent skill or trait. However, the distribution of student accuracies within a class is non-normal. Even at high average class accuracies (95%), the majority of accurately assessed students is offset by a small minority of students for whom the weakest topics were incorrectly identified. To offset this property of the algorithm, it is recommended for educators to target the >99% accuracy for course-crucial UCB formative assessments. It is believed that a UCB-based formative assessment has pedagogical potential for practical applications and should be further explored. Unlike IRT and Elo rating-based assessment methods, UCB-based assessment requires no question item calibration and does not depend on the debatable premise that different questions measure the same latent trait. As a consequence, the UCB method belongs to a different, sparsely explored class of easy-to-implement and maintain formative assessment solutions. It may prove to be a fresh and viable alternative to traditional linear assessment in situations where IRT and Elo methods were deemed too complex to implement and maintain. In the future, a comparative

study of the UCB assessment method with established item calibration-dependent methods may be of interest.

3.3. Elo-Rating and Multi-Armed Bandit-Based Algorithm

Elo-Rating and Multi-Armed Bandit-based algorithms can significantly reduce test creation complexity compared to IRT solutions and reduce test length (Melesko & Ramanauskaite, 2021). They combine the advantages of Elo rating systems, such as dynamic calibration of item difficulties, guesses parameters, and compatibility with latent trait theory, with an efficient exploration–exploitation trade-off solution. Unlike CCT-based tests and UCB-based algorithms explored in the previous section, the Elo–Bandit algorithm considers varying item complexity and item guess probabilities. The Elo–Bandit algorithm simultaneously measures item difficulties and students’ multivariate proficiencies as it explores the item pool. The algorithm functions in accordance with Hattie’s feedback model (Hattie & Timperley, 2007). First, the instructor forms an item pool divided into topics or competencies (“Where am I going?”). Second, students partake in formative assessment where questions are chosen and dynamically calibrated by the proposed algorithm (“How am doing?”). Third, the algorithm can reveal with weakest areas of knowledge and suggest the easiest items, learning objects, within that area of knowledge to learn (“Where to next?”).

This hybrid testing algorithm is tested using simulations and is applied to real student data. Several components are needed to apply and evaluate the algorithm: a way to model and generate synthetic students, a way to assess data from real students, components that will perform assessment (testing), a benchmark estimation component, and components for measuring assessment performance against a baseline. The schematic representation of the utilized system is presented in Fig. 3.7.

3.3.1. Elo–Bandit Algorithm

The focus of the section is the newly proposed in this dissertation algorithm based on Elo rating and upper confidence bound algorithms, further referenced as the Elo–Bandit Algorithm (EBA). It builds on a less precise, classical test theory (Melesko & Novickij, 2019).

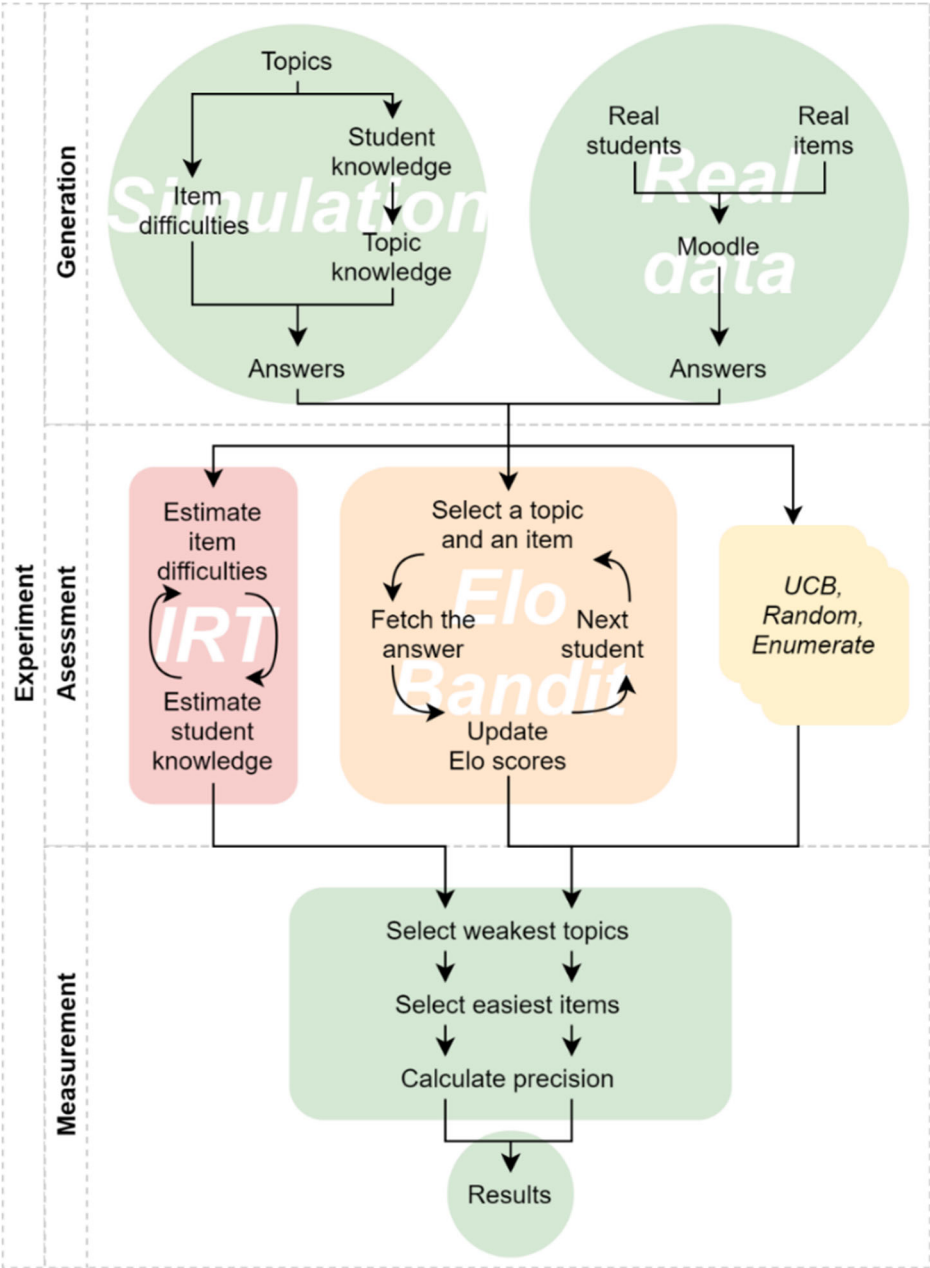


Fig. 3.7. Schematic representation of research methodology and its principal components

The problem of formative assessment can be modelled as taking an item pool on a number of topics as input, performing an assessment of students, and returning individualized learning suggestions for each student. Suggestions should reflect the topic in most dire need of instruction and item difficulty (easier items first). During the assessment, the algorithm must negotiate the exploration of topics on which student knowledge is yet unknown and investigate weak topics (“exploitation” in multi-armed bandit terminology) to precisely determine which items are known to a student. To evaluate students’ knowledge of each topic and assess item difficulty, the proposed EBA algorithm uses the Elo rating system. Table 3.4 shows the proposed EBA algorithm. In the pseudo-code, U_s denotes a set of all test items i_c which were already chosen and have been used by the student s ; w_{it} is a weight that indicates how much the item i measures knowledge in topic T ; t is a set of all items that can be used to measure knowledge on the topic; finally, b_t is the lower confidence bound for a given topic Elo. The explanation for the used equations follows below.

Firstly, the algorithm attempts to select the least familiar topic, then asks the easiest question within that topic. The weakest topic selection is performed using the following equation:

$$t_n = \underset{t \in T}{\operatorname{argmin}} \left(sElo_t - C \sqrt{\frac{\log n}{N_t}} \right), \quad (3.3)$$

where t is a chosen topic for a quiz question n , with $sElo_t$ difficulty estimation. The second parameter in brackets describes uncertainty in difficulty estimation and is controlled by exploration constant C , a total number of questions asked n , and a number of questions asked on the topic N_t . The uncertainty parameter decreases as the algorithm probes student’s knowledge on the topic and $sElo_t$ estimation becomes more accurate (Girisgin et al., 2015). Each student is assigned a starting Elo rating on each topic t , $sElo_t = 1000$.

The algorithm takes a pool of items as input, with weights describing their belonging to several topics and the guess probability factor. For example, a multiple-choice quiz item with one correct answer option out of four has a guess factor of $P_g = 0.25$. If no item difficulty calibration data is provided, EBA then assigns each item an initial Elo rating $iElo = 1000$. The probability of correct answer P_c to an item with $iElo$ score on a single topic t can be calculated using a shifted logistic function:

$$P_c = P_g + \frac{1 - P_g}{1 + 10^{\frac{iElo - sElo_t}{400}}}. \quad (3.4)$$

The base of the exponent and the denominator in the exponent are the same as in the Elo score used in chess matchmaking, where they were chosen for human readability.

Table 3.4. Proposed Elo–Bandit algorithm’s pseudo-code for the test item selection

Algorithm Elo–Bandit	
Input item-topic weights w_{it} , number of previously served items n , number of previously served items for each topic N_t , student’s Elo rating for each topic $sElo_t$, selected item’s difficulty Elo rating Elo_i , student’s answer to the selected item X_{si} .	
Output updated number of previously served items n , updated number of previously served items for each topic N_t , updated student’s Elo rating for each topic $sElo_t$, updated selected item’s difficulty Elo rating Elo_i ,	
1	// Find unused items i for each topic T with non-zero item-topic weights
2	For each topic do
3	For each item do
4	If $w_{it} \neq 0$ and $i \notin U_s$ then
5	$t \cup \{i\}$
6	// Calculate lower confidence bound of estimated Elo b_t
7	For each $t \neq \emptyset$ do
8	$b_t = sElo_t - C \sqrt{\frac{\log n}{N_t}}$
9	Choose a topic with the smallest b_t value $T_c = \underset{t \in T}{argmin}(b_t)$
10	Choose the item with the smallest Elo from T_c , $i_c = \underset{i \in T_c}{argmin}(Elo_i)$
11	Add the item to the list of seen items by the student, $U_s \cup \{i_c\}$
12	Calculate the probability of a correct answer, $P_c = P_g + \frac{1 - P_g}{1 + 10^{\frac{Elo_{i_c} - sElo_t}{400}}}$
13	// Update item and student knowledge Elo estimates based on the answer
14	For each topic do
15	$sElo'_t = sElo_t + w_{it}k(X_{si} - P_c)$
16	$Elo'_i = Elo_i - j(X_{si} - P_c)$

The new student's Elo rating on a given topic $sElo_t'$ after a student gives either correct, $X_{ans} = 1$, or incorrect, $X_{ans} = 0$, answer to an item is calculated based on the expected probability of correct answer:

$$sElo_t' = sElo_t + k(X_{ans} - P_c). \quad (3.5)$$

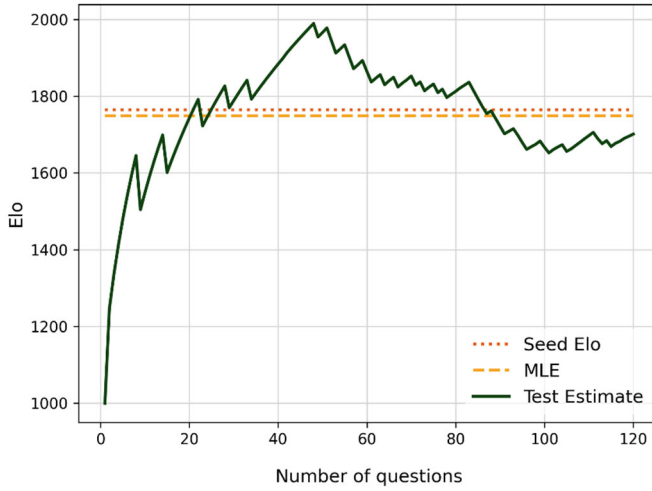


Fig. 3.8. Item Elo rating estimation approaching MLE estimation. Also plotted seed difficulty Elo used to generate items

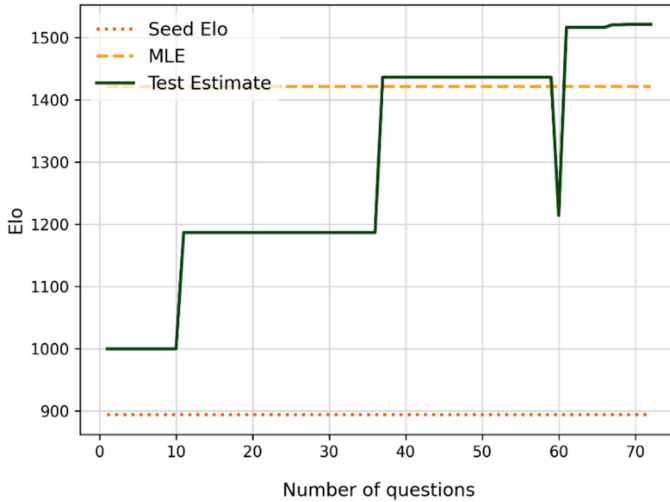


Fig. 3.9. Student's topic knowledge estimation approaching the MLE value

Constant k in Equation 3 specifies the sensitivity of the Elo rating to change or the maximum change the Elo rating can achieve given the most unlikely outcome. If k is small, the Elo estimation converges slowly; if k is large, then the Elo score oscillates with large amplitude around the true score without ever reaching it. A common solution is to use a decaying constant, which converges fast on the true value when there are few data points to estimate the Elo rating, but ultimately settles on a more accurate value than a large k value. There are more mathematically rigorous methods of modelling uncertainty inherent in the Elo rating estimation and using it as the basis for the constant, but they do not produce superior results (Nižnan, Pelánek, & Rihák, 2015). Convergence of item Elo rating estimation $iElo$ and student's topic Elo rating $sElo$ towards a value estimated using the Maximum Likelihood Estimation (MLE) method can be seen in Fig. 3.8.

Seed Elo used to generate synthetic data is also plotted in the figure. A more detailed description of seed Elo and synthetic student generation is presented in the following sections.

3.3.2. Other Formative Assessment Algorithms

Described in this section, the Elo–Bandit algorithm is compared to the Upper Confidence Bound-based algorithm and two classical formative assessment methods, choosing questions at random and sequential enumeration through topics.

The upper confidence bound algorithm identifies and explores the weakest topic within the quiz material without considering item difficulty or the order in which those items ought to be studied to improve topic knowledge (Melesko & Novickij, 2019). The UCB algorithm uses the same principle to select a topic as EBA, Equation 1. Except, it is based on the classical test theory and uses a mean answer value instead of the Elo rating estimation to denote students' knowledge of the topic. In practice, a UCB-based algorithm tends to first explore all topics, as an exploration parameter starts large. As the number of answered topic questions increases, the uncertainty parameter decreases, and the mark on the topic decides which topic shall be further explored by the algorithm.

The random algorithm simply selects a random question out of the pool of items yet unseen by the student. The enumerate topics algorithm servers a random question from each unexhausted topic in sequence. Its operation resembles UCB when its operation is dominated by the exploration component.

3.3.3. Simulated Students' Data

Before applying the algorithm proposed in this dissertation to assess real students, several simulations were performed. The number of synthetic students in each experiment was set to 120 to be close to an expected number of real students. A large

number would give EBA an advantage, as it would mean that each question item would receive more answers from students simplifying its estimation procedure. Each simulated assessment had items on multiple topics. Each topic was represented by a vector of weights equal in length to the number of items within the quiz. Each element of the topic vector denoted the item's relevancy to the topic. The number of items on each topic was set to be equal when possible. When the number of items was indivisible by the number of topics, the remaining items were assigned to separate topics.

The difficulty of simulated items, seed Elo, was generated using a normal distribution with mean, $\mu = 1000$, and standard deviation, $\sigma = 300$. Each synthetic student was assigned a general level of knowledge $sElo$ from a normal distribution, $\mu = 1000$ and $\sigma = 400$. Then that value was used as a basis for topic knowledge $sElo_t$, $\mu = sElo$ and $\sigma = 200$. Finally, the student's capacity to answer an item was calculated by estimating the Elo rating relevant to the item using item weights, w_{ti} , and topic knowledge:

$$aElo_{si} = \sum_{t=1}^T sElo_t \times w_{ti}. \quad (3.6)$$

Then, the correct answer was chosen as a Bernoulli random variable with the probability calculated using equation 2, guess factor $P_g = 0.25$.

