A Systematic Review on Assessment in Adaptive Learning: Theories, Algorithms and Techniques

Adel Ihichr, Omar Oustous, Younes El Bouzekri El Idrissi, Ayoub Ait Lahcen
National School of Applied Sciences, The Engineering Sciences Laboratory at Ibn Tofail University, Kenitra, Morocco

***Abstract*—Computerized knowledge assessments have become increasingly popular, especially since COVID-19 has transformed assessment practices from both technological and pedagogical standpoints. This systematic review of the literature aims to analyze studies concerning the integration of adaptive assessment techniques and algorithms in Learning Management Systems (LMS) to generate a global vision of their potential to enhance the quality and adaptability of learning, and to provide recommendations for their application. A review of international indexed databases, specifically Scopus, was conducted, focusing on studies published between 2000 and 2024. The PICO framework was used to formulate the search query and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to select 66 relevant studies based on inclusion and exclusion criteria such as publishing year, document type, subject area, language, and other factors. The results reveal that integrating adaptive assessments positively impacts the quality of learning by generating short tests dynamically adapted to students' skills, learning styles, and behaviors. Furthermore, the findings identify various techniques and algorithms used, as well as their main features and benefits. These tools tailor adaptive learning programs to meet students' specific needs, preferences, and proficiency levels, thereby enhancing student motivation and enabling them to engage with material that matches their knowledge and abilities. In conclusion, the systematic review emphasizes the significance of integrating adaptive assessments in educational environments and offers tailored recommendations for their implementation to provide adaptive learning. These recommendations can be adopted and reused as guidelines to develop new and more sophisticated assessment models.**

Keywords—Adaptive assessment; adaptive learning; test; education; techniques

I. Introduction

Adaptability changes the method of delivering education, and learners use adaptive learning systems (ALS) as part of blended learning or fully online learning, it can be obligatory with credit or optional courses without any credit [1]. The aim of ALS is to alter instructions using a set of predefined rules to provide learning materials adapted to the needs and behavior of the student [2]. In addition, learner assessment is one of the principal moments in the educational process [3], it offers the possibility to construct a continuous and reversible process, over the learning life cycle of the student. According to OECD [4], assessment is the process of measuring and/or collecting and using evidence and proof about the outcomes of students.

Many people usually imagine when they first hear the word assessment that it solely refers to the collecting and analyzing of some information about a learner. However, assessment can also involve interpreting and acting on information gathered about a learner's understanding and/or performance in relation to educational goals.

Given the critical importance of adaptive assessment in enhancing educational outcomes and the evolving landscape of personalized learning [5], it is imperative to develop effective assessment mechanisms that can adapt to individual learner needs. Adaptive assessment systems have emerged as a promising approach to improve the precision and relevance of student evaluations. These systems utilize advanced algorithms and theoretical frameworks to dynamically adjust the difficulty and content of assessments based on real-time analysis of student performance and behavior [6]. Furthermore, assessment can also involve interpreting and acting on information gathered about a learner's understanding and/or performance in relation to educational goals. The principal challenge posed is not just the availability of learning content to learners but also the ability to present knowledge in the right place, time, and form.

Despite the promising potential of adaptive assessment to transform educational practices, comprehensive and systematic research in this domain remains limited. This study aims to address this gap by conducting a thorough literature review of the current state-of-the-art in adaptive assessment. The findings will provide valuable insights for researchers and practitioners, helping them understand current research trends, identify gaps, and develop more effective adaptive assessment tools. Ultimately, this research seeks to contribute to the advancement of adaptive assessment systems, which are essential to modern educational environments, thereby enhancing student learning experiences and outcomes.

This research stems from the need to understand the most recent advances in adaptive assessment within educational practices. The main contribution of this study is to identify how various assessment implementations and methods naturally emerge to meet the demand for measuring learning outcomes.

To achieve this objective, a systematic review of numerous studies collected from the Scopus database was conducted, employing inclusion and exclusion criteria. Subsequently, through the analysis of our selected studies, a framework was derived for an effective integration of adaptivity in knowledge assessment. The necessity for developing adaptive assessment-driven framework stems from the significant dropout rates observed in traditional assessment offered to students. In response to this challenge, a framework that offers adaptive, personalized advanced learning assessment was embraced, helping to place the student at the center of the learning process to maximize learning outcomes.

Following the review and examination of the 66 studies included in the research, three key themes were identified. These themes encompass the role of knowledge assessment in learning and education as a scaffolding instrument, the identification of the most used adaptive assessment theories, algorithms, and the proposed guidelines for defining the main characteristics of an effective adaptive assessment.

The paper is organized into several key sections to offer a comprehensive understanding of adaptive assessment in educational practices. Section II examines adaptive learning in detail, addressing its definition, role, advantages, and various assessment methods. Section III outlines the research approach, detailing the formulation of research questions, search strategy, and criteria for selecting studies. Section IV provides a thorough analysis of the included studies, covering statistical descriptions, key theories and models underlying adaptive assessment, and recommendations for future directions. Section V presents a discussion that synthesizes the findings, highlighting the strengths, and practical implications of the reviewed adaptive assessment methods. Finally, the conclusion summarizes in Section VI, the significance of adaptive assessment in enhancing ALS, providing a clear and detailed perspective on the role and influence of adaptive assessment in contemporary education. This structure is designed to offer a comprehensive overview and critical analysis of adaptive assessment's impact on educational practices.

II. Literature Review

Adaptive learning has revolutionized educational practices by tailoring the learning experience to individual student needs. This section explores the different aspects of adaptive learning, focusing on its definition, role, advantages, approaches, and various assessment methods.

A. *Adaptive Learning*

Adaptive learning, by definition, is a methodology to adjust and personalize a learning process notably content to a learner, in order to fit to different situations and circumstances [7]; thus, adaptive learning gives the possibility to make a learner a collaborator in the process of education instead of a passive recipient of information [8] because each learner has a specific optimal learning path composed of many connected dots and each dot represents a knowledge component or a skill [9]. Adaptive learning considers many aspects, such as the learner's current level [10], individual needs [11], the learning style [12], and different interactions with students to individualize all components of the learning process: content, interface, learning style and assessment [13]. Therefore, adaptive learning allows a student to progress to more challenging material based on their performance, while providing additional support to others to help them master skills. This approach ensures that learners are not constrained by the class pace but can follow their own individual learning pace [9]. Thus, the instructor and learner can make the right decision at the right moment by following the learning curve and learning trajectories [14]. In summary, adaptive learning seeks to tailor the educational experience to help students achieve their learning objectives by emulating the personalized, one-on-one interaction between a teacher and a learner.

B. *The Role of Assessment in Adaptive Learning*

"What we assess is what we value. We get what we assess, and if we don't assess it, we don't get it" is a statement declared by psychologist Lauren Resnick [15] which shows the importance of the assessment as the main and the critical step in the educational process, over twenty years ago, evaluation and assessment were considered a challenging research issue in education [16]. Many Adaptive Learning Systems (ALS) focus more on assessment than on content [17]. Assessment data collected from students' responses is used to personalize the learning process, creating tailored paths based on results. These paths are continuously updated with each assessment [13]. Thus, learners' assessment is considered a crucial factor for a successful e-learning process.

C. *Types of Knowledge Assessment*

Knowledge assessment can be globally divided into three types: summative, formative and pre-assessment. Generally, summative assessment is more dominant than formative assessment in e-learning [9] but the last two types have shown more potential, especially in adaptive learning [18] and researchers have shifted the focus to these two types.

1) Summative assessment: Also known as assessment of learning, this type of evaluation occurs at the end of the learning cycle to measure learner achievement and verify mastery of a curriculum unit [19]. It assesses the cumulative progress of the learner.

2) Formative assessment: Also called assessment for learning, it occurs throughout the learning cycle to provide both tutors and learners with feedback information to assist their learning experience and improve it [19]. This type assesses the quality of learning by positioning assessment between teaching and learning (current progress) instead of being positioned only at the end of the process [9], when formative assessments are so deeply embedded in the learning environment that the separation between assessment and learning is completely fuzzy and so unnoticeable to the students, this concept is known as stealth assessment [20]. Formative assessment allows us to refine the learning trajectory, increase the engagement and enthusiasm of a student to achieve a course and develop his self-regulation.

3) Pre-assessment also known as diagnostic assessment, this kind of assessment measures each learner's prerequisites, providing the teacher with a clear picture of the learners' level of skills [21]. It is frequently succeeded by a sort of compensatory instruction to eliminate obstacles and offer different kinds of remedial activities [18]. It is usually initiated at the entrance of a learning course. It is designed to gauge the current student's prior learning to detect learners' needs, competencies, preconceptions, and prior knowledge, to orient them toward the most suitable curriculum.

D. *Learning Outcomes Assessed in Online Learning*

Wei et al. [8] defined three principal learning outcomes assessed in online learning, to know cognitive, behavioral, and effective learning outcomes (see Fig. 1), based on a study of 65 peer-reviewed articles. The study demonstrated that these three aspects are correlated.

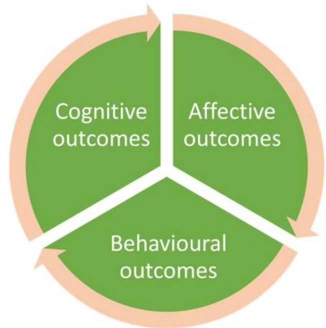

Fig. 1. Learning outcomes assessed.

1) Cognitive outcomes: Correspond basically to obtained knowledge and intellectual skills. Knowledge is all information that a student has learned from a topic, but intellectual skills concern different abilities, such as reasoning, thinking processes and decision making.

2) Behavioral outcomes: Refer to measuring learner engagement by monitoring his behavior while course learning such as duration and times of learning, participation in forums and discussion, submission of tasks, completion of tests.

3) Affective outcomes: Correspond to perceptions of students, their satisfaction with the course, appreciation of their learning experience and benefits expected by them after enrolling in a course.

In his paper, Shepard [22] suggested that most learning and assessment in higher education concentrates on cognitive outcomes rather than on affective outcomes, and behavior.

E. Methods of Assessment

A variety of assessment methods were developed by researchers to optimize the cognitive load, implement adaptive learning systems, and measure learning outcomes, such as the following:

1) Computer-graded tests: This method is gradually replacing the traditional method called instructor graded assignments, which allow instructors to diagnose their students closely and with more efficiency [23]. This method is used to recognize and accredit student learning, and it allows the generation of immediate feedback that is linked to the parameter configuration so that each answer is either right or wrong.

2) Self-assessment: This method is implemented in WEAS [24] and MAPS [25]. It allows students to self-assess their performance based on criteria and standards already fixed by teachers. It is a useful manner to prepare for the exam [26]. Generally self-assessment is combined with feedback and hints [27], so students participate in active learning by improving their knowledge and detecting possible misconceptions.

3) Peer assessment: This method is implemented in MAPS [25]. It is considered an educational activity where the assessor and the assessed have similar statuses, and each one assesses learning outcomes, quality, and the level of the other; it requires a mutual relation based on trust. Peer review, peer grading, peer feedback and peer evaluation are all synonyms of peer assessment [28]. Lu and Zhang [29] demonstrate that students benefit more from acting as assessors than from being assessed. Li [30] has proven that quality increases if peer assessment is done anonymously. One of the advantages of this method is the creation of a competitive atmosphere between students despite their social aspects.

4) First-step rapid diagnostic assessment: Kalyga [31] described this method as a diagnostic process in which each student studies a task for a limited time and is asked to put his first step toward the solution, so the tutor can determine the level of mastery of the student depending on his first step. If the learner possesses the required knowledge component, the final answer will be provided immediately; otherwise, the search process will commence.

5) Concept map: It is a graph in which nodes symbolize concepts, and the directed labeled links that interconnect the nodes represent relations between concepts. This research in [32] uses concept maps as a method for assessing students' comprehension of content by allowing them to display their understanding of concepts and connections between concepts in a graphic format. Using concept maps to assess learner knowledge at the level of understanding has many benefits, and this method is implemented in the Concept Mapping Tool.

F. From Traditional Assessment to Adaptive Assessment

The previous section highlighted the importance of knowledge assessment in providing adaptive learning. Now, this adaptability is also integrated into the assessment itself to achieve adaptive assessment. In contrast to conventional testing, which is based on fixed items that every examinee must tackle regardless of their knowledge level, adaptive assessment allows for the selection of questions based on the examinee's performance. An adaptive assessment provides a short, personalized test, its items change depending on the response of the student, and the difficulty of each question is correlated to the answer to the previous question [13]. In addition, the decision to stop testing is dynamically related to the student's performance shown in the test [33]. In brief, adaptive assessment seeks to avoid presenting easy questions to capable students, who are likely to answer correctly. Similarly, challenging questions are not presented to struggling students, who may find them difficult.

G. Advantages of Adaptive Assessment

Adaptive assessment retains the advantages of classical assessment in enhancing the learning process, but there are additional advantages that make adaptive assessment more efficient and effective, such as: Energizes and individualize the assessment process [13], reduce the length of the test by at least 60 percent, so it can detect the level of a learner with fewer questions [34], reduces the duration of the test by reducing the number of questions and items [35], increases motivation of the learner by suggesting easier questions [36], gives self-reliance [36], avoids annoying students by providing tests adapted to their level of knowledge and skill [21], provides more detailed statistics used to refine learning trajectory and to correct curriculum [13], helps ITS to make a rapid diagnosis of a student's characteristics and knowledge level to update their models [37], allow tutors to better differentiate between candidates by considering time analysis of responses [38], reduce cheating because the test is tailored to each learner due to presence of questions banks which offer possibility to vary questions [39], give the possibility to each learner to tackle his course and tests at any time, and from any place, and enhancing the quality of feedback given to learners in real time.