3.3.4. Real Students' Data

Two datasets were used to test the presented algorithms. A newly presented 2021 dataset was from a formative assessment of real students, which was performed using the Moodle system. The data was collected using a quiz covering 15 topics of a cloud computing course. The size of the item pool was 80 items. The quiz was held on 11 February 2021 and had no summative impact on the final grade. In total, 106 undergraduate students from Vilnius Gediminas Technical University participated and agreed to share their data for scientific analysis. Each student was served all items from the pool, and their answers were recorded in a comma-separated file. The file was then used as a proxy for students to test different algorithms. This way, each algorithm received the same input data. All 80 items were multiple-choice questions with one correct answer, guess factor $P_g = 0.25$. The algorithm accuracy was measured in relation to item difficulty and student topic knowledge as estimated by MLE-based IRT with the entirety of the dataset as input. While algorithms had to operate on incomplete data, first asking each student one question, then two, and so on.

The second dataset explored in this section was already used in a publication (Melesko & Novickij, 2019). This dataset was included to also measure the Elo–

Bandit algorithm against previously presented the Upper Confidence Bound-based algorithm using the same data and same Classical Test Theory-based benchmarks. Data was collected in a quiz held at Vilnius Gediminas Technical University, Lithuania, on 25 April 2019. It had 104 undergraduate student participants, who each answered 60 items covering 15 topics. In the case of the CTT-based experiment, the topic knowledge of students was assumed to be the mean of all answers on said topic.

3.3.5. Validation Measurements and Metrics

Item Response Theory with parameters calculated using iterative maximum likelihood estimation method was used to provide a benchmark against which the performance of algorithms is measured. Benchmark values would represent idealized IRT results if a full-length IRT test was applied to calibration data. The MLE method maximizes the likelihood function so that the observed data is the most probable. It begins with some a priori value for a student's knowledge on each topic and item difficulty Elo ratings. Then students' knowledge ratings are recalculated to best reflect observed quiz outcomes. Afterward, item difficulty ratings are recalculated in light of the newly estimated student's topic knowledge. This process was iteratively repeated until the mean change in Elo ratings fell below one point.

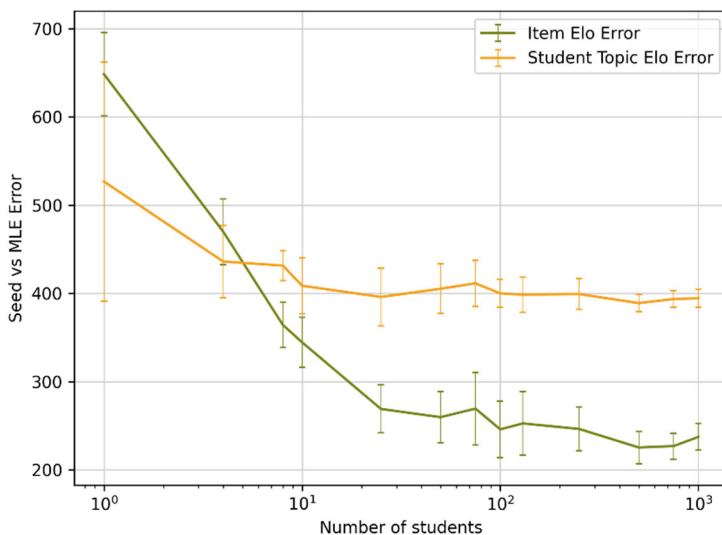


Fig. 3.10. Mean absolute error of MLE calculated values as compared to seed values used in the generation of synthetic students as a function of group size

For synthetic students and quizzes, where seed Elo rating values used in the generation were known, the mean error between MLE and seed Elo rating was calculated and is presented in Fig. 3.10. Error bars represent the standard deviation from the mean plotted. It is the mean of the absolute difference between seeded Elo values and estimated with MLE. The number of topics within the quiz was set to 15. The quiz length in the graph on the right was set to 80 items. The number of students in the graph on the right is 120. These numbers were chosen to resemble the parameters of the experiment with real student data. However, it is important to note that synthetic student generation parameters may be significantly different from latent trait distribution in real students.

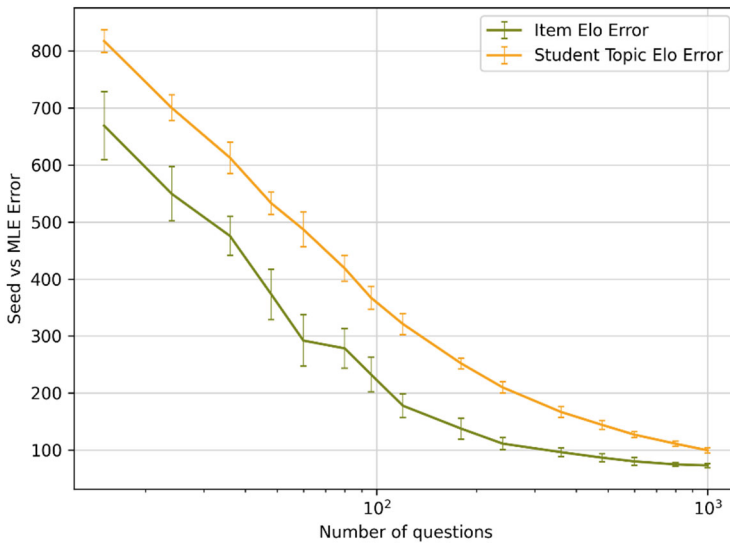


Fig. 3.11. Mean error of MLE calculated values as compared to seed values used in the generation of synthetic students as a function of quiz length

Assessment of accuracy for LTT-based experiments was measured against IRT MLE values. Latent values were not used because they are unknown to real students, and without reference to real values, synthetic scores can be manipulated: if the deviation in topic knowledge is small, separating the best-known topics from the worst will be harder. The set of suitable items according to IRT for a student was established by selecting topics least familiar to a student (lowest $sElo_i$ values) using complete knowledge of quiz answers, then the easiest items (lowest $iElo$ values) from those topics were added to the set. Delta values for the selection of suitable items were chosen so that an average of 15 items were chosen for a

student in synthetic and real experiments. This number was chosen for two reasons. Items in real quizzes mostly covered a narrow band of material. Selecting a few of them would mean students would quickly amend the gaps in their knowledge and would need to either retake formative assessment or study without guidance. Thus, wasting time. In this setting, a hypothetical average student with equal knowledge on every topic would have 1 item to study from each topic of the course. Also, the number of suitable items depends on the proficiency of a student. If the chosen number is too low, the formative assessment will fail to provide sufficient material to the most knowledgeable students.

Once the set of suitable items for each student is known, the assessment performance is measured using precision or the Positive Predictive Value (PPV). PPV describes the probability of providing suitable learning material to a student based on data gathered by formative assessment. PPV is calculated, $PPV = TP / (TP + FP)$. Where TP is True Positive, the number of items in the set of suitable items according to a formative assessment algorithm at some length of a quiz. TP is an intersection of the set of MLE-selected suitable items and algorithm-selected items. $TP + FP$ (false positive) is simply the number of elements in the set of items selected by the algorithm.

The accuracy assessment for CTT-based experiments targeting the weakest topic prediction was also measured using the precision metric. Except, item difficulty was not considered, and the student's topic knowledge was measured by a mark for that topic. The median number of suitable topics for a student is set to one to be consistent with the previous study.

3.3.6. Results of Elo-Bandit assessment experiments

The formative assessment precision in establishing personalized suitable items for each synthetic student is presented in Fig. 3.12 as measured against IRT MLE values. It presents mean precision $\pm$ sample standard deviation for ten groups of students taking ten different quizzes. Measurements were performed for ten synthetic groups with 120 students each.

The Elo-Bandit algorithm performs significantly better than UCB and classical test methods, i.e., random and enumerate. These methods quickly approach 50% precision as compared to IRT but never achieve 60% or higher precision. EBA achieves 80% precision by question 21, and 90% by 45, as marked in the plot. No algorithm, however, achieves ideal IRT accuracy, but Elo Bandit comes close to the ideal MLE-based IRT assessment with 98% precision. Simulated results depend on the assumptions about values and distribution of latent traits, which only approximate real students.

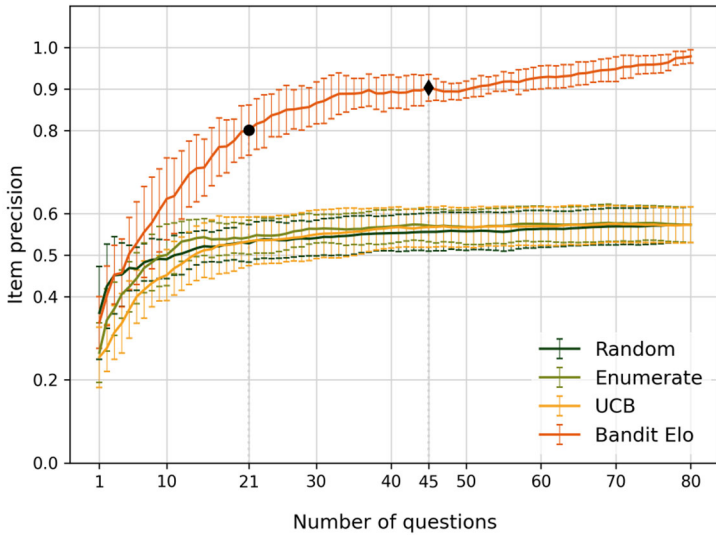


Fig. 3.12. Formative assessment precision selecting items as a function of quiz length for synthetic students. Simulated data from ten groups taking ten different quizzes, 120 students in each group

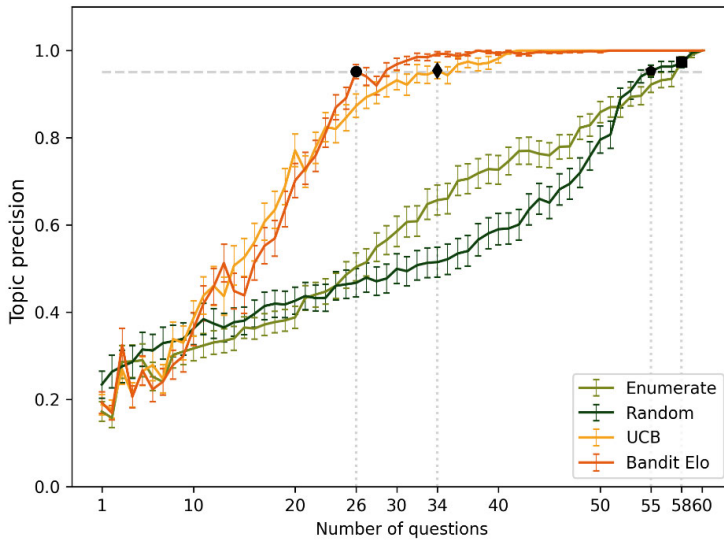


Fig. 3.13. Formative assessment precision selecting topics in CCT as a function of quiz length, 2019 dataset. The horizontal dashed line denotes 95% precision

The same algorithms were compared using real student data from the 2019 dataset (104 students) to measure their precision in selecting topics using classical test theory (Fig. 3.13). The figure presents the measured precision of suitable topic detection dependence on quiz length involving real students' results presented as the mean precision $\pm$ Standard Error of the Mean (SEM).

Elo–Bandit has virtually the same performance as the UCB-based algorithm in this application. Even though it breaches the 95% precision mark sooner, it lags in reaching 70% accuracy. Both algorithms offer about a 50% reduction in quiz length compared to classical testing methods when searching for the weakest topics. Enumerate and Random algorithms perform similarly. Enumerate has a better performance in medium quiz lengths.

A more detailed look at the precision distribution for UCB and Elo–Bandit algorithms for the 2019 dataset CCT model is presented in Fig. 3.14.

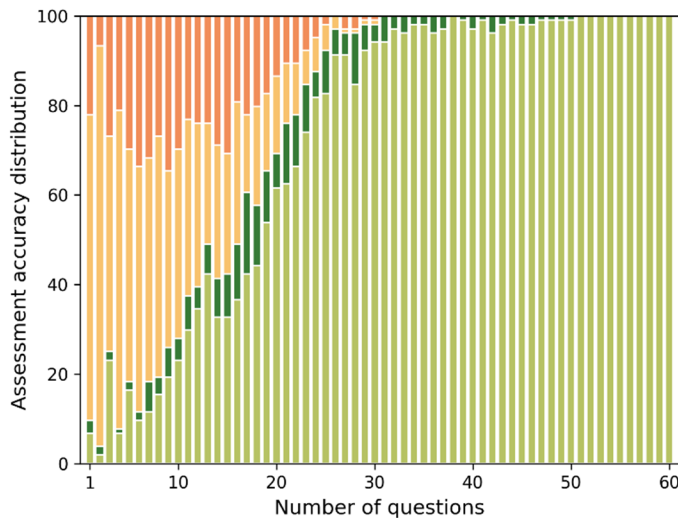


Fig. 3.14. Distribution of students assessed for topic precision for the Elo–Bandit algorithm. Where red <1%, orange 1–50%, dark green 50–99%, and green >99% precision

Both algorithms display very similar behaviour. Elo–Bandit shows slightly fewer students with wildly incorrect (less than one percent precision) estimates at quiz lengths shorter than 30 questions. This is because UCB uses the average score to measure topic knowledge. This estimate overshoots real value with a low sample size, while Elo rating estimation approaches the final estimate more smoothly. However, UCB demonstrates fewer students in the 50–99% category past 30 question mark because it uses the same mean score as the benchmark ground truth

CCT measure. Therefore, given full information, it will always reach full agreement with the benchmark.

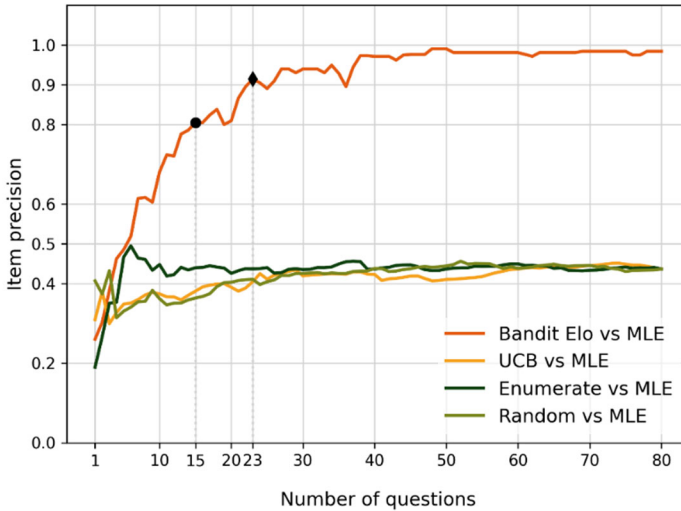


Fig. 3.15. Mean algorithm precision estimating personalized suitable items as measured against IRT MLE

The latent trait theory model was applied in estimating item suggestions on real student data, 2021 dataset (Fig. 3.15). A formative assessment quiz included 80 items on 15 topics and was taken by 106 students. Results confirm observations made using synthetic students. Elo–Bandit is a promising alternative to IRT-based tests, achieving almost the same precision of 97% with half the test length. Elo–Bandit has significantly better precision in finding suitable items than CCT-based algorithms. The Elo–Bandit algorithm shows better performance on real data compared to simulation. Offering more than twice precise results for tests longer than 15 questions. The 80% precision was achieved at 15 questions in comparison to 23 questions for simulation, and 90% precision at 23 in contrast to 45 question simulation value. Constants for the algorithm were calibrated using simulated data. A possible explanation is provided in the following discussion section.

The Pearson coefficient correlation matrix for the 2021 dataset is presented in Fig. 3.16. The correlation matrix demonstrates a generally low correlation even for inter-topic questions, except for items 64–69. Strongly correlated items would improve the performance of Elo–Bandit and UCB algorithms. Probing an item from a strongly correlated topic brings more information about other topic items,

so fewer questions are needed to identify familiar and unfamiliar subjects for a student.