H. Approaches Followed in Implementing Adaptive Assessment

From the articles of Lendyuk et al. [36] and Al-Rajhi et al. [40], three approaches of tests in adaptive assessment can be deduced; after a short description of each one, Table I provides a comparison between these three approaches:

1) Pyramidal testing: It is used to assess learners without giving them a preliminary test, so all examinees take the same test with a middle level of difficulty, then the next task and question is related to the previous answer, if it is correct then the difficulty will increase and vice versa. Many variations of pyramidal testing have been developed such as constant step sizes, variable step sizes, truncated pyramids, and multiple-item pyramids.

2) Flexilevel testing: It programs the first task with a definite level of difficulty chosen by the tutor. Each level has one item, and the difficulty of the next item depends on the answer of the learner. It increases if the answer is correct and vice versa.

3) Stradaptive test: Stratified adaptive, in which many levels or strata of difficulties are defined. Each one group's test items have approximately the same average difficulty. These strata are classified in order of difficulty, so the next item is selected from the upper strata if the previous answer is correct; otherwise, the system suggests an item from the bottom.

TABLE I. COMPARISON BETWEEN TYPES OF ADAPTIVE TESTS

Type of adaptive test	Preliminary test	Difficulty level of the first question	Number of items by difficulty level	Next question
Pyramidal testing	No	Middle level	Depends on its variations	Depends on response
Flexilevel testing	Yes	Difficulty defined by the tutor	One	Depends on response
Stradaptive test	Yes	is typically set to be of moderate difficulty	Many	Next question from strata upper or downer depending on the response

III. MATERIALS AND METHODS

The integration of adaptive assessment in learning management systems has opened a new chapter in educational research. Historically, studies on educational assessment have primarily focused on traditional methods. However, with the advent of adaptive assessment, the landscape of evaluating student performance and learning outcomes has transformed significantly. Before delving into the detailed findings of our systematic review, we will provide, in this section, an overview of the research approach, including the formulation of research questions, the search strategy, and the criteria for selecting relevant studies.

A. General Background

This study aims for a systematic review that delves into how adaptive assessment is used in adaptive learning, shedding light on its contributions and implications. Insights were drawn from indexed articles and reviews from the esteemed Scopus database. Using the PICO framework and logical operators (AND, OR, NOT), a search question was elaborated to guide the research endeavours.

B. Research Question

The objective of this study is addressed by answering the following research questions:

- RQ1: In studies involving knowledge assessment, what is the role of assessment in adaptive learning, methods used in implementing assessment, and benefits of adaptive assessment?
- RQ2: What are the theories, algorithms and techniques underlying adaptive assessment in learning and how do these elements differ from each other?

C. Search Strategy

Following the PICO framework, the scientific articles were gathered from the largest database of scientific publications, namely Scopus. The authors focused on research in assessment and its features. A combination of keywords was used in the research study, taken from the title, abstract and keywords such as" adaptive learning" AND" assessment". The Boolean operators, parentheses, and stars were used wherever possible. Keyword synonyms were also used to obtain a more comprehensive search (as detailed in Table II). This meticulous approach aimed to maximize the potential results of our study.

D. Selection Strategy

1) Quantitative filtering: Following the formulation of the search query (Table II), a quantitative selection approach was employed, utilizing tools like Zotero software. The PRISMA framework was adhered to for analyzing and filtering the found studies based on inclusion and exclusion criteria outlined in (Table III).

2) Qualitative filtering: After the quantitative filtering, a qualitative selection was conducted based on:

- Title analysis according to the presence of the study's keywords.
- Abstract analysis based on sample and results.
- Content reading and synthesizing.

Table IV shows an overview of the number of articles found and included.

TABLE II. KEYWORDS IN THE SEARCH QUERY

learning **AND**	**OR** education **OR** student* **OR** knowledge
assessment **AND**	**OR** evaluation **OR** test*
Adapt*	**OR** adaptive learning
Search query	
SQ1: TITLE-ABS-KEY((learning OR education OR student*) AND (assessment OR evaluation OR test*) AND ("adapt*")) AND ABS ("adaptive learning" AND assessment)).	
SQ2: KEY((learning OR education OR student*) AND (assessment OR evaluation OR test*) AND ("adapt*")) AND ABS ("adaptive learning"))}	
SQ3 : TITLE(learning OR education OR knowledge) AND (assess* OR evaluat* OR test*) AND (adapt*))	

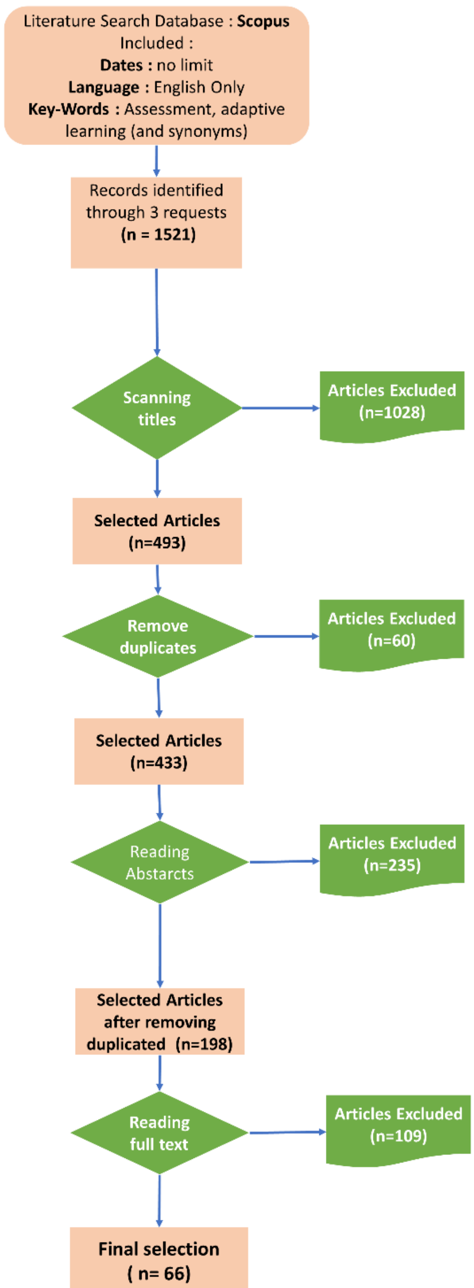

Fig. 2. Articles selection process and inclusion criteria.

TABLE III. INCLUSION AND EXCLUSION CRITERIA

Including Criteria	Excluding Criteria
Indexed in Scopus,	Not indexed in Scopus
Computer science and education subject areas	Other subject areas
English language	Not in English
Studies related to assessment knowledge	Studies not related to assessment knowledge

TABLE IV. SCOPUS SEARCH REQUESTS AND NUMBER OF RESULTS

Research field	Request	Result	Limited to English	After scanning title
TITLE-ABS-KEY	SQ1	406	396	122
KEY and ABS	SQ2	464	453	140
TITLE	SQ3	679	672	231
Total	-	1549	1521	493
Total after the merge	-	-	-	433

The flow chart diagram which is given in Fig. 2 describes the filtering process based on the PRISMA framework.

IV. ANALYSIS AND FINDINGS

Adaptive assessment leverages advanced theoretical frameworks to enhance the precision and effectiveness of evaluating student knowledge. This section offers a comprehensive analysis of the included studies, detailing statistical descriptions and exploring the primary theories that underpin adaptive assessment.

A. Statistical Description of the Included Studies

1) Databases: the scientific articles were gathered from the largest database of scientific publications, namely Scopus. The authors focused on research in assessment and its features.

2) Publishing year: Fig. 3 below shows a representation of selected studies according to publishing year, and it can be noticed that most of the papers were published after 2011, which explains the growing interest in improving assessment as a main component of the adaptive learning process.

3) Countries: As shown in Fig. 4, United States is having maximum number of included studies (27.5%), followed by Spain (07%), then Greece and China.

B. Theories Used in Adaptive Assessment

To the best of our knowledge, IRT (item response theory) and KST (knowledge space theory) are the two most powerful theoretical frameworks used in the development of efficient and effective adaptive assessment tools. This section first presents IRT, then KST, and finally a comparative table between the two theories in terms of the specific goals and requirements of knowledge assessment is drawn up.

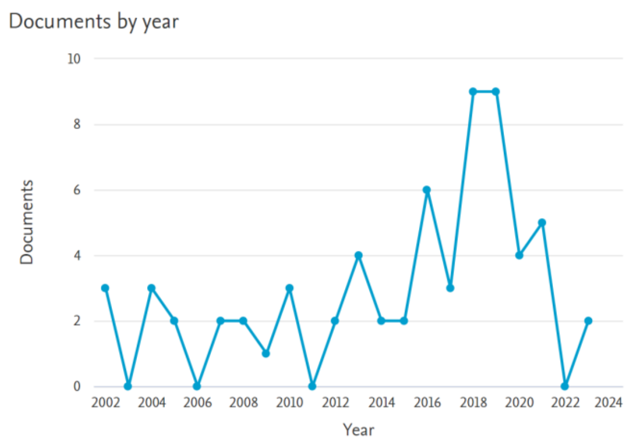

Fig. 3. Division of the included studies according to publishing year.

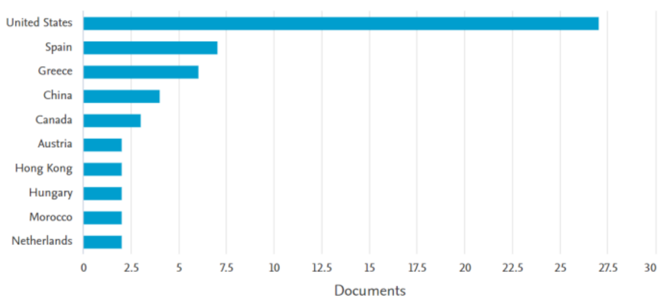

Fig. 4. Document by country.

Item response theory (IRT): is one of the most used theories in adaptive assessment, and its origins date back to Rasch and Lord in the 1950s. This theory supposes that an answer to a question is related to an unknown latent numerical θ, corresponding to the knowledge of the topic being assessed [41], so it describes how students interact with questions in tests; in other words, IRT tries to link observable actions as answers, responses to unobservable characteristics. This psychometric theory is used to estimate learner knowledge, and to develop learners' cognitive or non-cognitive measurement, to select the appropriate next question at each moment and to decide when the test is over [38]. Lendyuk et al. [36] explain that this variable nature of latent parameters provides the possibility for adaptable assessment, and mention that this combination between learner level and item difficulty on single measuring is the best advantage of IRT. Various IRT models exist, including Rasch, the 1PL model, the 2PL model, and GRM, chosen depending on item characteristics [42]. IRT is based on four principal assumptions: monotonicity (if the trait level increases, the probability of a correct answer also increases); one-dimensionality (one dominant latent trait to measure); local independence (for each level of ability, responses to separate items are mutually independent); and invariance (item parameters can be estimated from any position on the item response curve). See the handbook of Van der Linden [43] for more information about IRT.

1) Knowledge Space Theory (KST): In 1985, Doignon and Falmagne invented knowledge space theory (KST) which is based on probabilistic and combinatoric models [44]. In KST, each domain of knowledge is a collection of skills and concepts that must be learned by a student, and some skills are prerequisites of others, so if a learner acquires a skill, it becomes easier to master another; in other words, KST recognizes skills that are achievable without mastering any other skills [45]. KST is based on data collected from student answers to a set of questions reflecting different ability levels. These questions are not necessarily arranged in hierarchical order, and the answer can be correct or incorrect, so each student has a response state; for example, a learner who responds to questions 3, 4 and 5 correctly has a response state (3, 4, 5). For a test that contains seven questions, there are 27 possible response states, from a null state to the full response state in which the student responds to all questions correctly. After KST forms a subset called the knowledge structure, it contains possible knowledge states [46]. The KST provides an accurate statement of what the student knows, does not know, and is ready to learn next. There are many research articles that explain the use of KST in assessment such as article published by Arasasingham et al. [46] used Knowledge Space Theory to assess student understanding of stoichiometry, and the paper of Doble et al. [47] that examined several reliability measures for developing KST-based adaptive evaluation measures. Additionally, Fang et al. [48] construct student models based on knowledge space theory and can identify the student's present knowledge level through both initial and regular evaluations, including student task progression.

IRT and KST are both used in adaptive assessment but differ in their approach, focus, and other dimensions. Table V provides a comparison between the two theories.

C. Algorithms and Techniques Used in Adaptive Assessment

Knowledge assessment based on efficient techniques enhances the reliability of intelligent tutoring systems (ITS) by reducing the impact of human factors. This section divides and summarizes prominent algorithms and techniques used for assessing learner knowledge according to their technical differences, presenting a new taxonomy (see Fig. 5). The proposed taxonomy categorizes existing techniques into four groups: (1) techniques based on Bayesian Networks, (2) techniques based on logistic models, (3) techniques based on artificial intelligence, and (4) techniques based on learning styles and others.

TABLE V. COMPARISON BETWEEN IRT AND KST

	IRT	KST
Year of appearance	1980	1999
Process	Psychometric paradigms	Combinatorics, statistics, and stochastic processes
Unit of assessment	Item	Response state
Focus	modelling the connection between the test items and the examinee skills and performance	modelling the basic knowledge structure of a domain
Approach	Quantitative,statistical approach	Qualitative,behavioural approach
Objectif	modelling the statistical relationship between test question properties (as example: difficulty) and the likelihood that a student will respond correctly to the question.	modelling the prior connections between the concepts of a domain and represent them in the format of an oriented graph, and use this graph structure to orient the choice of the test items
Item selection	IRT can use either nonprobabilistic or probabilistic algorithms to select items tailored to level examiners and consider the characteristics of the items.	KST typically uses a nonprobabilistic algorithm to select items tailored to level examinee and the underlying knowledge structure of the domain.
Item types	Practically always made of multiple-choice questions or other dichotomous item types	Almost never is made of multiple-choice questions but can be used with a wider variety of item types, including open-ended.
The number of significantly different categories of test scores	Relatively small	Relatively big
Item selection	IRT can use either nonprobabilistic or probabilistic algorithms to select items tailored to the level examiners and consider the characteristics of the items.	KST typically uses a nonprobabilistic algorithm to select items tailored to the level examinee and the underlying knowledge structure of the domain.
Item types	Practically always made of multiple-choice questions or other dichotomous item types	Almost never is it made of multiple-choice questions but can be used with a wider variety of item types, including open-ended.
The number of significantly different categories of score in tests	Relatively small	Relatively large