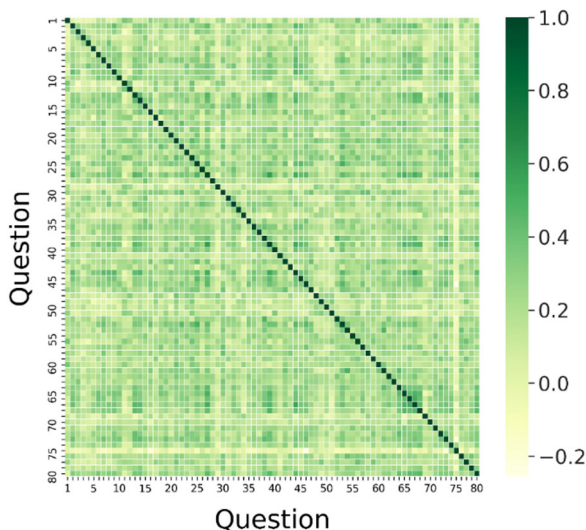


Fig 3.16. Item Pearson coefficient correlation matrix for 2021 dataset used to test the Elo–Bandit algorithm

3.4. Discussion of the Elo–Bandit algorithm

The Elo–Bandit algorithm shows promise as an alternative for IRT-based and CTT-based tests in formative assessment applications. Despite the pedagogical value provided by formative assessments (Dunn & Mulvenon, 2009; T. R. Guskey & Bailey, 2001), they have opportunity cost: assessments take away time that could be spent on directly studying course material through lectures and self-study. Time spent on each activity has a demonstrable effect on learning outcomes (Bratti & Staffolani, 2013). In addition to time savings, shorter assessment length may combat decision fatigue, possibly leading to more efficient post-test learning (Pignatiello, Martin, & Hickman, 2020). The Elo–Bandit algorithm provides an opportunity to benefit from formative assessment while saving time and mitigating fatigue. Potential quiz length reduction observed in the experimental setting was about 50%. In addition, informative assessment precision is less important than summative assessment. Imprecision in summative assessment may lead to failure to obtain employment or admission to the university to a deserving candidate. A degree of imprecision in formative assessment may cause a teaching system to serve a suboptimal learning object. Therefore, it might be rational to aim

at lower precisions, such as 80 or 90 percent, when it allows for significant quiz length reduction. In the presented 2021 case study with an item pool consisting of 80 questions, the median time of assessment amongst 106 students was 39 minutes or about 30 seconds per item. Settling for 80% or 90% precision would allow for cutting the median assessment time to less than 8 and 12 minutes, respectively. Such dramatic reduction in quiz length may allow instructors to incorporate formative assessment in the learning process where previously it was impractical.

Perhaps the key factor limiting the IRT-based assessment in classrooms is the need for calibration (Lord, 2012). A teacher cannot simply write and add items to a question bank as they teach. In addition to requiring a large student pool, unavailable to many teachers, item difficulty parameters are also dependent on the student background exacerbating the problem. The algorithm proposed in this dissertation does not require calibration. To the best of the author's knowledge, no similar formative assessment solution has been proposed and applied using data from real students. However, the Elo rating has been used to estimate the learning object's difficulty. In 2019, Mangaroska et al. published data from 22 students indicating positive results in the programming exercise recommendation system (Mangaroska et al., 2019). Pankiewicz and Bator found the Elo rating to be a superior task difficulty measure to proportion correct and learner feedback (Pankiewicz & Bator, 2019). Finally, based on simulations, Antal found that the Elo rating may be preferable to IRT for adaptive item sequencing in practice systems as it does not require item pretesting (Antal, 2016). These findings agree with the positive results obtained in this study: the Elo rating is a viable alternative to the IRT assessment's informative context.

Simulations showed significant algorithm performance differences in comparison with real data implying that generation parameters used for the synthetic student did not reflect students in the classroom. The algorithm's accuracy gets worse with a decrease in item difficulty and correlation. Real items were largely uncorrelated but more difficult, resulting in shorter quiz lengths. This also means that the Elo–Bandit assessment is most effective when used for unfamiliar material, as is often the case for formative assessment; a mean number of correct answers for real students was $\mu = 40$. But the proposed algorithm may be impractical if applied in the context of summative assessment without modifications.

To summarize, measuring topic level precision, with item difficulty ignored, the Elo–Bandit algorithm did not demonstrate increased performance to justify the increased complexity over the UCB algorithm but was still twice as quick as classical testing methods. In the context where item difficulty and precise item level suggestions are needed, CCT-based algorithms can't replace IRT and Elo-based assessment methods without further modifications, as their measurement accuracy was twice worse even at full quiz length. The Elo–Bandit algorithm may

also offer formative quiz length reduction in learning environments where item pretesting is not feasible.

The proposed Elo–Bandit algorithm can reduce formative assessment time by up to 70% when targeting 90% item level precision. This feature is student-friendly as it saves them time in comparison to traditional methods and helps to avoid fatigue. The proposed algorithm is also beneficial to instructors because it does not require item pretesting.

The synthetic data used in preliminary experiments did not accurately match data from real students. The assumed distribution of item difficulties and student proficiencies did not match with the one observed in empirical data. However, the proposed Elo–Bandit algorithm does not require normal distribution of item difficulties and demonstrated better performance with real data. This shows the algorithm’s flexibility and practical value in real classrooms.

The current version of the Elo–Bandit algorithm is designed to be brief and identify the weakest topics and easiest items for short-term learning path generation. In the future, it could be extended to generate a long-term learning path and retain a model of students’ knowledge between assessments. This would reduce assessment length even further and mitigate the need for repeated assessments providing even greater benefits to students in the form of saved time and mental resources.

3.5. Conclusions of Chapter 3

1. Novel application of the UCB algorithm is an efficient replacement for classical formative assessment methods, capable of quickly identifying weak areas of students’ knowledge. Aiming for 80% or 90% precision in a 15-topic 60-item quiz would allow for cutting median assessment time to less than 8 and 12 minutes, respectively, from 39 minutes.
2. The novel Elo–Bandit algorithm is a viable replacement for IRT-based formative assessment eliminating the need for costly item difficulty pretesting, with the added benefit of reducing the testing length: the reduction of up to 70% when targeting 90% item level precision.

General Conclusions

1. Literature review shows a disconnect between effective tutoring tools and their use in learning systems: feedback from the formative assessment is effective but underused, and learning styles are ineffective but popular.
2. The proposed smart learning system provides learning personalization based on established methods, such as Felder and Silverman learning styles and novel formative assessment algorithms without the expense of individual human tutors.
3. Novel UCB algorithm application is an efficient replacement for classical formative assessment methods, capable of quickly identifying weak areas of students' knowledge: aiming for 80% or 90% precision in a 15 topic 60 item quiz would allow for cutting median assessment time to less than 8 and 12 minutes respectively from 39 minutes.
4. The novel Elo-Bandit algorithm is a viable replacement for IRT-based formative assessment eliminating the need for costly item difficulty pretesting, with the added benefit of reducing the testing length by 70% when targeting 90% precision.

References

- Abdi, S., Khosravi, H., & Sadiq, S. (2021). Modelling learners in adaptive educational systems: A multivariate glicko-based approach. *ACM International Conference Proceeding Series*, 497–503. <https://doi.org/10.1145/3448139.3448189>
- Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2019). A multivariate elo-based learner model for adaptive educational systems. *ArXiv*.
- Alshmrany, S. (2022). Adaptive learning style prediction in e-learning environment using levy flight distribution based CNN model. *Cluster Computing*, 25(1), 523–536. <https://doi.org/10.1007/s10586-021-03403-3>
- Antal, M. (2016). On the use of ELO rating for adaptive assessment. *Studia Informatica*, 29–41.
- Assami, S., Daoudi, N., & Ajhoun, R. (2019). Ontology-based modeling for a personalized mooc recommender system. *Smart Innovation, Systems and Technologies*, 111, 21–28. https://doi.org/10.1007/978-3-030-03577-8_3
- Assessment Reform Group. (1999). Assessment for Learning: Beyond the Black Box. In *Reports from the Assessment Reform Group*. Qualifications and Curriculum Authority.
- Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3), 235–256. <https://doi.org/10.1023/A:1013689704352>

- Azevedo, R., & Bernard, R. M. (1995). A Meta-Analysis of the Effects of Feedback in Computer-Based Instruction. *Journal of Educational Computing Research*, 13(2), 111–127. <https://doi.org/10.2190/9lmd-3u28-3a0g-ftqt>
- Baker, F. B., & Kim, S.-H. (2004). Item Response Theory – Parameter Estimation Techniques - Second Edition, Revised and Expanded. *Measurement Theory in Action: Case Studies and Exercises*.
- Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23(1), 537–553. <https://doi.org/10.1007/s10639-017-9616-z>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy and Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Birnbaum, A. (1968). Some Latent Trait Models and their Use in Inferring an Examinee's Ability. *Statistical Theories of Mental Test Scores*.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *International Journal of Phytoremediation*, 21(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Bloom, B.S. (1971). Mastery learning: Theory and practice. In *Mastery learning*.
- Bloom, Benjamin S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Bloom, Benjamin S. (1969). Some theoretical issues relating to educational evaluation. *Educational Evaluation: New Roles, New Means: The 63rd Yearbook of the National Society for the Study of Education*, (part II), 26–50.
- Br. (2003). Science of the Brain. In R. Morris & M. Filenz (Eds.), *Neuroscience*. Retrieved from <http://brain.mcmaster.ca/BrainBee/Neuroscience.Science.of.the.Brain.pdf>
- Bratti, M., & Staffolani, S. (2013). Student Time Allocation and Educational Production Functions. *Annals of Economics and Statistics*, 103–140. <https://doi.org/10.2307/23646328>
- Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4321 LNCS, 3–53. https://doi.org/10.1007/978-3-540-72079-9_1
- Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1), 1–122. <https://doi.org/10.1561/22000000024>
- Burbaitė, R. (2014). *Išplėstiniai generatyviniai mokymosi objektai informatikos mokymuisi: koncepcija, modeliai ir realizacija*.

- C. Agbonifo, O., & A. Obolo, O. (2018). Genetic Algorithm-based Curriculum Sequencing Model For Personalised E-Learning System. *International Journal of Modern Education and Computer Science*, 10(5), 27–35. <https://doi.org/10.5815/ijmecs.2018.05.04>
- Campbell, J. C., Hindle, A., & Stroulia, E. (2015). Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data. *The Art and Science of Analyzing Software Data*, 139–159. <https://doi.org/10.1016/B978-0-12-411519-4.00006-9>
- Cesa-Bianchi, N., & Fischer, P. (1998). Finite-time regret bounds for the multiarmed bandit problem. *Proceedings of the Fifteenth International Conference on Machine Learning*, 100–108.
- Chen, X., Zou, D., Xie, H., & Wang, F. L. (2021). Past, present, and future of smart learning: a topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, 18(1), 529–552. <https://doi.org/10.1186/s41239-020-00239-6>
- Choi, Y., & McClenen, C. (2020). Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks. *Applied Sciences (Switzerland)*, 10(22), 1–17. <https://doi.org/10.3390/app10228196>
- Ciloglugil, B., & Inceoglu, M. M. (2018). An adaptive e-learning environment architecture based on agents and artifacts metamodel. *Proceedings - IEEE 18th International Conference on Advanced Learning Technologies, ICALT 2018*, 72–74. <https://doi.org/10.1109/ICALT.2018.00024>
- Clement, B. (2014). Online Optimization of Teaching Sequences with Multi-Armed Bandits. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, 269–272.
- Clement, B., Roy, D., Lopes, M., Oudeyer, P., Clement, B., Roy, D., ... Optimization, P. O. O. (2014). Online Optimization and Personalization of Teaching Sequences To cite this version : *DI: Digital Intelligence-1st International Conference on Digital Cultures*.
- Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2013). Multi-Armed Bandits for Intelligent Tutoring Systems. *ArXiv Preprint ArXiv:1310.3174*, 7(2), 20–48. <https://doi.org/10.5281/zenodo.3554668>
- Coelho, O. B., & Silveira, I. (2017). Deep Learning applied to Learning Analytics and Educational Data Mining: A Systematic Literature Review. *Anais Do XXVIII Simpósio Brasileiro de Informática Na Educação (SBIE 2017)*, 1, 143. <https://doi.org/10.5753/cbie.sbie.2017.143>
- Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Learning styles and pedagogy in post-16 learning A systematic and critical review. *Learning and Skills Research Centre*.
- Conlan, O., Hockemeyer, C., Lefrere, P., Wade, V., & Albert, D. (2001). Extending educational metadata schemas to describe Adaptive Learning Resources. *Proceedings of*

- the ACM Conference on Hypertext*, 161–162. <https://doi.org/10.1145/504255.504258>
- Consortium, I. M. S. G. L., & others. (2006). IMS Meta-data Best Practice Guide for IEEE 1484.12. 1-2002 Standard for Learning Object Metadata, Version 1.3 Final Specification. Retrieved November 8, 2021, from IMS Global Learning Consortium website: https://www.imsglobal.org/metadata/mdv1p3/imsmd%5C_bestv1p3.html
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 253–278. <https://doi.org/10.1007/BF01099821>
- Crocker, L., & Algina, J. (2008). Introduction to classical and modern test theory-Procedures for Estimating Reliability. In *Harcourt Brace Jovanovich College*. ERIC.
- Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, 3(OCT), 429. <https://doi.org/10.3389/fpsyg.2012.00429>
- Desmarais, M. C., & Baker, R. S. J. D. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, Vol. 22, pp. 9–38. <https://doi.org/10.1007/s11257-011-9106-8>
- Doebler, P., Alavash, M., & Giessing, C. (2015). Adaptive experiments with a multivariate Elo-type algorithm. *Behavior Research Methods*, 47(2), 384–394. <https://doi.org/10.3758/s13428-014-0478-7>
- Doleck, T., Lemay, D. J., Basnet, R. B., & Bazalais, P. (2020). Predictive analytics in education: a comparison of deep learning frameworks. *Education and Information Technologies*, 25(3), 1951–1963. <https://doi.org/10.1007/s10639-019-10068-4>
- Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-Agent Systems: A Survey. *IEEE Access*, 6, 28573–28593. <https://doi.org/10.1109/ACCESS.2018.2831228>
- Dunn, K. E., & Mulvenon, S. W. (2009). A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education. *Practical Assessment, Research & Evaluation*, 7. <https://doi.org/10.1002/ir>
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, 16(SUPPL. 1), 5–18. <https://doi.org/10.1007/s11136-007-9198-0>
- Elmarzaki, H. A. A. (2018). *Increasing the Architectures Design Quality for MAS: An Approach to Minimize The Effects of Complexity*. <https://doi.org/10.5121/csit.2018.80206>
- Elo, A. (1987). The Rating of Chess Players, Past and Present. In *New York: Arco*.
- Ennouamani, S., & Mahani, Z. (2018). An overview of adaptive e-learning systems. *2017 IEEE 8th International Conference on Intelligent Computing and Information Systems, ICICIS 2017, 2018-January*, 342–347. <https://doi.org/10.1109/INTELCIS.2017.8260060>
- Even-Dar, E., Mannor, S., & Mansour, Y. (2002). PAC bounds for Multi-armed Bandit

- and Markov Decision Processes. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2375, 255–270. https://doi.org/10.1007/3-540-45435-7_18
- Felder Richard M., N. C. S. U., & Silverman Linda K., I. for the S. of A. D. (1988). Learning and Teaching Styles in Engineering Education. *Engr. Education*, 78(7)(June), 674–681.
- FIDE. (2020). FIDE handbook. In *FIDE Charter* (p. March 1). Retrieved from <https://handbook.fide.com>
- Filatov, V., Zolotukhin, O., Yerokhin, A., & Kudryavtseva, M. (2019). Personalized Adaptation of Learning Environments. *Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL*. <https://doi.org/10.1109/CAOL46282.2019.9019525>
- Garcia-Rudolph, A., Opisso, E., Tormos, J. M., Madai, V. I., Frey, D., Becerra, H., ... López, J. (2021). Toward personalized web-based cognitive rehabilitation for patients with ischemic stroke: Elo rating approach. *JMIR Medical Informatics*, 9(11), e28090. <https://doi.org/10.2196/28090>
- Gareis, C. R. (2007). Reclaiming an important teacher competency: The lost art of formative assessment. *Journal of Personnel Evaluation in Education*, 20(1–2), 17–20. <https://doi.org/10.1007/s11092-007-9044-5>
- Giannakos, M. N., Mikalef, P., & Pappas, I. O. (2021). Systematic Literature Review of E-Learning Capabilities to Enhance Organizational Learning. *Information Systems Frontiers*, 1–17. <https://doi.org/10.1007/s10796-020-10097-2>
- Girisgin, S. A., Kalkan, E., Ergin, M., Keskin, F., Dundar, Z. D., Kebapcioglu, S., ... Cander, B. (2015). An experimental study: Does the neuroprotective effect increase when hypothermia deepens after traumatic brain injury? In *Iranian Red Crescent Medical Journal* (Vol. 17). [https://doi.org/10.5812/ircmj.17\(4\)2015.21233](https://doi.org/10.5812/ircmj.17(4)2015.21233)
- Glickman, M. (1995). A comprehensive guide to chess ratings. *American Chess Journal*, 3, 59–102. Retrieved from <http://www.glicko.net/research/acjpaper.pdf>
- Gomede, E., de Barros, R. M., & de Souza Mendes, L. (2020). Use of deep multi-target prediction to identify learning styles. *Applied Sciences (Switzerland)*, 10(5), 1756. <https://doi.org/10.3390/app10051756>
- Graf, S., Liu, T. C., & Kinshuk. (2010). Analysis of learners' navigational behaviour and their learning styles in an online course. *Journal of Computer Assisted Learning*, 26(2), 116–131. <https://doi.org/10.1111/j.1365-2729.2009.00336.x>
- Green, S., Hurst, L., Nangle, B., & Cunningham, P. (1997). Software agents: A review. *Technical Report*.
- Guskey, T. (2005). Formative Classroom Assessment and Benjamin S. Bloom: Theory, Research, and Implications. *Annual Meeting of the American Educational Research*

Association, (April), 1–11. <https://doi.org/April 2005>

Guskey, T. R., & Bailey, J. M. (2001). *Developing Grading and Reporting Systems for Student Learning*. Retrieved from http://books.google.com/books?id=O37oL0PL8wUC&printsec=frontcover&q=inauthor:Guskey&hl=&cd=11&source=gbs_api

Hagan M.T., H. B. D., & Beale, M. H. (1996). Neural Network design. In *Neural Network Design*.