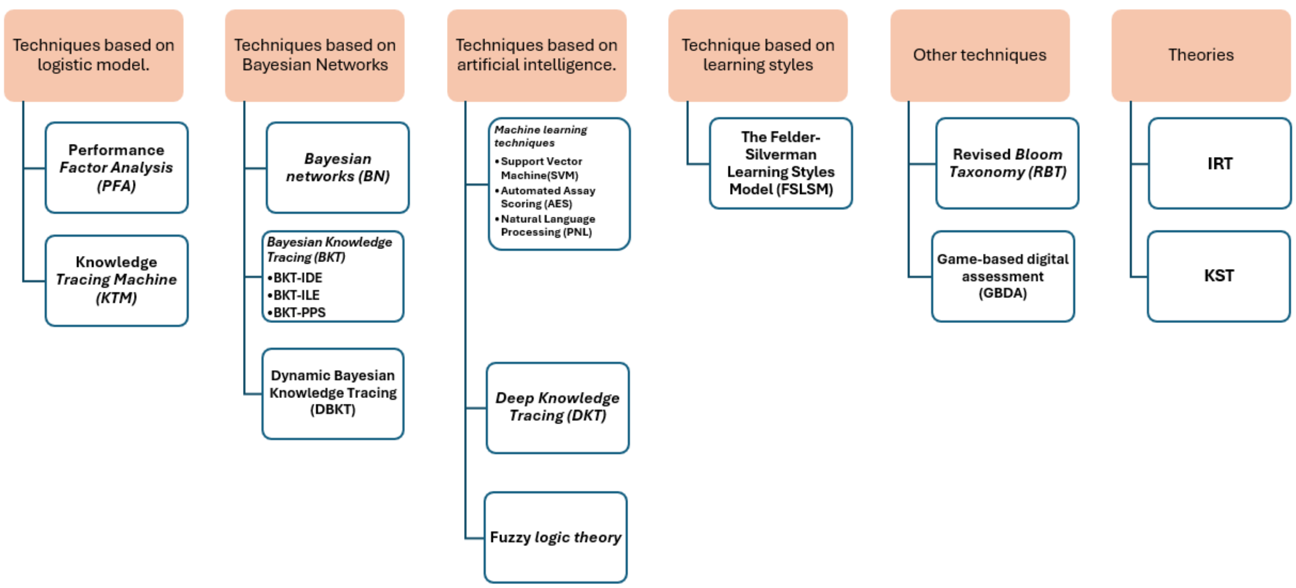

Fig. 5. Knowledge adaptive assessment techniques.

1) *Techniques Based on Bayesian networks:*

a) Bayesian Networks (BN): A Bayesian network is a technique based on Bayes' theorem that maps out cause and effect relationships in the form of a graphic representation. It is used to predict the probability that a factor is the most contributing factor in the occurrence of an event [49]. BN is well-suited for modeling content domains in learning assessments at various levels. Culbertson [50] details this in his state-of-the-art review, which describes the application of BN across 40 educational assessment systems in diverse domains. BN allows students to predict mastering in un-assessed sub-domains by utilizing results from assessed sub-domains, so assessment will be more effective in making precise and quick teaching decisions and optimizing the time invested in testing [50], especially when there is scarcity and uncertainty in data [51]. Collins et al. [52] used BN to provide an adaptive assessment of several features in a unique test. Xing et al. [53] suggest in their paper a Bayesian network model to assess dynamically the engagement of students in engineering design tasks DBNs are more powerful than BN in their ability to update in realtime the estimation of learner performance across multiple tests. DBNs can be used to infer previous, current and the future learner states. DBNs are a way to expand a static BN to model probability distributions over several points in time [42]. This technique serves two important functions in machine learning: classification and pattern discovery to capture and analyze information over time.

b) Bayesian Knowledge Tracing (BKT): BKT is a widespread approach based on BNs. This approach tries to model a learner's skill by calculating the probability of mastering a skill based on a set of parameters: guessing, slipping, the probability that a skill is already mastered, and the probability that the skill will be learned. BKT takes into account the time sequence for estimating the new skill and supposes that each skill learned is never forgotten [54]. BKT can involve sudden changes in knowledge. Researchers have developed several variants of BKT (Table VI), such as BKT-IDE [55], in which sliding and guessing are linked to the question to account for its difficulty, BKT-ILE [56] which considers the student's initial performance to be dependent on the question, and BKT-PPS [55], in which the accuracy of the learner's first tense is dependent on his or her initial performance. On the other hand, a version of the BKT was developed, Dynamic Bayesian Knowledge Tracing (DBKT) is used to model relationships and hierarchies between prior knowledge based on dynamic Bayesian networks. In DBKT, a student's knowledge mastery is also mapped by binary latent variables and can be inferred from the learning experiences of the student [57]. This approach considers various prior knowledge together in a unique model.

c) Comparaison between BKT and DBKT: Bayesian Knowledge Tracing (BKT) and Dynamic Bayesian Knowledge Tracing (DBKT) are two distinct methods for assessing learner knowledge, differing in several aspects. In general, DBKT provides a more complete and flexible assessment approach by considering prior knowledge, time dynamics and background characteristics. Table VII synthesizes some of the main distinctions between BKT and DBKT.

2) Techniques Based on logistic model:

a) Performance Factor Analysis (PFA): PFA is an algorithm following a statistical modeling approach that uses the logistic model to be sensitive to the most sought-after indicator in an evaluation: the performance measure or student's ability [58]. This technique is primarily sensitive to the relative ratio of correct to incorrect responses in an assessment, which allows for fine-tuning of the assessment's adaptability, and this sensitivity to accuracy can be used to measure learning, particularly its quality. In the article by Maier et al. [59], it is demonstrated that several issues need to be considered when utilizing PFA, such as parameter degeneration, especially for benchmarks where the learner's initial data are limited.

b) Knowledge Tracing Machine (KTM): The KTM model benefits from factoring machines (FM) to extend other logistic models (such as PFA) to larger scales. FMs were initially suggested as a broad predictor that runs on any real-valued vector of characteristics, that can model all the interactions across variables using factorized parameters. FMs are used to encode additional data about the student or the task in the model; in this way, KTM is particularly suited to modeling the student's knowledge mastery using a sparse set of weights for all the characteristics involved in the assessment [57]. In their article, Vie and Kachima [60] note that KTMs could be used to provide adaptive testing by selecting the next most appropriate item to be presented to a student, depending on the previous responses.

Both PFA and KTM can employ logistic regression models to predict the likelihood of a learner answering a question correctly; however, their approaches differ. Table VIII provides a comparison of the two techniques.

TABLE VI. COMPARISON BETWEEN BKT-IDE, BKT-ILE, AND BKT-PPS

	BKT-IDE	**BKT-ILE**	**BKT-PPS**
Question Difficulty	Related to sliding and guessing	Dependent on initial performance	Not directly addressed in the current question
Prior Performance	Not directly considered	Dependent on question difficulty	Used to adjust the first attempt probability

TABLE VII. COMPARING BKT AND DBKT

	BKT	**DBKT**
Based on	BN	DBN
Field of knowledge	Modeling a particular skill or concept	Takes into account several skills or concepts linked together in a single model
Time Dynamics	It is hypothesized that learning is independent of time	Considers the time dynamic of student learning and uses a sequence of responses to infer the state of knowledge over time.
Background characteristics	Does not expressly consider contextual factors such as feedback or advice.	Integrates background characteristics that may impact the student's learning experience
Model Flexibility	More flexible complex than DBKT	Less flexible than BKT

TABLE VIII. COMPARING PFA AND KTM

	Probabilistic Factor Analysis (PFA)	**Knowledge Tracing Model (KTM)**
Approach	Latent variable	History of answers
Logistic Model	Estimating the probability based on latent state of knowledge, item difficulty, and guessing parameters	Estimating the probability based on the current knowledge and item difficulty
Model Flexibility	Less flexible	Flexible
Item Selection	Yes	Yes
Feedback	No	Yes
Data Requirements	Large amount of data	Small amount of data

3) Techniques Based on artificial intelligence:

a) Machine learning techniques: Machine learning (ML), considered a part of artificial intelligence (AI), is one of the most challenging application areas in the field of learning assessment. It can be applied to improve item-based assessments [61]. In addition, the support vector machine, a supervised classification technique in machine learning, can diagnose students' knowledge mastery, especially in smaller test programs such as classroom assessments [62]. In brief, ML can generate computerized adaptive assessments that continuously provide feedback to instructors and staff on students' learning progress, the support they need, and their advancement toward learning goals.

Furthermore, the utilization of natural language is regarded as the most valuable technique for evaluating learning outcomes because it allows learners to display a deep comprehension of a given concept, but evaluating, rating, and giving feedback on writing assignments consume much time and effort, and can be biased by an unjust human assessor. In this context, researchers provide automated essay scoring (AES) as an emerging and growing technology of assessment in which computers replicate a written assignment's human evaluation by using multiple grading approaches, such as statistics, machine learning, and natural language processing (PNL) techniques [63]. ML can help rank students' handwritten assessments [64]. Using artificial intelligence techniques such as natural language processing and deep learning, AES systems can assess different dimensions of an assignment [65], such as grammar, syntax, and content, by examining learners' writing skills and recognizing their individual weaknesses and strengths.

b) Deep Knowledge Tracing (DKT): Deep Knowledge Tracing (DKT) [66], a pioneering algorithm that uses flexible deep recurrent neural networks to model student learning and trace their knowledge, is used to extract latent structure between assessments. DKT relies on RNN and LSTM models, which offer significant advantages by capturing complex representations of knowledge from assessments over time. This capability allows for substantially improved prediction performance across datasets from previous assessments. In addition, the learned model can be used to design the next assessment more suited to the student. DKT can suggest assessments that are more adaptive to individual needs and skip or delay questions that appear to be too easy or too difficult.

c) Fuzzy logic theory: Fuzzy logic theory, introduced by Zadeh in 1965, and adapted to assessment by Biswas in 1995 [67] has been widely used in educational assessment, Fuzzy logic is an artificial intelligence technique that can be considered ideally suited to provide a personalizable test, as it can successfully address the uncertainty and human subjectivity that characterize the identification of learner knowledge and learning needs [68]. Since all educational assessments deal with uncertainty, the ability of fuzzy logic to weigh these uncertainties makes it an excellent AI core in assessment systems. This application increases average class success and reduces test anxiety [69]. The authors of this article [70] utilize fuzzy logic to model learners' skills and knowledge, employing fuzzy sets to determine the difficulty and order of the questions on the test.

4) Technique *Based on learning styles:*

a) FSLSM.: The Felder-Silverman Learning Styles Model (FSLSM) is employed by some assessment tools to provide the appropriate items according to the student's capabilities and preferences based on learning styles that are divided into four dimensions, namely, processing, perceiving, inputting, and understanding [40]. According to a recent survey conducted by Nabizadeh et al. [71], FSLSM is the model most frequently adopted by many LMSs to understand the student's preferred learning style and tailor evaluations more closely to their particular skills and preferences, which can help to improve learning outcomes. It is important to note, however, that the use of learning styles in assessment is somewhat debated, and some researchers such as Abyaa et al. [72] and Kirschner [73], have advocated that there is limited proof to sustain the idea that adapting education to fit individual learning styles truly enhances learning outcomes.

5) Other techniques:

a) Revised Bloom Taxonomy (RBT): Bloom's Revised Taxonomy (RBT) is a framework used to articulate and classify learning objectives for assessment. It helps in defining and organizing what students are expected to learn as a result of instruction, guiding the creation of tests that align with these learning goals. It produces a classification of learning goals organized into six levels: Remember, Understand, Apply, Analyze, Create and Evaluate. Many searchers have implemented the RBT to develop effective and efficient assessments that reflect the six levels. It explores hierarchical cognitive processes, through a mix of question types so that the system can choose the proper one for each level to adapt its assessments. It generates three lower-level knowledge questions and adjusts their difficulty based on the student's background from previous tests [74]. For example, if the student has scored low on the assessments, the system selects items of average to medium difficulty, while if the student is more experienced, questions of high difficulty are chosen. It is necessary to learn the knowledge and skills of the previous level to progress to a higher level of knowledge. Finally, adaptive assessment using RBT has many strengths such as easy detection of students' deficiencies and a gradual rise in question difficulty to revise the whole course content and better structure the learning process to enhance students' new abilities.

b) Game-based digital assessment (GBDA): is regarded as one of the pivotal approaches to stimulating authentic and accurate behavioral outcomes by conducting a stealth assessment [75], it can also be used to screen for reading difficulties with less time and cost, while enabling the content of educational games to be tailored to individual learners [76].Furthermore, Alonso-Fernandez et al. [77] used gameplay traces to assess the increase in awareness (affective dimension) as the difference between the post-test mean score and the pretest mean score for each player.

D. Knowledge Assessment Framework to Generate Adaptive Learning

Among the assessment tools studied, such as PIAT [40], AskMe [78], and ASSA [79], most follow a framework similar to the one depicted in Fig. 6 for monitoring students' progress. This framework employs a loop strategy. In the initial step, a test estimates the learner's current knowledge. Based on the test results, the system suggests appropriate curriculum materials with difficulty levels that align with the learner's predicted knowledge. To address any gaps identified by the tests, this loop is repeated until the system determines that the learner has acquired sufficient skills to complete a topic. Thus, each student engages in a series of tasks that are dynamically generated based on their responses. These tasks and items, stored in a database, represent all possible knowledge levels aligned with the content.