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., & Gifford, J. A. (1978). Developments in Latent Trait Theory: Models, Technical Issues, and Applications. *Review of Educational Research*, 48(4), 467–510. <https://doi.org/10.3102/00346543048004467>

Hardwick, J. P., & Stout, Q. F. (1992). Bandit strategies for ethical sequential allocation. *Computing Science and Statistics*, 23, 421–424.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>

Herbert, R. (1952). Some Aspects of the Sequential Design of Experiments. *Bulletin of the American Mathematical Society*, 58(5), 527–535. <https://doi.org/10.1090/S0002-9904-1952-09620-8>

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A Systematic Review of Deep Learning Approaches to Educational Data Mining. *Complexity*, Vol. 2019. <https://doi.org/10.1155/2019/1306039>

Hohensinn, C., & Kubinger, K. D. (2011). Applying item response theory methods to examine the impact of different response formats. *Educational and Psychological Measurement*, 732–746. <https://doi.org/10.1177/0013164410390032>

Hou, J., Koppat, M. W., Quecedo, J. M. H., Stoyanova, N., Kopotев, M., & Yangarber, R. (2019). Modeling language learning using specialized elo ratings. *ACL 2019 - Innovative Use of NLP for Building Educational Applications, BEA 2019 - Proceedings of the 14th Workshop*, 494–506. <https://doi.org/10.18653/v1/w19-4451>

Howard-Jones, P. A. (2014). Neuroscience and education: Myths and messages. *Nature Reviews Neuroscience*, Vol. 15, pp. 817–824. <https://doi.org/10.1038/nrn3817>

Huang, C. J., Chen, H. X., & Chen, C. H. (2009a). Developing argumentation processing agents for computer-supported collaborative learning. *Expert Systems with Applications*, 36(2 PART 2), 2615–2624. <https://doi.org/10.1016/j.eswa.2008.01.036>

Huang, Y. M., Lin, Y. T., & Cheng, S. C. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers and Education*, 53–67. <https://doi.org/10.1016/j.compedu.2008.06.007>

Jia, J., & Le, H. (2020). The design and implementation of a computerized adaptive testing system for school mathematics based on item response theory. *Communications in*

Computer and Information Science, 1302, 100–111. https://doi.org/10.1007/978-981-33-4594-2_9

Jindal, R., & Borah, M. D. (2013). A Survey on Educational Data Mining and Research Trends. *International Journal of Database Management Systems*, 53. <https://doi.org/10.5121/ijdms.2013.5304>

Jones, L. V., & Thissen, D. (2006). 1 A History and Overview of Psychometrics. *Handbook of Statistics*, 26, 1–27. [https://doi.org/10.1016/S0169-7161\(06\)26001-2](https://doi.org/10.1016/S0169-7161(06)26001-2)

Juškevičienė, A. (2014). *Antrosios kartos saityno priemonės mokymuisi*. 146.

Kang, X., & Zhang, E. (2019). A universal defect detection approach for various types of fabrics based on the Elo-rating algorithm of the integral image. *Textile Research Journal*, Vol. 89, pp. 4766–4793. <https://doi.org/10.1177/0040517519840636>

Kausar, S., Huahu, X., Hussain, I., Wenhao, Z., & Zahid, M. (2018). Integration of Data Mining Clustering Approach in the Personalized E-Learning System. *IEEE Access*, 72724–72734. <https://doi.org/10.1109/ACCESS.2018.2882240>

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers and Education*, 106, 166–171. <https://doi.org/10.1016/j.compedu.2016.12.006>

Kolen, M. J. (1981). Comparison of Traditional and Item Response Theory Methods for Equating Tests. *Journal of Educational Measurement*, 18(1), 1–11. <https://doi.org/10.1111/j.1745-3984.1981.tb00838.x>

Kuleshov, V. (2000). Algorithms for the multi-armed bandit problem. *Journal of Machine Learning Research*.

Lai, T. L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1), 4–22. [https://doi.org/10.1016/0196-8858\(85\)90002-8](https://doi.org/10.1016/0196-8858(85)90002-8)

Lan, A. S., & Baraniuk, R. G. (2016). A contextual bandits framework for personalized learning action selection. *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016*, 424–429.

Lee, W. C., Kim, S. Y., Choi, J., & Kang, Y. (2020). IRT Approaches to Modeling Scores on Mixed-Format Tests. *Journal of Educational Measurement*, 230–254. <https://doi.org/10.1111/jedm.12248>

Li, X., Xu, H., Zhang, J., & Chang, H. H. (2021). Optimal Hierarchical Learning Path Design With Reinforcement Learning. *Applied Psychological Measurement*, 45(1), 54–70. <https://doi.org/10.1177/0146621620947171>

Lin, F. (2020). Adaptive quiz generation using Thompson sampling. *Third Workshop Eliciting Adaptive Sequences for Learning (WASL 2020) Cyberspace*. Retrieved from <https://humanlearn.io/>

Liu, M., Kitto, K., & Buckingham Shum, S. (2021). Combining factor analysis with

writing analytics for the formative assessment of written reflection. *Computers in Human Behavior*, 120, 106733. <https://doi.org/10.1016/j.chb.2021.106733>

Liu, Y., Mandel, T., Brunskill, E., & Popovi, Z. (2014). Trading Off Scientific Knowledge and User Learning with Multi-Armed Bandits. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, (Edm), 161–168.

Lord, F. M. (2012). Applications of item response theory to practical testing problems. In *Applications of Item Response Theory To Practical Testing Problems*. <https://doi.org/10.4324/9780203056615>

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.

Magnus, B. (2018). The Basics of Item Response Theory Using R. In *Measurement: Interdisciplinary Research and Perspectives* (Vol. 16). <https://doi.org/10.1080/15366367.2018.1462078>

Mangaroska, K., Vesin, B., & Giannakos, M. (2019). Elo-rating method: Towards adaptive assessment in e-learning. *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019*, 380–382. <https://doi.org/10.1109/ICALT.2019.00116>

Margienè, A., & Ramanauskaitė, S. (2021). Toward adaptability of e-evaluation: Transformation from tree-based to graph-based structure. *Applied Sciences (Switzerland)*, 11(9), 4082. <https://doi.org/10.3390/app11094082>

Marzano, R.J., Gaddy, B. B., & Dean, C. (2000). What Works In Classroom Instruction. *Aurora CO Midcontinent Research for Education and Learning*, (2723), 800–933. Retrieved from http://www.mcrel.org/PDF/Instruction/5992TG_What_Works.pdf

Marzano, Robert J, Pickering, D., & Pollock, J. E. (2001). Classroom instruction that works: Research-based strategies for increasing student achievement. In *Representations* (Vol. 80014). Retrieved from <http://www.amazon.com/Classroom-Instruction-That-Works-Research-Based/dp/0871205041>

Massa, L. J., & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences*, 16(4), 321–335. <https://doi.org/10.1016/j.lindif.2006.10.001>

McCabe, G. P., Sneath, P. H. A., & Sokal, R. R. (1975). Numerical Taxonomy: The Principles and Practice of Numerical Classification. *Journal of the American Statistical Association*, 70(352), 962. <https://doi.org/10.2307/2285473>

McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers and Education*, pp. 299–312. [https://doi.org/10.1016/S0360-1315\(02\)00032-5](https://doi.org/10.1016/S0360-1315(02)00032-5)

Mecca, G., Raunich, S., & Pappalardo, A. (2007). A new algorithm for clustering search results. *Data and Knowledge Engineering*, 62(3), 504–522. <https://doi.org/10.1016/j.datak.2006.10.006>

- Meiser, T., Plieninger, H., & Henninger, M. (2019). IRTree models with ordinal and multidimensional decision nodes for response styles and trait-based rating responses. *British Journal of Mathematical and Statistical Psychology*, 72(3), 501–516. <https://doi.org/10.1111/bmsp.12158>
- Milosevic, D., Brkovic, M., Debevc, M., & Krneta, R. (2007). Adaptive Learning by Using SCOs Metadata. *Interdisciplinary Journal of E-Skills and Lifelong Learning*, 3, 163–174. <https://doi.org/10.28945/391>
- Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Studies in Educational Evaluation*, Vol. 68, p. 100963. <https://doi.org/10.1016/j.stueduc.2020.100963>
- Mizumoto, A., Sasao, Y., & Webb, S. A. (2019). Developing and evaluating a computerized adaptive testing version of the Word Part Levels Test. *Language Testing*, 36(1), 101–123. <https://doi.org/10.1177/0265532217725776>
- Moss, C. M., & Brookhart, S. M. (2019). *Advancing formative assessment in every classroom: A guide for instructional leaders*. ASCD.
- Mousavinasab, E., Zarifsanaiy, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, Vol. 29, pp. 142–163. <https://doi.org/10.1080/10494820.2018.1558257>
- Mui, J., Lin, F., & Dewan, M. A. A. (2021). Multi-armed Bandit Algorithms for Adaptive Learning: A Survey. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12749 LNAI, 273–278. https://doi.org/10.1007/978-3-030-78270-2_49
- Newton, P. M. (2015). The learning styles myth is thriving in higher education. *Frontiers in Psychology*, 6(DEC), 1908. <https://doi.org/10.3389/fpsyg.2015.01908>
- Newton, P. M., & Salvi, A. (2020). How Common Is Belief in the Learning Styles Neuromyth, and Does It Matter? A Pragmatic Systematic Review. *Frontiers in Education*, Vol. 5, p. 270. <https://doi.org/10.3389/educ.2020.602451>
- Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Nižnan, J., Pelánek, R., & Rihák, J. (2015). Student Models for Prior Knowledge Estimation. *Proceeding of the 8th International Conference on Educational Data Mining, EDM15*.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., & Damasevicius, R. (2021). Artificial

Neural Networks for Educational Data Mining in Higher Education: A Systematic Literature Review. *Applied Artificial Intelligence*, 35(13), 983–1021. <https://doi.org/10.1080/08839514.2021.1922847>

Pankiewicz, M., & Bator, M. (2019). Elo Rating Algorithm for the Purpose of Measuring Task Difficulty in Online Learning Environments. *E-Mentor*, 82(5), 43–51. <https://doi.org/10.15219/em82.1444>

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles concepts and evidence. *Psychological Science in the Public Interest, Supplement*, 9(3), 105–119. <https://doi.org/10.1111/j.1539-6053.2009.01038.x>

Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis - A new alternative to knowledge tracing. *Frontiers in Artificial Intelligence and Applications*, 200(1), 531–538. <https://doi.org/10.3233/978-1-60750-028-5-531>

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers and Education*, 98, 169–179. <https://doi.org/10.1016/j.compedu.2016.03.017>

Pignatiello, G. A., Martin, R. J., & Hickman, R. L. (2020). Decision fatigue: A conceptual analysis. *Journal of Health Psychology*, Vol. 25, pp. 123–135. <https://doi.org/10.1177/1359105318763510>

Qi, X., Liu, C., & Schuckers, S. (2018). IoT edge device based key frame extraction for face in video recognition. *Proceedings - 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID 2018*, 641–644. <https://doi.org/10.1109/CCGRID.2018.00087>

Rafferty, A. N., Ying, H., & Williams, J. J. (2018). Bandit assignment for educational experiments: Benefits to students versus statistical power. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10948 LNAI, 286–290. https://doi.org/10.1007/978-3-319-93846-2_53

Rafferty, A. N., Ying, H., & Williams, J. J. (2019). Statistical Consequences of using Multi-armed Bandits to Conduct Adaptive Educational Experiments. In *Journal of Educational Data Mining* (Vol. 11). Retrieved from <https://jedm.educationaldatamining.org/index.php/JEDM/article/view/357%0Ahttps://jedm.educationaldatamining.org>

Rawson, K., Stahovich, T. F., & Mayer, R. E. (2017). Homework and achievement: Using smartpen technology to find the connection. *Journal of Educational Psychology*, 109(2), 208–219. <https://doi.org/10.1037/edu0000130>

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21(1), 25–36. <https://doi.org/10.1177/0146621697211002>

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2015). Matching learning style to instructional method: Effects on comprehension. *Journal of Educational Psychology*,

107(1), 64–78. <https://doi.org/10.1037/a0037478>

Romero, Cristobal, & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>

Romero, Cristbal, & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, Vol. 40, pp. 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>

Sadler, Royce, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144. Retrieved from [http://michiganassessmentconsortium.org/sites/default/files/Formative Assessment and Design of Instructional Systems.pdf](http://michiganassessmentconsortium.org/sites/default/files/Formative%20Assessment%20and%20Design%20of%20Instructional%20Systems.pdf)

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *International Journal of Phytoremediation*, 21(1), 77–84. <https://doi.org/10.1080/0969595980050104>

Schivinski, B., Brzozowska-Woś, M., Buchanan, E. M., Griffiths, M. D., & Pontes, H. M. (2018). Psychometric assessment of the Internet Gaming Disorder diagnostic criteria: An Item Response Theory study. *Addictive Behaviors Reports*, 8, 176–184. <https://doi.org/10.1016/j.abrep.2018.06.004>

Schlechter, P., Wilkinson, P. O., Knausenberger, J., Wanninger, K., Kamp, S., Morina, N., & Hellmann, J. H. (2021). Depressive and anxiety symptoms in refugees: Insights from classical test theory, item response theory and network analysis. *Clinical Psychology and Psychotherapy*, 28(1), 169–181. <https://doi.org/10.1002/cpp.2499>

Schunk, D. H. (2012). Learning theories: An educational perspective. In *Reading*.

Scriven, M. (1967). The methodology of evaluation. *Social Science Education Consortium*.

Segal, A., Ben David, Y., Williams, J. J., Gal, K., & Shalom, Y. (2018). Combining difficulty ranking with multi-armed bandits to sequence educational content. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10948 LNAI, 317–321. https://doi.org/10.1007/978-3-319-93846-2_59

Sensuse, D. I., Hasani, L. M., & Bagustari, B. (2020). Personalization Strategies Based on Felder-Silverman Learning Styles and Its Impact on Learning: A Literature Review. *2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020*, 293–298. <https://doi.org/10.1109/IC2IE50715.2020.9274670>

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>

Shatnawi, S., Gaber, M. M., & Cocca, M. (2014). Automatic content related feedback for MOOCs based on course domain ontology. *Lecture Notes in Computer Science (Including*

Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8669 LNCS, 27–35. https://doi.org/10.1007/978-3-319-10840-7_4

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>

Soloman, B. a, Carolina, N., & Felder, R. M. (2012). Index of Learning Styles Questionnaire. *Learning*, pp. 1–5. Retrieved from <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>

Tomasik, M. J., Berger, S., & Moser, U. (2018). On the development of a computer-based tool for formative student assessment: Epistemological, methodological, and practical issues. *Frontiers in Psychology*, 9(NOV). <https://doi.org/10.3389/fpsyg.2018.02245>

Tyner, M. (1977). Dynamic Parameter Estimation in Linear Systems. *Am Soc Eng Educ COED Trans*, 9(4), 45–50.

Urbonienė, J. (2014). *Adaptyviųjų programavimo mokymo priemonių projektavimas*.

Vaughan Williams, L., Liu, C., Dixon, L., & Gerrard, H. (2021). How well do Elo-based ratings predict professional tennis matches? *Journal of Quantitative Analysis in Sports*, 17(2), 91–105. <https://doi.org/10.1515/jqas-2019-0110>

Veenman, M. V. J., Prins, F. J., & Verheij, J. (2003). Learning styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology*, Vol. 73, pp. 357–372. <https://doi.org/10.1348/000709903322275885>

Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., & Duval, E. (2012). Context-aware recommender systems for learning: A survey and future challenges. *IEEE Transactions on Learning Technologies*, Vol. 5, pp. 318–335. <https://doi.org/10.1109/TLT.2012.11>

Vermorel, J., & Mohri, M. (2005). Multi-armed bandit algorithms and empirical evaluation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/11564096_42

Véron, M., Marin, O., & Monnet, S. (2014). Matchmaking in multi-player on-line games: Studying user traces to improve the user experience. *Proceedings of the 24th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video, NOSSDAV 2014*, 7–12. <https://doi.org/10.1145/2578260.2578265>

Wainer, H. (2001). Computer-adaptive testing: A primer. *Language Learning & Technology*.

Wang, W., Song, L., Wang, T., Gao, P., & Xiong, J. (2020). A Note on the Relationship of the Shannon Entropy Procedure and the Jensen–Shannon Divergence in Cognitive Diagnostic Computerized Adaptive Testing. *SAGE Open*, 10(1). <https://doi.org/10.1177/2158244019899046>

- Wauters, K., Desmet, P., & Noortgate, W. (2011). Monitoring learners' proficiency: Weight adaptation in the Elo rating system. *EDM 2011 - Proceedings of the 4th International Conference on Educational Data Mining*, 247–251.
- Wauters, Kelly, Desmet, P., & Van Den Noortgate, W. (2012). Item difficulty estimation: An auspicious collaboration between data and judgment. *Computers and Education*, 58(4), 1183–1193. <https://doi.org/10.1016/j.compedu.2011.11.020>
- Westby, C. (2019). The Myth of Learning Styles. *Word of Mouth*, Vol. 31, pp. 4–7. <https://doi.org/10.1177/1048395019879966a>
- Williams, J. J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K. Z., Lasecki, W. S., & Heffernan, N. (2016). AXIS: Generating explanations at scale with learnersourcing and machine learning. *L@S 2016 - Proceedings of the 3rd 2016 ACM Conference on Learning at Scale*, 379–388. <https://doi.org/10.1145/2876034.2876042>
- Wu, J., Huang, Z., Liu, Q., Lian, D., Wang, H., Chen, E., ... Wang, S. (2021). Federated Deep Knowledge Tracing. *WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 662–670. <https://doi.org/10.1145/3437963.3441747>
- Xu, D., & Wang, H. (2006). Intelligent agent supported personalization for virtual learning environments. *Decision Support Systems*, 42(2), 825–843. <https://doi.org/10.1016/j.dss.2005.05.033>
- Yang, F., Li, F. W. B., & Lau, R. W. H. (2010). An open model for learning path construction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6483 LNCS, 318–328. https://doi.org/10.1007/978-3-642-17407-0_33
- Yanson, R., & Johnson, R. D. (2016). An empirical examination of e-learning design: The role of trainee socialization and complexity in short term training. *Computers and Education*, 101, 43–54. <https://doi.org/10.1016/j.compedu.2016.05.010>
- Zhang, D., & Dong, Y. (2004). Semantic, hierarchical, online clustering of web search results. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3007, 69–78. https://doi.org/10.1007/978-3-540-24655-8_8
- Zhou, Y., Li, X., Cao, Y., Zhao, X., Ye, Q., & Lv, J. (2021). *LANA: Towards Personalized Deep Knowledge Tracing Through Distinguishable Interactive Sequences*. Retrieved from <http://arxiv.org/abs/2105.06266>
- Žilinskienė, I. (2013). *Adaptyvus mokomųjų modulių personalizavimo metodas*.
- Zunaira Fatima Syeda, Uzma Shahzadi, & Ghazanfar Ali. (2020). Rasch Calibration of Achievement Test: An Application of Item Response Theory. *Sjesr*, 3(2), 426–432. [https://doi.org/10.36902/sjesr-vol3-iss2-2020\(426-432\)](https://doi.org/10.36902/sjesr-vol3-iss2-2020(426-432))

List of Scientific Publications by the Author on the Topic of the Dissertation

Papers in the Reviewed Scientific Journals

Meleško J.; Ramanauskaitė S. Time saving students' formative assessment: Algorithm to balance number of tasks and result reliability // *Applied sciences*. Basel: MDPI. ISSN 2076-3417. vol. 11, iss. 13 (2021), p. 1–15. DOI: 10.3390/app11136048. [Citav. rod.: 2,679 (2021, InCites JCR SCIE)]

Meleško, J.; Novickij, V. Computer adaptive testing using Upper-Confidence Bound algorithm for formative assessment // *Applied sciences*. Basel: MDPI AG. eISSN 2076-3417. 2019, vol. 9, iss. 20, art. no. 4303, p. 1–12. DOI: 0.3390/app9204303. [Citav. rod.: 2,474 (2019, InCites JCR SCIE)]

Meleško, J.; Kurilov, J. Adaptive tutoring system with application of intelligent agents // *International Journal of Smart Education and Urban Society* (IJSEUS). Hershey: IGI Global. ISSN 2574-8254. eISSN 2574-8270. 2018, vol. 9, iss. 2, p. 1–11. DOI: 10.4018/IJSEUS.2018040101.

Kurilov, J.; Meleško, J.; Krikun, I. Intelligent multi-agent learning system applying educational data mining = Intelektinė daugiaagentė mokymosi sistema, naudojanti edukacinių duomenų tyrybą // *Informacijos mokslai*. Vilnius: Vilniaus universiteto leidykla. ISSN

1392-0561. eISSN 1392-0561. 2017, T. 79, p. 30–43. DOI: 10.15388/Im.2017.79.11381.

Meleško, J.; Kurilov, J.; Krikun, I. Daugiaagentė sistema matematikos ir informatikos mokymui = Multi-agent system for computer science and engineering education // *Lietuvos matematikos rinkinys*. Ser. B. Vilnius: Matematikos ir informatikos institutas. ISSN 0132-2818. eISSN 2335-898X. 2017, T. 58, p. 57–61. DOI: 10.15388/LMR.B.2017.10

Papers in Other Editions

Meleško, J.; Kurilov, J. On personalized multi-agent learning system: artificial neural network agent // *ICERI 2017: 10th annual International Conference of Education, Research and Innovation*, November 16–18, 2017, Seville, Spain: conference proceedings. Valencia: IATED, 2017. ISBN 9788469769577. p. 3883–3891. (ICERI Proceedings).

Meleško, J. On personalized multi-agent learning system: learning outcomes assessment agent // *ICERI 2017: 10th annual international conference of education, research and innovation*, 16–18 November, 2017, Seville, Spain: conference proceedings. Valencia: IATED Academy, 2017. ISBN 9788469769577. p. 3892–3899. (ICERI Proceedings).

Kurilov, J.; Krikun, I.; Meleško, J. On using learning analytics to personalize learning // *ICERI 2016: 9th annual international conference of education, research and innovation*, Seville, 14th–16th of November, 2016: conference proceedings. Seville: IATED Academy, 2016. ISBN 9788461758951. p. 6987–6996. (ICERI proceedings, ISSN 2340-1095). [IATED digital library; Conference Proceedings citation Index - Social Science & Humanities (Web of Science)]

Meleško, J.; Kurilov, J. Personalized intelligent multi-agent learning system for engineering courses // *Advances in information, electronic and electrical engineering* (AIEEE): 2016 IEEE 4th workshop. New York: IEEE, 2016. eISBN 9781509044733. p. 1–6. DOI: 10.1109/AIEEE.2016.7821821. [IEEE Xplore; Conference Proceedings Citation Index - Science (Web of Science); INSPEC]

Kurilov, J.; Kurilova, J.; Kurilova, I.; Meleško, J. Personalized learning system based on students' learning styles and application of intelligent technologies // *ICERI 2016 : 9th annual international conference of education, research and innovation*, Seville, 14th-16th of November, 2016 : conference proceedings. Seville: IATED Academy, 2016. ISBN 9788461758951. p. 6976–6986. (ICERI Proceedings, ISSN 2340-1095).

Meleško, J.; Kurilov, J. Semantic technologies in e-learning: learning analytics and artificial neural networks in personalized learning systems // *WIMS 2018. 8th international conference on Web Intelligence, Mining and Semantics*, Novi Sad, Serbia, June 25–27, 2018. New York: Association for Computing Machinery, 2018, art. no. 34. ISBN 9781450354899. p. 1–7. (ACM International Conference Proceeding Series). DOI: 10.1145/3227609.3227669.

Summary in Lithuanian

Ivadas

Problemos formulavimas

Pastaruosius du dešimtmečius aktyviai plėtojamas pažangiųjų programinių agentų ir daugiaagenčių sistemų taikymas švietimo sistemoje. Intelektualieji agentai gali pritaikyti mokymosi medžiagą skirtingiems mokinių poreikiams. Pritaikomosios sistemos naudojimas švietimo srityje, suteikia daugialypę galimybę tobulinti žinių įgijimą ir gerinti formuojamojo vertinimo veiksmingumą bei tikslumą.

Mokslinių tyrimų, analizuojančių personalizuotas el. mokymosi sistemas, pagrįstas mokymosi stiliais, yra, tačiau jie dažnai orientuoti į konkretų vieną uždavinį, o ne į daugiaagenčius sprendimus. Mokinio ir el. mokymosi aplinkos kompleksiskumas reikalauja sudėtingesnių kombinuotų adaptyviojo mokymosi sprendimų. Siekiant sukurti optimalų (tinkamumo konkrečiam mokinio profiliui požiūriu) mokymosi objektą, el. mokymosi sistemai realizuoti reikia integruoti kelis skirtingus sprendimus. Yra metodų, skirtų mokymosi stiliams ir mokyimo medžiagos tinkamumui nustatyti, kuriuos būtų galima integruoti į modernią mokymosi sistemą, tačiau literatūroje trūksta dėmesio šiuolaikiškiems formuojamojo vertinimo metodams.

Formuojamasis, arba ugdomasis, vertinimas skirtas sukurti mokinio grįžtamąjį ryšį, kad juo būtų galima patobulinti trūkstamas mokinio kompetencijas, užtikrinti sklandesnį jų įgijimo procesą. Klausimų parinkimo problema ugdomajam testui nėra paprasta užduotis ir lemia procedūros tikslumą bei ilgį. Norint tiksliai nustatyti, kokios yra

dabartinės mokinių žinios ir jų spragos, kokios rekomenduojamos žinių tobulinimo kryptys, galima pasitelkti formuojamojo vertinimo ir išmaniųjų mokymo sistemų teikiamas galimybes.

Darbo aktualumas

Standartizuotas mokymas mokyklose ir universitetuose mažai atsižvelgia į individualius mokinių poreikius. Be to, grįžtamasis ryšys pažymio forma, kaip tai daroma tradiciniuose testuose, mokiniui yra mažai naudingas, nes neatitinka rekomenduojamų grįžtamojo ryšio rekomendacijų. Kompiuterizuotos mokymo aplinkos su integruotomis daugiaagentėmis sistemomis gali pagerinti mokymosi procesą. Metodai, pagrįsti formuojamojo vertinimo paradigma, yra ypač daug žadantys. Disertacijoje pristatomi nauji formuojamojo vertinimo metodai ir sukurtas algoritmas yra tiesiogiai naudingi edukaciniame procese. Metodai trumpina formuojamųjų testų atlikimo laiką ir supaprastina jų sudarymą.

Tyrimo objektas

Tyrimo objektas – adaptyvieji formuojamojo vertinimo algoritmai ir jų integravimas į daugiaagentes mokymosi sistemas.

Darbo tikslas

Disertacijos tikslas – sukurti trumpinančius testavimo laiką ir nereikalaujančius kalibravimo ugdomojo vertinimo algoritmus ir taip padidinti el. mokymosi sistemų adaptyvumą.

Darbo uždaviniai

Siekiant išspręsti problemą ir pasiekti disertacijos tikslą, iškelti šie uždaviniai:

1. Išanalizuoti mokymosi stilių ir ugdomojo vertinimo pritaikymą ir efektyvumą intelektualiose el. mokymosi sistemose.
2. Pasiūlyti daugiaagentės mokymosi sistemos koncepcinį modelį, kuris padės integruoti skirtingus el. mokymosi personalizavimo aspektus.
3. Sukurti adaptyvius, į studentą orientuotus formuojamojo vertinimo metodus, skirtus el. mokymosi sistemoms.
4. Palyginti sukurtus formuojamojo vertinimo algoritmus su daugiadimense testo uždavinių teorija (angl. MIRT) pagrįstomis metodikomis, studentų darbo krūvio ir žinių spragų nustatymo aspektais.

Tyrimų metodika

Tyrimė taikyti skaitmeniniai (simuliacijos ir kintamųjų optimizavimas) ir eksperimentiniai metodai. Buvo pasiūlytos daugiaagentės mokymosi sistemos, paremtos duomenų gavybos, semantinio klasterizavimo ir neuroninių tinklų principais. Formuojamojo vertinimo algoritmai sukurti taikant viršutinės patikimumo ribos (UCB) ir Elo vertinimo metodus bei įvertinti MIRT ir didžiausio tikėtinumo vertinimo (MLE) modelių kontekste.

Darbo mokslinis naujumas

Mokslinio naujumo aspektai informatikos inžinerijos srityje yra šie:

1. Formuojamasis vertinimas buvo sumodeliuotas kaip paieškos ir eksploataavimo problema, UCB algoritmas pritaikytas studentams vertinti klasikinės testų teorijos kontekste. Gauti teigiami rezultatai rodo stiprų testo trukmės sumažėjimą.
2. Sukurtas naujas Elo ir UCB įkvėptas algoritmas, pritaikytas ugdomajam studentų vertinimui latentinių bruožų teorijos kontekste. Šis metodas sujungia IRT ir tradicinių testavimo metodų pranašumus: jam nereikia kalibruoti klausimų ir jis žymiai sutrumpina testo laiką.

Darbo rezultatų praktinė vertė

Siūlomos daugiaagentės sistemos ir formuojamo vertinimo algoritmai gali būti naudojami kuriant naujas išmaniojo mokymosi platformas, kurios atitinka nustatytas strategines kompetencijomis grindžiamo mokymosi ir vertinimo vizijas. Sukurtų metodų derinys su šiuo metu plačiai paplitusiais mobiliaisiais įrenginiais ir nuotoliniu mokymu gali padėti laiku suteikti mokiniui personalizuotą ir automatizuotą grįžtamąjį ryšį. Mobilieji įrenginiai įgalina pedagogus naudotis skaitmeniniais mokymo tobulinimo metodais, pasiūlytais disertacijoje. Algoritmų taikymas išspręs tokias problemas, kaip testų kūrimo sudėtingumas ir ilgas testavimo laikas. Tai paspartins mokymąsi per formuojamųjų vertinimų teikiamą personalizuotą grįžtamąjį ryšį.

Ginamieji teiginiai

Išanalizavus tyrimo metu nustatytą informaciją ir įvertinus darbo rezultatus, buvo suformuluoti toliau pateikti ginamieji teiginiai:

1. Siūloma išmanioji el. mokymosi sistema integruoja naujai pasiūlytus, į studentą orientuotus formuojamojo vertinimo algoritmus.
2. Siūlomi formuojamojo vertinimo algoritmai atitinka Hattie trijų žingsnių veiksmingo grįžtamojo ryšio modelį.
3. Vertinimo, kaip UCB problemos, modeliavimas leidžia greičiau nustatyti silpniausias mokinio žinių spragas, lyginant su tradiciniais formuojamaisiais testais.
4. Elo grindžiamas formuojamasis vertinimas sutrumpina testo trukmę ir leidžia pasiekti panašų tikslumą kaip ir IRT metodai, bet kartu nereikalauja išankstinio užduočių kalibravimo.

Darbo rezultatų aprobavimas

Disertacijos tema paskelbta vienuolika straipsnių ir pranešimų. Tyrimų rezultatai paskelbti dviejuose straipsniuose žurnale su citavimo indeksu, referuojamame *Clarivate AnalyticsWoS* (Meleško, 2019, 2021), trys straipsniai – kituose recenzuojamuose mokslo žurnaluose (Meleško 2017a, 2018a; Kurilov 2017). Rezultatai buvo paskelbti šešiuose konferencijų straipsniuose (Kurilov, 2016a, 2016b; Meleško, 2016, 2017b, 2017c, 2018b), iš

kurių penki įtraukti į *Clarivate Analytics Web of Science* duomenų bazę. Disertacijos rezultatai aprobuoti keturiose tarptautinėse konferencijose, vykusiose Ispanijoje (ICERI 2016 ir ICERI 2017), Serbijoje (WIMS 2018) ir Lietuvoje (AIEEE 2016).

Disertacijos struktūra

Disertaciją sudaro įvadas, trys skyriai, bendrosios išvados, literatūros sąrašas, autoriaus mokslinių publikacijų sąrašas ir santrauka lietuvių kalba.

Disertaciją sudaro 128 puslapiai, 27 paveikslų, 12 lentelės, 22 lygtys ir 182 literatūros šaltiniai.

1. Mokymo adaptyviųjų sistemų literatūros analizė

Išmanojo mokymosi tyrimų vis daugėja. Tačiau yra keletas šios bendros tendencijos išimčių, pavyzdžiui, tyrimai RFID tematika. Specialiųjų poreikių ir mokinių bendradarbiavimo mokymo temų populiarumas nebeauga ir nebėra toks aktualus moksliniu požiūriu. Greičiausiai augančios mokslinių tyrimų sritys yra el. mokymosi sistemų inžinerija ir afektinė bei biometrinė kompiuterija (Chen et al., 2021).