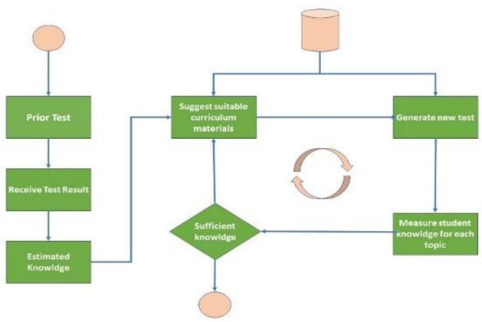

Fig. 6. A framework schematic representation of an adaptive assessment.

E. Recommendations

The future of adaptive assessment is full of promise regarding individualization, precision, and the incorporation of different learning modes into the design of the assessment. It is evident that technology such as big data and AI are reshaping the future of learning assessment. The present adaptive assessment model is gradually becoming obsolete as teachers and learners adopt intelligent solutions to enhance their testing experience. These solutions have the potential to enhance engagement and accessibility in the assessment process. This paper will explore potential future directions for designing the next generation of adaptive assessments, considering recent technological advancements.

First, AI provides a variety of new tools and technologies that can help to enhance engagement, such as gamification, conversational AI, or virtual and augmented reality. These tools can help make the testing process more engaging and enjoyable, by encouraging students to participate earnestly in the assessment and evaluation process. These technologies can help to assess students on real-world complex concepts and skills, such as engineering tasks, surgery, and aviation, and can provide assessment tools that can interact with humans through natural language such as ChatGPT.

Second, technology-driven approaches can help in removing obstacles by allowing students to take tests regardless of their geographic location. In addition, the inclusion of rich media features, such as videos, interactive simulations, and interactive games, can be particularly helpful for students with learning disabilities, who may have trouble with real-world assessment models by benefiting from assistive technologies that help them to break down barriers to assessment.

Third, IoT devices can be employed to capture a student's attention, which is essential in learning assessment. This will facilitate capturing student behaviors in online learning assessment strategies.

Fourth, intelligent assessment should involve tools/mechanisms to identify cheating, plagiarism as well as when learners are memorizing the answers to assignments.

In addition, the next generation of adaptive assessments should meet the following characteristics:

- Capable of accommodating thousands, if not millions, of students taking tests simultaneously, so all schools and universities must be equipped with a sophisticated technology infrastructure to enable computer-based adaptive assessment at an accessible cost.
- Provide large databases of items in all domains, these items must be scalable and can be fitted into any adaptive learning system based on universal open standards,
- Generalizable to other fields outside of science, technology, engineering, and mathematics (STEM) disciplines to cover other areas such as business, literature, education, arts, and humanities.
- Be able to motivate learners to assume more control and responsibility on assessment tests.
- Prioritize the development of digital environments that are secured, and with transparent policies describing the usage and protection of data from further unauthorized or abusive access.
- Strive to ensure equity among students when subjected to different assessments.
- Use adaptive assessment data to sustain an educational norm and to inform the development of policy.
- Improve the new skills of all educational stakeholders, such as digital literacy, as assessment processes are challenged by a multiform dimension that is not restricted to writing or reading words and demands new IT competencies.
- Collecting feedback from all stakeholders enables continuous refinement and improvement of assessment tools, resulting in a more focused, tailored, and favorable testing environment.

As the world becomes increasingly digital, further innovations in this area are anticipated, aiming to make assessments for learning more accessible, engaging, and effective for all learners. This progression may lead to a transition from traditional adaptive assessments to what could be termed "Deep Assessment".

V. Discussion

In this systematic review, we examined several key aspects of adaptive assessment: (1) the primary algorithms utilized in adaptive learning systems (ALS); (2) the effectiveness, strengths, and functioning of these assessment methods; and (3) a comparative analysis of these models, including a comprehensive summary of prominent algorithms and techniques for assessing learner knowledge, which led to the development of a new taxonomy. Our review revealed that different algorithms are used to track and evaluate student knowledge, each possessing distinct characteristics and capabilities.

The primary objective of this review is to synthesize the various models and theoretical frameworks that influence the effectiveness of adaptive assessment in educational practices. Our findings indicate that IRT and DKT are among the most significant algorithms used in adaptive assessment systems.

IRT, an algorithm with origins dating back to Rasch and Lord in the 1950s, remains one of the most widely used theories in adaptive assessment. Based on a logistic model, IRT has endured over the years through continuous development and enhancement by researchers. Various IRT models, such as the Rasch model, 1PL, 2PL, and GRM, have evolved into powerful tools for adaptive learning systems. This ongoing development has made IRT increasingly robust, offering an optimal solution for adaptive learning systems and ensuring its relevance and effectiveness in modern educational practices. Numerous educational systems and tools implement IRT, including SIETTE [80], CONCERTO [81], PASS [40], YIXUE [82], APelS [83], Persofit [84], eDia [18], The MISTRAL [85], and ALEAS [86].

In contrast, Deep Knowledge Tracing (DKT) is a very recent algorithm introduced in 2015 that benefits significantly from advancements in artificial intelligence. DKT utilizes deep recurrent neural networks (RNNs and LSTMs) to model student learning and trace their knowledge. By capturing complex knowledge representations over time, DKT significantly enhances prediction performance across assessment datasets, leveraging the growth of AI to provide cutting-edge solutions for adaptive learning. However, our review did not find any ALS currently implementing DKT. This absence can be attributed to the recency of the algorithm and its ongoing development, highlighting an area for future research and potential application.

Our systematic review has several unique advantages and particularities compared to other reviews in the field. While several researchers have examined the use of assessment in adaptive learning systems, our review integrates and builds upon these studies to offer a more comprehensive perspective. For example, Wei [8] explored multiple types of assessment instruments and approaches beneficial for teachers and tutors, and Nikou and Economides [19] reviewed mobile-based assessment across major educational technology research journals. Shute and Rahimi [9] focused on computer-based assessment for learning in elementary and secondary education, highlighting the potential for integrating instruction and assessment. Xiong and Suen [10] examined assessment approaches for open online education from both formative and summative perspectives, and Goss [87] reviewed student learning outcomes in higher education and academic libraries. Additionally, Moris et al. [5] demonstrated the value of formative assessment and feedback in higher education.

Our review offers several distinct advantages. First, it provides comprehensive integration by combining findings from multiple studies, resulting in a holistic overview that includes various types of adaptive assessment methods, models, and theoretical frameworks. Second, we develop a new taxonomy of prominent algorithms and techniques, offering a structured and detailed classification that can guide future research and applications in adaptive assessment. Third, our review spans multiple disciplines, ensuring that the findings are applicable across various educational contexts and not limited to a single field.

In conclusion, the evolution of adaptive assessment technologies, from the longstanding and continually improving IRT models to the innovative and AI-driven DKT, highlights the dynamic nature of this field. As educational practices increasingly incorporate these advanced algorithms, future research should focus on integrating the strengths of both traditional and modern approaches to further enhance the precision, adaptability, and effectiveness of adaptive learning systems.