Išmanojo mokymosi diegimas pramonėje vyksta informacinių technologijų sektoriuje. Daugiausia mokslinių tyrimų, susijusių su pramone, atliekama vartotojų įtraukimo ir mokymosi skatinimo srityse. Tyrimai taip pat rodo ribotą susidomėjimą didelės apimties duomenų rinkimu švietimo tikslais ir socialiniu mokymusi bendradarbiaujant tematikomis (Giannakos et al., 2021).

Formuojamasis vertinimas yra labai svarbus mokymosi proceso komponentas, nes juo vertinamos mokinio žinios, nustatomas kelias mokymosi tikslams pasiekti ir stebima pažanga (Black & Wiliam, 1998; Dunn & Mulvenon, 2009). Šis procesas pagreitina mokymąsi ir daro jį efektyvesnį (Guskey, 2005). Priešingai, apibendrinamasis vertinimas dažnai turi baudžiamąjį atspalvį, nes mokiniai už klaidas baudžiami žemesniu pažymiu (Scriven, 1967). Užuoat vertinus mokinio žinių lygį, formuojamojo vertinimo tikslas – teikti konstruktyvų grįžtamąjį ryšį, siekiant paspartinti mokymąsi (Moss & Brookhart, 2019; Sadler, 1998), suformuluoti mokymosi tikslus ir stebėti pažangą (Tomasik et al., 2018). Pasak Gareiso, tai galima pasiekti atsižvelgiant į tris veiksnius, kurie formuojamąjį vertinimą daro veiksmingą (Assessment Reform Group, 1999). Pirma, formuojamasis vertinimas yra veiksmingiausias, kai yra naudojamas siekiant pritaikyti mokinio mokymosi kelią, atsižvelgiant į jo esamą supratimo lygį. Antra, formuojamojo vertinimo rezultatai neturėtų daryti įtakos mokinio vertinimui, kaip tai daroma apibendrinamojo vertinimo atveju. Tai padeda mokiniams prisiimti atsakomybę už savo mokymosi procesą ir sutelkti dėmesį į naujų žinių įgijimą. Trečia, formuojamasis vertinimas turėtų būti naudojamas grįžtamajam ryšiui suteikti ir nukreipti mokinius mokymosi tikslų link (Sadley & Royce, 1989).

Pasak Marzano ir kt., grįžtamasis ryšys yra esminis pedagoginis metodas, padedantis gerinti mokinių pasiekimus (Marzano et al., 2001). Tyrimai rodo, kad grįžtamasis ryšys gali pagerinti mokymosi rezultatus su vidutiniu efekto dydžiu 0,8, kai studentai yra testuojami iškart po konsultacijos (Azevedo & Bernard, 1995; Shute, 2008). Tačiau dėl šių rezultatų vis dar diskutuojama (Bennett, 2011). Vienas iš veiksnių, lemiančių grįžtamojo ryšio kokybę, yra tai, kaip jis tenkina esminį pedagoginį poreikį – mažinti atotrūkį tarp

esamų ir pageidaujamų mokinio gebėjimų (Hattie & Timperley, 2007). Kitas veiksnys – vėlavimas tarp mokinio žinių įvertinimo ir grįžtamojo ryšio, kuris, vertinant testus rankiniu būdu, gali trukti kelias dienas ar net savaites. Šis vėlavimas gali neigiamai paveikti teigiamą grįžtamojo ryšio poveikį (Gareis, 2007). Galiausiai grįžtamasis ryšys pažymio forma nesuteikia galimybės siekti didesnio profesinio meistriškumo (Guskey & Bailey, 2001), taigi neatlieka vienos iš pagrindinių formuojamojo vertinimo ir grįžtamojo ryšio funkcijų (Nicol & MacFarlane-Dick, 2006).

Vertinant formuojamojo vertinimo praktinę vertę, reikia atsižvelgti į laiką, sugaištą jam atlikti. Mokinių žinios nuolat kinta, todėl mokiniai turi būti pakartotinai vertinami (Choi & McClenen, 2020). Tai atima daug laiko. Kartu, jei mokiny bus priverstas demonstruoti jau įsisavintas žinias, gali dingti motyvacija atlikti formuojamąjį vertinimą. Todėl būtina užtikrinti pusiausvyrą tarp reikiamo užduočių skaičiaus ir vertinimo tikslumo (silpniausios mokinio žinių sritys ir geriausias mokymosi kelias).

Klasikinė testų teorija (KTT) jau seniai naudojama kuriant ir naudojant testus (Novick, 1966). Ji remiasi prielaida, kad testai matuoja tikrojo rezultato ir klaidingo rezultato santykį. Latentinų bruožų teorija (LTT), dar žinoma kaip atsako į testą teorija (angl. *Item Response Theory*, IRT), taiko labiau niuansuotą požiūrį, pavyzdžiui, statistinius modelius, kuriuose atsižvelgiama į skirtingą testų sunkumą ir daugiamatės latentinių bruožų erdvės egzistavimą (Hambleton, Swaminathan, Cook, Eignor & Gifford, 1978; Meiser, Plieninger & Henninger, 2019). Ji grindžiama idėja, kad teisingo atsakymo į elementą tikimybė yra mokinio ir elemento asmeninių latentinių savybių funkcija (Kolen, 1981; Reckase, 1997). Šis metodas dažnai taikomas atliekant kompiuterizuotą adaptyvųjį testavimą (Jia & Le, 2020; Lord, 2012). Atliekant adaptyvųjį testavimą mokiniui pateikiami elementai parenkami atsižvelgiant į anksčiau matytus elementus ir mokinio atsakymus. Šis metodas gali sutrumpinti testavimo laiką ir padidinti tikslumą, pavyzdžiui, maksimaliai padidinant informacijos gavimą iš kiekvieno elemento (McDonald, 2002; Wang, Song, Wang, Gao & Xiong, 2020). Tačiau IRT grindžiamiems testams reikia kalibruoti elementų rinkinį, kad būtų nustatyti elementų parametrai, naudojami apskaičiuojant teisingo atsakymo tikimybę (Edelen & Reeve, 2007; Syeda, Shahzadi & Ali, 2020). Tai padidina testo kūrimo ir pakartotinio taikymo sudėtingumą. Praktiškai taikant šis sudėtingumas gali trukdyti pedagogams naudoti IRT grindžiamus formuojamojo vertinimo testus.

Elo reitingais pagrįstos sistemos gali atlikti panašų vaidmenį kaip ir IRT pagrįsti sprendimai. Elo reitingavimas iš pradžių buvo sukurtas panašių įgūdžių šachmatininkams parinkti (Elo, 1987). Tie patys principai gali būti naudojami vertinant elementų sudėtingumą, mokinių įgūdžius ir juos derinant ugdymo aplinkoje (Abdi, Khosravi & Sadiq, 2021; Mangarosa et al., 2019; Pelánek, 2016). Ir IRT, ir Elo reitingavimo sistemos paprastai naudoja logistinę funkciją, kad nustatytų teisingo atsakymo į klausimą tikimybę. Tačiau Elo reitingu pagrįstos sistemos dinaminio būdu nustato elemento parametrus, kiekvieno atsakymo elemento sunkumo nustatymas iš karto koreguojamas testo metu. Todėl, prieš pradedant testą, nereikia atlikti atskiros kalibravimo etapo. Kaip ir IRT atveju, buvo pasiūlytos daugiamatės, daugiau nei vieną latentinį požymį matuojančios Elo reitingų sistemos (Abdi, Khosravi, Sadiq & Gasevic, 2019; Doeblner et al., 2015).

Yra ir kitų metodų, kaip modeliuoti studentus ir žinių įgijimą (Desmarais & Baker, 2012), pavyzdžiui, veiklos veiksnių analizė (Liu, Kitto & Buckingham Shum, 2021;

Pavlik et al., 2009), žinių sekimas (Corbett & Anderson, 1994; Wu et al., 2021; Zhou et al., 2021) ir kompiuteriniu mokymu pagrįsti sprendimai (Lin, 2020; Y. Liu et al., 2014).

Pirmojo skyriaus išvados:

1. Individualizuotas mokymas gali turėti teigiamą poveikį studentų mokymosi efektyvumui, kas meta tyrimų šioje srityje daugėja.
2. Mokymosi stiliais pagrįstas personalizavimas yra plačiai taikomas, tačiau jam trūksta empirinio pagrindimo.
3. Mokiniai patys blogai identifikuoja efektyviausius mokymosi stilius, todėl jiems turėtų padėti įrankiai.
4. Pagrindinė mokinių kompetencijų diagnostavimo priemonė turėtų būti formuojamasis vertinimas, kuris turi daug privalumų, lyginant su apibendrinamuoju vertinimu.
5. Efektyvaus formuojamojo vertinimo raktas yra grįžtamasis ryšys, kurį užtikrina trijų žingsnių procesas: Kur aš einu? Kaip man sekasi?, Kur toliau?

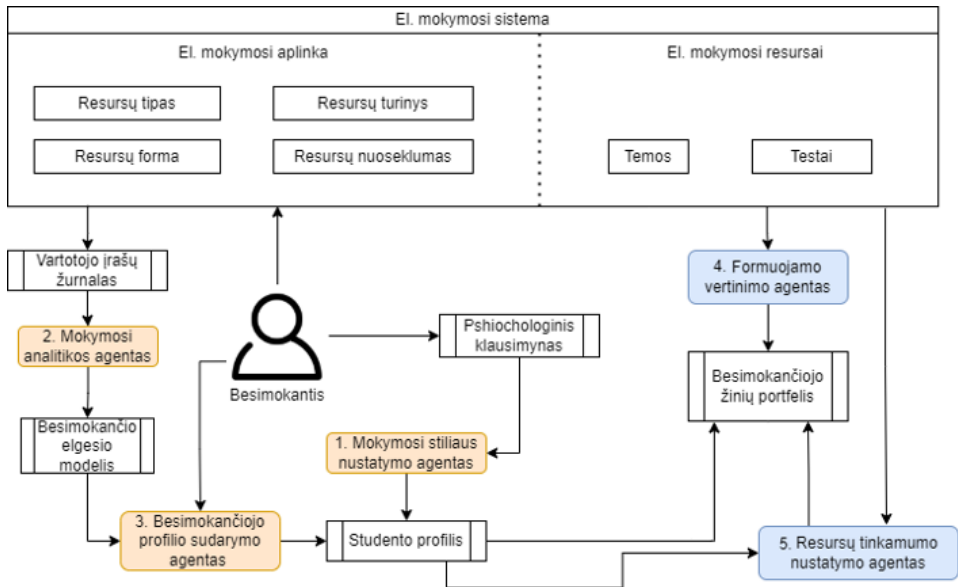
2. Personalizuotos daugiaagentės el. mokymosi sistemos projektavimas

Sistema, formuodama rekomendacijas, turėtų atsižvelgti į žinias apie mokinius ir pedagoginę medžiagą. Todėl išteklių tipas, forma, seka ir turinys turėtų būti pritaikyti prie asmeninių mokinių poreikių. Norint įgyvendinti personalizavimą įvairiuose pedagoginiuose elementuose, geriausia naudoti kelis agentus, kurių kiekvienas turi aiškiai apibrėžtą funkciją. Toks modulinis požiūris gali būti vertingas praktiškai įgyvendinant ir taikant siūlomą sistemą, taip pat pradiniuose planavimo ir iteracinio tobulinimo etapuose. Pagrindinė siūlomos sistemos inovacija tai, kad ji naudoja formuojamąjį vertinimą mokymuisi personalizuoti. Tokių sistemų šiuo metu nėra, nes trūksta tinkamų ugdomojo vertinimo algoritmų.

Formuojamojo vertinimo agentas (4) yra pagrindinis, kai laikomasi gerosios grįžtamojo ryšio praktikos. Tai pirmasis žingsnis personalizavimo link, nes jis gali užtikrinti, kad kiekvienas mokinytis tiesiogiai matytų savo silpnąsias vietas ir kelią, kaip geriau įsisavinti dalyką. Žinoma, laikas ir pastangos, skirtos formuojamojo vertinimo testui sudaryti ir atlikti, gali būti panaudotos tiesiog mokantis. Formuojamojo vertinimo agentas gali sutaupyti laiko mokytojui, nes savarankiškai ir individualizuotai vertina studentų žinias. Disertacijos 3 skyriuje pasiūlyti algoritmai leidžia sutaupyti mokinių laiką sutrumpinant testo trukmę. Agentas ir mokymosi sistema pavaizduoti S2.1 pav.

Keletas papildomų agentų gali pagerinti sistemos personalizavimo laipsnį ir papildyti formuojamąjį vertinimą. Mokymosi stiliaus nustatymo agentas su specialiu psichologiniu klausimynu turėtų būti taikomas siekiant gauti tikimybinis mokinių mokymosi stilius. Mokymosi stilių identifikavimo programinis agentas (1) turėtų būti sukurtas tam, kad gautų šias mokinių mokymosi stilių reikšmes. Mokinio pirmenybė pagal aktyvaus ir refleksyvaus mokymosi ašį rodo, kokio tipo ištekliams turėtų būti teikiama pirmenybė: užduotims, reikalaujančioms aktyvaus dalyvavimo ar pasyvaus susivokimo.

Psichologinių klausimynų rezultatai ne visada yra tikslūs ir juos gali būti sunku dinamiškai atnaujinti. Galimos to priežastys – ilga klausimyno trukmė ir netikslūs atsakymai. Duomenys apie realių mokinių elgesį mokymosi aplinkoje, gauti remiantis mokymosi analitika, turėtų būti naudojami mokinių profiliams koreguoti pagal gautus duomenis. Taigi reikėtų sukurti mokymosi analitikos programinį agentą (2), kuris diagnozuotų mokinių profilius, stebėdamas jų realų elgesį mokymosi aplinkoje.



S2.1 pav. Daugiaagentės el. mokymosi sistemos koncepcinis modelis

Mokymosi stilių identifikavimo ir mokymosi analizės agentų pateikti duomenys turi būti suderinti naudojant mokinio profilio kontrolės agentą (3). Šis agentas sukurs ir atnaujins mokinių, remdamasis jų pateiktais duomenimis, taip pat duomenimis, surinktais iš jų veiklos mokymosi aplinkoje. Šiuos duomenis taip pat galima papildyti papildomais parametrais ir savybėmis, pavyzdžiui, žiniomis, interesais, tikslais, kognityviniais bruožais, mokymosi elgsenos tipu ir kt. Kad personalizavimo sistema veiktų, jai nebūtina žinoti visų mokinio bruožų, tačiau tam tikromis aplinkybėmis pedagogai gali norėti remtis viena ar kita pedagogine strategija, kad papildytų veiksmingą grįžtamojo ryšio modelį. O antrinių agentų funkcija – sudaryti jiems tokią galimybę.

Toliau reikėtų sukurti pedagoginio tinkamumo programinį agentą (5), kuris įgyvendintų rekomendavimo sistemą. Šis pedagoginio tinkamumo programinis agentas turėtų susieti optimalius mokymosi komponentus su konkrečiais mokiniais pagal tikimybinis tinkamumo indeksus. Visi mokymosi komponentai rekomendacinėje sistemoje turėtų būti susieti su konkrečiais mokiniais pagal jų tikimybinio tinkamumo indeksus. Kuo aukštesnis tikimybinio tinkamumo indeksas, tuo geriau mokymosi komponentas atitinka konkretaus

mokinio poreikius. Atsižvelgiant į mokinių žinių rinkinį (vertinamą formuojamojo vertinimo agento (4)) ir medžiagos tinkamumą, galima sukurti optimalų mokymosi scenarijų (t. y. aukščiausios kokybės mokymosi scenarijų) konkrečiam mokiniui. Mokinių kompetencijų, supratimo, įgūdžių ir kt. lygis tiesiogiai priklauso nuo optimalių mokymosi scenarijų taikymo realioje pedagoginėje praktikoje lygmenis. Greitai nustačius trūkstamas studentų žinias, naudojant formuojamojo vertinimo agentą (4) ir pateikus jiems tinkamiausią medžiagą, padidėtų mokymosi efektyvumas.

Taigi mokymosi programinė įranga ir visas mokymosi procesas turėtų būti individualizuotas atsižvelgiant į pagrindines mokinių savybes ir poreikius. Mokiniai turi skirtingus poreikius ir savybes, t. y. ankstesnes žinias, intelekto lygį, interesus, tikslus, kognityvinius bruožus (darbinės atminties gebėjimą, indukcinio mąstymo gebėjimą ir asociatyvius mokymosi įgūdžius), mokymosi elgesio tipą (pagal savireguliacijos lygį) ir mokymosi stilius. Mokymosi personalizavimas – tai asmeninių mokymosi paketų (scenarijų, modulių), pagrįstų rekomendavimo sistema, tinkančių konkrečioms mokiniams pagal jų asmeninius poreikius, kūrimas ir įgyvendinimas.

El. mokymosi sistemos modeliui sudaryti buvo atlikta keletas apklausų, eksperimentų su skirtingų universitetų studentais. Vienos apklausos metu buvo apklausti 104 Vilniaus Gedimino technikos universiteto bakalauro studijų studentai. Apklausos metu vertinta, kokio tipo, aplinkos testai jiems yra priimtinausi. Nustatyta, kad studentai iš esmės neturi stiprios nuomonės apie testo tipą. Todėl, atsižvelgiant į daugybę kompiuterinio testavimo privalumų, tokių kaip testo trukmė, piniginių išlaidų ir mokytojo darbo krūvio sumažėjimas, priimtinausias sprendimas – pereiti prie išimtinai kompiuterinio formuojamojo testavimo formos.

Su 22 Vilniaus universiteto studentais atlikus mokymosi stiliaus nustatymo eksperimentą nustatyta, kad:

1. 50 % studentų turi vidutinę priklausomybę atitinkamam mokymosi stiliui, o 45 % mokinių turėjo stipriai išreikštą konkretų mokymosi stilių.
2. Multimodalams (studentams, pirmenybę teikiantiems daugiau nei vienam mokymosi stiliui) priklausė 14 (63,6 %) studentų.
3. Tik 1 studentas neturi didelių preferencijų nė vienam iš mokymosi stilių.
4. Dominuoja kairėje skalės pusėje esantys mokymosi stiliai: 59,1 % studentų buvo įvertinti kaip aktyvūs, 86,4 % – vizualiniai, 72,7 % – sensoriniai ir 54,5 % – nuoseklūs.
5. Mokymosi stiliaus pirmenybės tarp lyčių reikšmingai nesiskyrė.

Mokiniai, turintys stiprią pirmenybę mokymosi stiliui, turi daugiau sunkumų mokantis pagal tradicinę mokymo programą nei mokiniai, turintys silpną pirmenybę mokymosi stiliui. Taigi mokiniams, teikiantiems stiprią pirmenybę mokymosi stiliui, gali būti ypač naudinga taikyti išmaniąsias technologijas ir adaptuoti mokymosi paketus (scenarijus) prie jų pirmenybę teikiančio mokymosi stiliaus. Šie rezultatai rodo, kad mokymosi stiliaus negalima ignoruoti, kuriant el. mokymosi sistemą. Tačiau mokymosi stilius nėra susijęs su konkrečiu mokinio turimu žinių bagažu nustatymu, todėl, siekiant visapusiškai adaptyvios, personalizuotos sistemos, būtina įtraukti programinius agentus, kurie atliktų ne tik resursų atranką pagal mokymosi stilių, bet ir vykdytų mokinio žinių lygio, nepakankamai išmanomų temų sąrašo nustatymo veiksmus.

Antrojo skyriaus išvados:

1. Siūlomoje išmaniojoje el. mokymosi sistemoje naudojami plačiai paplitę Felderio ir Silvermano mokymosi stiliai, o jų trūkumams eliminuoti taikoma ir mokymosi analitika.
2. Apklausti studentai neišreiškė stiprių pageidavimų dėl kompiuterizuoto mokymosi ir mokymosi su savo mobiliuoju įrenginiu, bet yra šiek tiek labiau linkę naudoti nešiojamuosius ar planšetinius kompiuterius.
3. Apklausoje dalyvavę studentai išreiškė aiškią pirmenybę mokymosi stiliams pagal Felderio ir Silvermano mokymosi stilių modelį, kur net 45 % studentų išreiškė stiprią pirmenybę tam tikram mokymosi stiliui.
4. Pasiūlytas el. mokymosi sistemos modelis orientuotas į efektyvų, bet retai taikomą formuojantį vertinimą. Tačiau tokiai vizijai būtini nauji formuojamojo testavimo algoritmai.

3. Formuojamuoju vertinimu pagrįstas mokymosi turinys personalizuotai aplinkai

Testuojant mokinius reikia tiksliai susieti užduotis ir klausimus su dalyko žiniomis, kuriuos testai siekia patikrinti. Formuojamas testas gali nustatyti silpnai žinomas mokiniams temas ir konstruktyviai informuoti juos apie gebėjimus, kuriuos dar reikia lavinti. Kai mokinys padaro klaidą, ji iš karto siejama su žinių trūkumų. Tokiu būdu mokinys gauna asmenines mokymosi gaires, o ne įvertinimą pažymiu. Testus į mokinių mokymosi scenarijų įtraukia mokymosi scenarijaus sudarymo agentas. Tai vykdoma kai mokymosi rezultatų vertinimo agentas nustato, kad mokinys baigė visus mokymosi objektus, kurie jam buvo pateikti. Klausimų atranka ir pateikimu siekiama išmokyti mokinius konkrečių kompetencijų, kurias testas turi įvertinti.

Disertacijoje siūlomo mokymosi rezultatų vertinimo agento tikslas – suteikti personalizuotai intelektinei mokymosi sistemai šiuos anksčiau trūkstamus privalumus:

1. Motyvuoti studentus demonstruojant jiems jų žinių augimą.
2. Padidinti mokymosi efektyvumą, identifikuojant silpnai mokino žinomas temas ir jam pranešant apie tai.
3. Netaikyti mokinių baudimo ir lyginimo pažymiais.
4. Didinti mokymosi kokybę reikalaujant, kad visi mokiniai įgytų visas būtinas kompetencijas.
5. Taikyti personalizuotą žinių įvertinimą, taupantį mokinių laiką ir jėgas.

Kito klausimo parinkimo problema, atliekant formuojamąjį įvertinimą, gali būti modeliuojama kaip MAB problema (Melesko & Novickij, 2019). MAB algoritmai formuojamojo vertinimo užduotims parinkti dar nebuvo taikyti. Tyrimai patvirtina, kad pagrindinis formuojamojo vertinimo elementas yra grįžtamasis ryšys, suvokimas, kad esamos mokinių žinios neatitinka jų tikslų ir kur reikia eiti toliau, kad šie trūkumai būtų pašalinti. Siūlomas vertinimo metodas tenkina šį poreikį, greitai nustatydamas trūkstamas žinių sritis ir

nuodugniai jas išnagrinėdamas, kad padėtų toliau mokytis. Į šį procesą galima pažvelgti per J. Hattie trijų klausimų grįžtamojo ryšio modelio prizmę (Hattie & Timperley, 2007). Norėdamas panaudoti algoritmą, mokytojas pirmiausia turi suformuoti temas arba kompetencijas (Kur aš einu?). Algoritmas greitai nustato trūkstamas žinių sritis (Kaip aš einu?) ir išsamiai išnagrinėja temą, padėdamas toliau mokytis (Kur toliau?). Šis metodas kartu su dabar plačiai paplitusiais mobiliaisiais įrenginiais gali sumažinti minėtas problemas, tokias kaip testų kūrimo sudėtingumas ir ilgas testavimo laikas, ir kartu pateikti tikslus formuojamojo vertinimo duomenis, suderinamus su Hattie trijų klausimų grįžtamojo ryšio, kompetencija pagrįsto mokymosi ir vertinimo metodikomis.

Pagrindinis užduoties iššūkis – surasti silpniausią mokinio išmanomą temą. Tam algoritmas turi ištirti (angl. *Explore*) įvairias temas ir „išnaudoti“ (angl. *Exploit*) konkrečią temą. Taip gali būti gautas tikslesnis silpnai mokino žinomos temos pažinimo lygis. Dilema – kada tirti nežinomas temas, o kada giliau nagrinėti silpnai mokiniui pažįstamas temas – yra MAB algoritmų pagrindas. Šioje disertacijoje pasirinktas euristinis šios problemos sprendimo principas yra optimizmas neapibrėžtumo akivaizdoje, o šiuo principu veikiantis algoritmas – viršutinė pasitikėjimo riba (angl. *Upper-Confidence Bound*, UCB). UCB algoritmas yra vienas paprasčiausių algoritmų, užtikrinančių mažesnę nei tiesišką gailesčio (angl. *Regret*) didėjimą. Algoritmas siūlo pasirinkti veiksmą, kurio viršutinė pasitikėjimo riba yra didžiausia, o mūsų modelio atveju – temą, kurios apatinė pasitikėjimo riba yra mažesnė. Tuomet pasirinktas klausimas n iš temos t bus:

$$t_n = \underset{t \in T}{\operatorname{argmin}} \left(\mu_t - C \sqrt{\frac{\log n}{N_t}} \right), \quad (\text{S3.1})$$

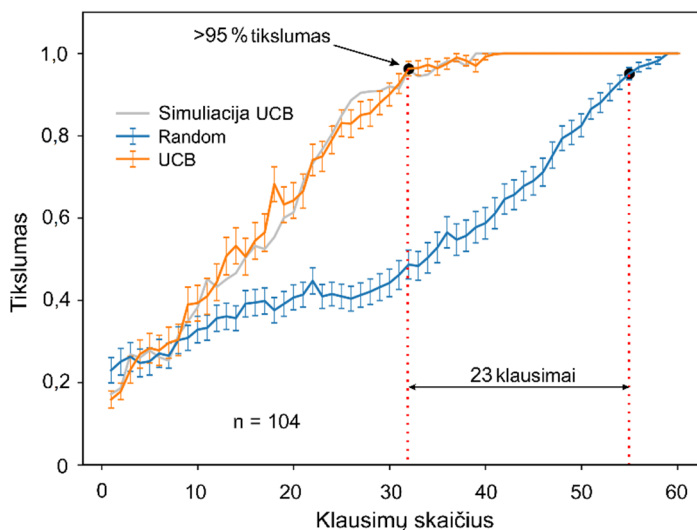
čia C – konstanta, kurią galima pasirinkti siekiant reguliuoti antrojo tyrimo komponento įtaką temos pasirinkimui; N_t – iki šiol užduotų klausimų šia tema skaičius. Daugėjant temos klausimų, mažėja formulės neapibrėžtumas ir tyrinėjimo komponentas (Girisgin et al., 2015). Taigi algoritmas ieškos silpniausių mokinio žinių temų, jas identifikavęs nuodugniai apklaus mokinį minėtomis temomis. Tai vertinga pedagoginiu požiūriu, nes nustatytas trūkstamas temines žinias galima koreguoti. Be to, algoritmas surinks nuodugnesnę informaciją apie silpniausią temą, ją „išnaudodamas“ (angl. *Exploit*), o tai bus naudinga koreguojant žinias po vertinimo. Tuo atveju, kai temos elementų fondas yra išnaudotas, algoritmas pasirenka temą su antra mažiausia verte, apskaičiuota pagal (S3.1) formulę.

Siekiant įvertinti UCB algoritmo naudojimo praktiškumą įvairiose klasėse, buvo ištirta testų erdvė. Iš lentelės S3.1 matyti, kad UCB testavimo metodas tampa vertingesnis, kai didėja klausimų ir temų skaičius. Šiame eksperimente imituojamų mokinių skaičius buvo 500.

S3.1 paveiksle pateikti eksperimento su mokiniais ($n = 104$) rezultatai. Formuojamojo vertinimo testą sudarė 15 temų klausimai, kiekvienoje temoje buvo po 4 klausimus, iš viso 60 klausimų. Rezultatai rodo, kad šiame scenarijuje taikant tradicinius testavimo metodus testo trukmę būtų galima sutrumpinti nuo 60 klausimų iki 55 klausimų, jei būtų siekiama 95 % tikslumo nustatant silpniausią temą. Tačiau taikant UCB adaptyvųjį vertinimą, testo trukmę būtų galima sumažinti beveik perpus – iki 32 klausimų. Empiriniai rezultatai (oranžinė ir mėlyna kreivės) atitinka modeliavimo projekcijas (pilka kreivė). Šiam konkrečiam testui ir mokinių grupei UCB galėtų sutrumpinti testo trukmę 23 klausimais, jei siekiame tokio paties (>95 %) tikslumo.

S3.1 lentelė. Atsitiktinių klausimų testo trukmės sumažėjimas, reikalingas 95 % tikslumui pasiekti, naudojant UCB algoritmą su optimalia tyrimo konstanta, palyginti su tradiciniu atsitiktinių klausimų metodu. Visi skaičiai pateikti optimaliai C vertei, čia T – temų skaičius; Q – bendras klausimų skaičius. Žalia spalva rodo, kad testo ilgis labai sumažėja, o raudona spalva žymima mažesnė nauda

T \ Q	4	8	16	32	64	128	256	512
2	33 %	0 %	8 %	0 %	24 %	-2 %	22 %	24 %
4	0 %	29 %	31 %	35 %	36 %	44 %	43 %	52 %
8		0 %	40 %	45 %	50 %	54 %	58 %	60 %
16			0 %	57 %	50 %	59 %	67 %	71 %
32				0 %	65 %	52 %	64 %	73 %
64					0 %	62 %	56 %	66 %



S3.1 pav. 104 mokinių formuojamojo vertinimo vidutinis tikslumas

Simuliuojamos viršutinės patikimumo ribos (UCB) kreivė atvaizduoja 104 sintetinius duomenis apie studentų testo rezultatus. Ji skirta parodyti, kad sintetiniai eksperimentai atitinka tikrovę. Paklaidų juostos rodo SEM.

Silpniausiai žinomos mokiniui temos nustatymo sudėtingumas priklauso nuo to, kaip stipriai koreliuoja tos pačios temos atsakymai. Kai atsakymų dispersija lygi nuliui, pakanka užduoti tik vieną klausimą, kad sužinotume mokinio žinias apie likusius temas elementus.

Tačiau kai atsakymai nesusieti, temos žinias įvertinti yra sunkiau ir tai gali paneigti siūlomo UCB testavimo modelio prielaidą. Taigi buvo atliktas eksperimentas, kuriuo siekta įvertinti vienos temos elementų žinių koreliacijos poveikį metodo veiksmingumui. Kaip ir buvo tikėtasi, silpnų temų nustatymas yra trivialus stipriai koreliuotų atsakymų

atveju. Tačiau net ir nekoreliuotų atsakymų atveju UCB rezultatai yra dvigubai geresni nei atsitiktinio klausinėjimo.

UCB algoritmas gali sutrumpinti vertinimo trukmę pusiau neprarandant testavimo tikslumo. Net labai trumpų testų, kuriuose yra nedaug temų, atveju algoritmas leidžia gerokai sutrumpinti testo trukmę (S3.1 lentelė). Tačiau testai, kuriuose kiekvienoje temoje yra tik po vieną klausimą, neturi jokio pranašumo. Tuomet temos sąvoka praranda savo pedagoginę prasmę. Yra įrodymų, kad mokymuisi skirtas laikas teigiamai veikia mokinių pasiekimus (Bratti & Staffolani, 2013), todėl pageidautina mažinti vertinti skiriamą laiką. Eksperimentai su universiteto studentais patvirtina modeliuojant padarytas išvadas (S3.1 pav.). Toks stiprus testo trukmės pakeitimas gali pakeisti dinamiką klasėje, nes dėstytojui nereikės visos pamokos skirti vien formuojamajam vertinimui. Tai savo ruožtu gali padidinti galimybes teikti individualizuotą grįžtamąjį ryšį mokiniams, susijusį su geresniais rezultatais (Marzano et al., 2000; Shute, 2008). Šis pranašumas neatsiperka dėl testo kūrimo sudėtingumo, priešingai nei IRT ir Elo reitingais grindžiamose sistemose, kuriose testo kūrimas kai kuriose situacijose gali būti pernelyg sudėtingas dėl elementų kalibravimo problemos (Pelánek, 2016; Wainer, 2001). Dėl šio esminio skirtumo mes nelyginame UCB ir IRT bei Elo algoritmų efektyvumą. Toks palyginimas neatskleistų nė vieno iš šių metodų. Jei UCB rezultatai yra prastesni (kaip galima daryti prielaidą), jo naudojimo paprastumas ir nepriklausomybė nuo klausimų elementų kalibravimo daro jį įdomia tradicinio vertinimo alternatyva.

Palyginti su IRT ir Elo algoritmais, UCB gaus mažiau informacijos apie mokinį matematine prasme (Huang et al., 2009; Pelánek, 2016) ir tai gali būti laikoma trūkumu. Tačiau ne visa informacija yra vienodai vertinga pedagoginiu požiūriu (Nicol & Macfarlane-Dick, 2006; Sadley & Royce, 1989). Tarkime, kad vertiname mokinių žinias dviejose temose. Artėjame prie testo pabaigos ir nustatėme, kad pirmosios temos žinios yra pakankamos, tačiau trūksta žinių apie antrąją temą. Dėl UCB algoritmo pobūdžio antroji tema yra labiau nagrinėjama nei pirmoji, todėl tikimės, kad nagrinėdami antrąją temą gausime mažiau informacijos. Tačiau pedagoginiu požiūriu informacija apie antrą temą gali būti vertingesnė, siekiant patenkinti vertinimo poreikius pagal trijų klausimų grįžtamojo ryšio modelį (Hattie & Timperley, 2007). Remiantis testo rezultatais, studentui reikia mokytis antrą temą. Jei ketiname tęsti jos mokymą, galime pasinaudoti papildomais diagnostiniais duomenimis, kad sutaupytime laiko ir pastangų nemokydami to, ką mokinys jau žino. Tikslesnių duomenų apie pirmąją temą iš karto panaudoti negalime.

Formuojamasis vertinimas, paremtas UCB algoritmu, duoda daug žadančių rezultatų, palyginti su tradiciniais vertinimo metodais. Taikant šį metodą galima gerokai sutrumpinti testo trukmę nesumažinant vertinimo tikslumo. Naudojant testus, kuriuos sudaro 8 klausimai, šis rodiklis sumažėja 29 %, o testus, kuriuos sudaro 512 klausimų, rodiklis sumažėja 73 %. Atsakymų į tos pačios temos klausimus dispersija turi nedidelę įtaką vertinimo tikslumui, kai empiriškai nustatytos vertės (0,1–0,25), todėl šis algoritmas tinka tais atvejais, kai klausimai nebūtinai matuoja tą patį latentinį įgūdį ar bruožą. Tačiau mokinių tikslumo pasiskirstymas klasėje nėra normalus. Net esant dideliame klasės tikslumo vidurkiui (95 %), daugumą tiksliai įvertintų mokinių atsveria nedidelė mažuma mokinių, kurių silpniausios temos buvo nustatytos neteisingai. Siekdami kompensuoti šią algoritmo savybę, rekomenduojame pedagogams siekti >99 % tikslumo atliekant esminius UCB formato vertinimus. Manome, kad UCB grindžiamas formuojamasis vertinimas turi praktinio

taikymo pedagoginį potencialą ir turėtų būti toliau tiriamas. Skirtingai nuo IRT ir Elo reitingais pagrįstų vertinimo metodų, UCB pagrįstam vertinimui nereikalingas klausimų kalibravimas ir jis nepriklauso nuo ginčytinos prielaidos, kad skirtingi klausimai matuoja tą patį latentinį požymį. Todėl UCB metodas priklauso kitai, menkai ištirtai lengvai įgyvendinamų ir prižiūrimų formuojamojo vertinimo sprendimų klasei. Jis gali būti nauja ir perspektyvi tradicinio linijinio vertinimo alternatyva tais atvejais, kai IRT ir Elo metodai laikomi per sudėtingais įgyvendinti ir prižiūrėti. Ateityje gali būti įdomu atlikti UCB vertinimo metodo lyginamąjį tyrimą su nustatytais metodais, priklausomais nuo elementų kalibravimo.

Naujai pasiūlytas Elo-Bandit algoritmas (EBA), pagrįstas Elo reitingo ir viršutinės patikimumo ribos algoritmais (Melesko & Ramanauskaitė, 2021). Jis remiasi klasikine testų teorija pagrįstu UCB testavimo algoritmu. Formuojamojo vertinimo problemą galima sumodeliuoti kaip įvesties duomenų rinkinio apie tam tikrą skaičių temų priėmimą, mokinių vertinimo atlikimą ir individualizuotų mokymosi pasiūlymų kiekvienam mokiniui grąžinimą. Pasiūlymai turėtų atspindėti temą, kurios labiausiai reikia mokytis, ir elementų sudėtingumą (pirmiausia pateikiami lengvesni elementai). Vertinant algoritmas turi derėti su temomis, kurių mokinio žinios dar nežinomos, ir su silpnų temų tyrimu, kad būtų tiksliai nustatyta, kurie elementai mokiniui yra žinomi. Siekiant įvertinti mokinio žinias kiekviena tema ir įvertinti elementų sudėtingumą, pasiūlytas EBA algoritmas naudoja Elo reitingų sistemą. S3.2 lentelėje pateiktas siūlomas EBA algoritmas.

Pirmiausia algoritmas bando pasirinkti mažiausiai žinomą temą, tada užduoda lengviausią tos temos klausimą. Silpniausia tema parenkama pagal šią lygtį:

$$t_n = \underset{t \in T}{\operatorname{argmin}} \left(sElo_t - C \sqrt{\frac{\log n}{N_t}} \right), \quad (\text{S3.2})$$

čia t – pasirinkta n testo klausimo tema; $sElo_t$ – sunkumo įvertinimas. Antrasis parametras skliausteliuose nusako sunkumo įvertinimo neapibrėžtumą ir yra valdomas tyrinėjimo konstantos C , bendro užduodamų klausimų skaičiaus n ir temos klausimų skaičiaus N_t . Neapibrėžtumo parametras mažėja, nes algoritmas tiria mokinio žinias apie temą ir $sElo_t$ įvertis tampa tikslesnis (Girisgin et al., 2015). Kiekvienam mokiniui priskiriamas pradinis Elo įvertinimas kiekvienoje temoje t , $sElo_t = 1000$.

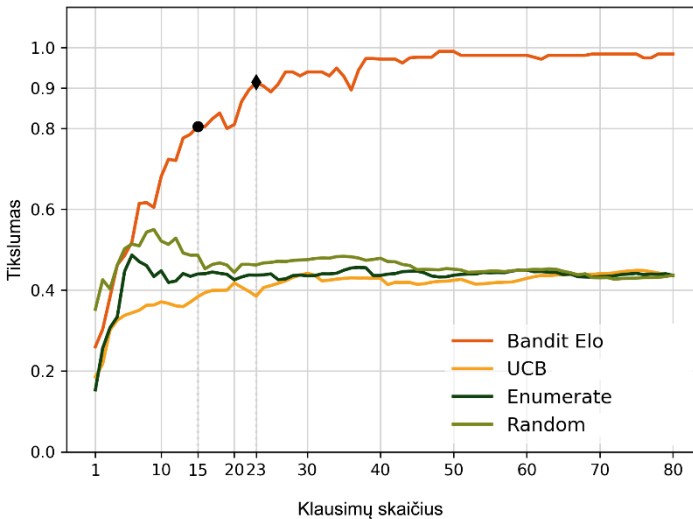
Algoritmas kaip įvesties duomenis priima elementų grupę, kurios svoriai apibūdina jų priklausymą tam tikroms temoms ir spėjimo tikimybės koeficientą. Pavyzdžiui, kelių atsakymų testo elemento, turinčio vieną teisingo atsakymo variantą iš keturių, spėjimo tikimybės koeficientas yra $P_g = 0,25$. Jei nepateikta jokių elementų sudėtingumo kalibravimo duomenų, EBA kiekvienam elementui priskiria pradinį Elo įvertinimą $iElo = 1000$. Teisingo atsakymo tikimybę P_c į $iElo$ balą, turinčio vienos temos t elemento teisingo atsakymo tikimybę, galima apskaičiuoti naudojant pasislinkusią logistinę funkciją:

$$P_c = P_g + \frac{1 - P_g}{1 + 10^{\frac{iElo - sElo_t}{400}}}. \quad (\text{S3.3})$$

Eksponentės pagrindas ir vardiklis eksponentėje yra tokie patys kaip ir šachmatų varžybose naudojamame Elo balų skaičiuje. Kai kur jie buvo pasirinkti, kad būtų suprantami žmogui. Naujo mokinio Elo įvertinimas tam tikra tema $sElo_t$, mokiniui pateikus teisingą $X_{ans} = 1$ arba neteisingą $X_{ans} = 0$ atsakymą į klausimą, apskaičiuojamas pagal tikėtiną teisingo atsakymo tikimybę:

$$sElo_t' = sElo_t + k(X_{ans} - P_c). \quad (S3.4)$$

S3.4 lygties konstanta k nurodo Elo reitingo jautrumą pokyčiams arba didžiausią Elo reitingo pokytį, kurį Elo reitingas gali padaryti atsižvelgiant į labiausiai neįtikėtiną rezultatą.



S3.2 pav. Vidutinis algoritmo tikslumas vertinant personalizuotą siūlomų mokytis užduočių parinktį, lyginant su IRT tikslumu

Algoritmas buvo pritaikytas ir jo siūlomų mokytis užduočių parinktės tikslumas įvertintas naudojant eksperimento metu gautus universiteto studentų duomenis (S3.2 pav.). Formuojamojo vertinimo testą sudarė 80 klausimų 15 temų; jį atliko 106 studentai. Rezultatai patvirtina stebėjimus, atliktus naudojant sintetinius duomenis. Elo-Bandit yra perspektyvi alternatyva IRT pagrįstiems testams, nes pasiekiamas beveik toks pat tikslumas – 97 %, o testo trukmė yra perpus trumpesnė. Elo-Bandit gerokai tiksliau suranda tinkamus elementus nei CCT pagrįsti algoritmai. Elo-Bandit algoritmas geriau veikia tikruosius duomenis, palyginti su modeliavimu. Siūlomi daugiau nei dvigubai tikslesni rezultatai ilgesniems nei 15 klausimų testams. 80 % tikslumas buvo pasiektas esant 15 klausimų, palyginti su 23 klausimų simuliacijos verte, ir 90 % tikslumas esant 23, palyginti su 45 klausimų simuliacijos verte, nepaisant to, kad algoritmo konstantos buvo kalibruotos naudojant imitacinius duomenis.

Elo-Bandit algoritmas yra perspektyvus kaip alternatyva IRT ir CTT grindžiamiems testams formuojamojo vertinimo programose. Nepaisant pedagoginės formuojamojo vertinimo vertės (Dunn & Mulvenon, 2009; Guskey & Bailey, 2001), jis turi alternatyviasias sąnaudas: vertinimas atima laiką, kurį būtų galima skirti tiesioginiam kurso medžiagos studijavimui per paskaitas ir savarankiškam mokymuisi. Kiekvienai veiklai skiriamas laikas turi akivaizdų poveikį mokymosi rezultatams (Bratti & Staffolani, 2013). Be sutaupyto laiko, trumpesnė vertinimo trukmė gali padėti kovoti su nuovargiu nuo sprendimų, o tai gali lemti efektyvesnę mokymąsi po testo (Pignatiello et al., 2020). Elo-Bandit algoritmas suteikia galimybę pasinaudoti formuojamuoju įvertinimu taupant laiką ir mažinant nuovargį. Eksperimentinėje aplinkoje pastebėtas potencialus testo trukmės sumažėjimas buvo apie 50 %. Be to, formuojamojo vertinimo metu rezultatų tikslumas yra mažiau svarbus nei apibendrinamojo vertinimo metu. Dėl netikslumo apibendrinamojo vertinimo metu nusipelniusiam kandidatui gali nepavykti įsidarbinti arba įstoti į universitetą. O dėl tam tikro netikslumo, atliekant formuojamąjį vertinimą, mokymo sistema gali tapti neoptimaliu mokymosi objektu. Todėl gali būti racionalu siekti mažesnio tikslumo, pavyzdžiui, 80 ar 90 procentų, kai tai leidžia gerokai sutrumpinti testavimo trukmę. Pateiktame 2021 atvejo tyrime, kurio klausimų fondą sudarė 80 klausimų, 106 mokinių vidutinė vertinimo trukmė buvo 39 minutės, t. y. apie 30 sekundžių vienam klausimui. Pasirinkus 80 proc. arba 90 proc. tikslumą, būtų galima sutrumpinti vidutinę vertinimo trukmę atitinkamai iki mažiau nei 8 ir 12 minučių. Taip smarkiai sutrumpinus testo trukmę, dėstytojai galėtų įtraukti formuojamąjį vertinimą į mokymosi procesą ten, kur anksčiau tai buvo nepraktiška.

Galbūt pagrindinis veiksnys, ribojantis IRT grindžiamą vertinimą klasėse, yra kalibravimo poreikis (Lord, 2012). Mokytojas negali paprasčiausiai rašyti ir papildyti klausimų banko. Be to, kad reikia didelio mokinių skaičiaus, kuris daugeliui mokytojų yra neprieinamas, klausimų sudėtingumo parametrai taip pat priklauso nuo mokinių kilmės, o tai dar labiau apsunkina šią problemą. Šioje disertacijoje pasiūlyto algoritmo nereikia kalibruoti. Kol kas joks panašus formuojamojo vertinimo sprendimas nebuvo pasiūlytas ir pritaikytas naudojant realius mokinių testavimo duomenis. Tačiau mokymosi objekto sudėtingumui įvertinti buvo naudojamas Elo reitingas. 2019 m. Mangaroska ir kt. (2019) paskelbė 22 studentų duomenis, rodančius teigiamus programavimo pratybių rekomendavimo sistemos rezultatus. Pankiewicz ir Bator (2019) nustatė, kad Elo įvertinimas yra geresnis užduoties sudėtingumo matas už teisingo ir mokinio grįžtamojo ryšio proporcijas. Galiausiai, remdamasis modeliavimu, Antal nustatė, kad Elo reitingas gali būti pranašesnis už IRT adaptyviam užduočių eiliškumui nustatyti praktinėse sistemose, nes nereikalauja išankstinio užduočių testavimo (Antal, 2016). Šios išvados sutampa su šiame tyrime gautais teigiamais rezultatais: Elo reitingas yra tinkama alternatyva IRT vertinimui formuojamajame kontekste.

Siūlomas Elo-Bandit algoritmas gali iki 70 % sutrumpinti formuojamojo vertinimo laiką, kai siekiama 90 % elementų lygio tikslumo. Ši funkcija yra palanki mokiniams, nes, palyginti su tradiciniais metodais, taupo jų laiką ir padeda išvengti nuovargio. Siūlomas algoritmas taip pat naudingas dėstytojams, nes nereikalauja išankstinio elementų testavimo.

Trečiojo skyriaus išvados:

1. Naujai pasiūlytas UCB algoritmo taikymas yra efektyvus klasikinių formaliojo vertinimo metodų pakaitalas, galintis greitai nustatyti silpnąsias mokinių žinių sritis. Siekiant 80 % tikslumo, jis leidžia sutrumpinti vidutinį vertinimo laiką nuo 39 minučių iki mažiau nei 8 minučių, o siekiant 90 % tikslumo – iki 12 minučių.
2. Naujai sukurtas Elo-Bandit algoritmas yra efektyvus IRT grindžiamo formuojamojo vertinimo pakaitalas, panaikinantys brangiai kainuojančio išankstinio testavimo poreikį, be to, jis padeda sutrumpinti testavimo trukmę 70 %, kai siekiama 90 % tikslumo.

Bendrosios išvados

1. Literatūros apžvalga parodė kad, mokymuisi personalizuoti plačiai naudojamų mokymosi stilių efektyvumas yra abejotinas, ir kad grįžtamasis ryšis yra ypač efektyvus (iki dviejų standartinių nuokrypių) mokymo procese, jis gali būti suteiktas naudojant ugdomąjį įvertinimą.
2. Siūloma išmanioji mokymosi sistema pagerina mokymosi personalizavimą. Ji leidžia naudotis Felderio ir Silvermano mokymosi stiliais ir naudoja formuojamojo vertinimo agentą pagrįstą naujai pasiūlytais algoritmais.
3. UCB algoritmo taikymas yra efektyvus klasikinių formuojamojo vertinimo metodų pakaitalas, galintis greitai nustatyti silpnąsias mokinių žinių sritis: siekiant 80 % tikslumo, jis leidžia sutrumpinti vidutinį vertinimo laiką nuo 39 minučių iki mažiau nei 8 minučių, o siekiant 90 % tikslumo – iki 12 minučių.
4. Naujas Elo-Bandit algoritmas yra efektyvus IRT metodo pakaitalas, taikytinas formuojamajam vertinimui. Jis nereikalauja kalibravimo, bet sutrumpina testo trukmę 70 %, kai siekiama 90 % tikslumo.

Jaroslav MELEŠKO

FORMATIVE ASSESSMENT METHODS FOR
INTELLIGENT LEARNING SYSTEMS

Doctoral Dissertation

Technological Sciences,
Informatics Engineering (T 007)

INTELEKTUALIOMS ELEKTRONINIO MOKYMOSI
SISTEMOMS SKIRTI FORMUOJAMOJO VERTINIMO
METODAI

Daktaro disertacija

Technologijos mokslai,
informatikos inžinerija (T 007)

2022 07 28. 11,0 sp. l. Tiražas 20 egz.
Leidinio el. versija <https://doi.org/10.20334/2022-032-M>
Vilniaus Gedimino technikos universitetas
Saulėtekio al. 11, 10223 Vilnius,
Spausdino UAB „Ciklonas“
Žirmūnų g. 68, 09124 Vilnius