VI. Conclusion and Future Research

This article has shed light on the essential role played by assessment, in particular adaptive assessment, in the implementation and progress of ALS. The COVID-19 pandemic has had a lasting impact on education [88], and it has proven that ALS will no longer be just an add-on, a nice thing to use, but will forever be a fundamental, essential part of teaching, learning and assessment. Therefore, the assessment of learning must change to prepare for a world more deeply infused with smart technologies and a mode of teaching that tends towards distance and personalization. However, there are many hurdles, limitations and barriers that hinder easy and smooth integration and use of technology in knowledge assessment such as system gaming which is defined as an attempt to pass a question or task by systematically taking advantage of the properties and regularities of the system rather than thinking about the test. However, there are several directions for further and highly evolving research in adaptive assessment that attempts to produce innovative solutions and smart algorithms to address these barriers to build trust in technology to make a fair and adaptive assessment capable of measuring a learner's performance and improving their educational experience. In addition, there is a need to develop research on the use of simulations and digital games in assessment and to develop methodologies or tools to assess learners in a mobile learning environment.

This work offers an initial exploration of adaptive assessment, aiming to provide a deeper understanding of its mechanisms, benefits, types, and the various techniques and algorithms used in its implementation. While not exhaustive, this article serves as a foundational reference and is intended to be a work in progress. Researchers are invited to contribute to its further development and refinement. Overall, this article suggests that adaptive assessment remains a promising field for measuring student knowledge in the right place, at the right time, and in the right form. However, it is important to approach the assessment of written production with caution in existing adaptive assessment methods. To fully address areas beyond

STEM, such as social-emotional learning, critical thinking, creativity, and executive functions, these assessments must be adapted and extended accordingly. To sum up, the study shows that assessments, especially adaptive ones, take a large place in adaptive learning environments. Furthermore, this review can be adopted and reused as a guideline to develop new and more sophisticated assessment models.

Given the rapid advancements in AI and the emerging era of big data, which are driving the evolution of adaptive learning systems (ALS) [89], Our future research is to enhance DKT by integrating the strengths of other models studied in this article and applying our proposed framework.

ACKNOWLEDGMENT

This work was supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA) and the CNRST of Morocco under Project No. 451/2020 (Smart Learning).

REFERENCES

[1] V. H. M. Dale and J. Singer, “Learner experiences of a blended course incorporating a mooc on haskell functional programming,” *Research in Learning Technology*, vol. 27, no. 0, Jul 2019. [Online]. Available: https://journal.alt.ac.uk/index.php/rlt/article/view/2248

[2] K. Wauters, P. Desmet, and W. Van den Noortgate, “Adaptive item-based learning environments based on the item response theory: possibilities and challenges: Adaptive itss based on irt,” *Journal of Computer Assisted Learning*, vol. 26, no. 6, p. 549–562, Dec 2010.

[3] H. S. Alenezi and M. H. Faisal, “Utilizing crowdsourcing and machine learning in education: Literature review,” *Education and Information Technologies*, vol. 25, no. 4, pp. 2971–2986, 2020.

[4] D. Nusche, L. Earl, W. Maxwell, and C. Shewbridge, *OECD Reviews of Evaluation and Assessment in Education: Norway 2011*, ser. OECD Reviews of Evaluation and Assessment in Education. OECD, Oct 2011.

[5] Morris R, Perry T, Wardle L (2021) Formative assessment and feedback for learning in higher education: A systematic review. Review of Education 9(3). https://doi.org/10.1002/rev3.3292,

[6] Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & education*, *57*(4), 2333-2351.

[7] H. Li, W. Cui, Z. Xu, Z. Zhu, and M. Feng, “Yixue adaptive learning system and its promise on improving student learning.” in *CSEDU (2)*, 2018, p. 45–52.

[8] X. Wei, “Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: A systematic literature review,” *Computers & Education*, p. 24, 2021.

[9] V. Shute and S. Rahimi, “Review of computer-based assessment for learning in elementary and secondary education: Computer-based assessment for learning,” *Journal of Computer Assisted Learning*, vol. 33, no. 1, p. 1–19, Feb 2017.

[10] Y. Xiong and H. K. Suen, “Assessment approaches in massive open online courses: Possibilities, challenges and future directions,” *International Review of Education*, vol. 64, no. 2, p. 241–263, Apr 2018.

[11] E. O’Donnell, S. Lawless, M. Sharp, and V. P. Wade, “A review of personalised e-learning: Towards supporting learner diversity,” *International Journal of Distance Education Technologies (IJDET)*, vol. 13, no. 1, p. 22–47, 2015.

[12] H. M. Truong, “Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities,” *Computers in Human Behavior*, vol. 55, p. 1185–1193, Feb 2016.

[13] N. Morze, L. Varchenko-Trotsenko, T. Terletska, and E. SmyrnovaTrybulska, “Implementation of adaptive learning at higher education institutions by means of moodle lms,” *Journal of Physics: Conference Series*, vol. 1840, no. 1, p. 012062, Mar 2021.

[14] B. Vesin, K. Mangaroska, and M. Giannakos, “Learning in smart environments: user-centered design and analytics of an adaptive learning system,” *Smart Learning Environments*, vol. 5, no. 1, p. 24, Dec 2018.

[15] L. Resnick, *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Routledge, 2018.

[16] P. Brusilovsky, C. Karagiannidis, and D. Sampson, “Layered evaluation of adaptive learning systems,” *International Journal of Continuing Engineering Education and Lifelong Learning*, vol. 14, no. 4/5, p. 402, 2004.

[17] M. Mohammad Bagheri, “Intelligent and adaptive tutoring systems: How to integrate learners,” *International Journal of Education*, vol. 7, no. 2, p. 1, Apr 2015.

[18] B. Csapo and G. Moln´ ar, “Online diagnostic assessment in support´ of personalized teaching and learning: The edia system,” *Frontiers in Psychology*, vol. 10, p. 1522, Jul 2019.

[19] S. A. Nikou and A. A. Economides, “Mobile-based assessment: A literature review of publications in major referred journals from 2009 to 2018,” *Computers & Education*, vol. 125, p. 101–119, Oct 2018.

[20] V. J. Shute and Y. J. Kim, “Formative and stealth assessment,” *Handbook of research on educational communications and technology*, pp. 311–321, 2014.

[21] M. Boussakuk, A. Bouchboua, M. El Ghazi, M. El Bekkali, and M. Fattah, “Design of computerized adaptive testing module into our dynamic adaptive hypermedia system,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 18, p. 113, Sep 2021.

[22] K. Shephard, “Higher education for sustainability: seeking affective learning outcomes,” *International Journal of Sustainability in Higher Education*, vol. 9, no. 1, p. 87–98, Jan 2008.

[23] R. Dabrowski, J. W. LeLoup, and L. E. MacDonald, “Effectiveness of computer-graded vs. instructor-graded homework assignments in an elementary spanish course: A comparative study at two undergraduate institutions,” *IALLT Journal of Language Learning Technologies*, vol. 43, no. 1, p. 78–100, 2013.

[24] L. He and P. Brandt, “Weas: a web-based educational assessment system,” in *Proceedings of the 45th annual southeast regional conference on - ACM-SE 45*. Winston-Salem, North Carolina: ACM Press, 2007, p. 126.

[25] C.-h. Chen, “The implementation and evaluation of a mobile self- and peer-assessment system,” *Computers & Education*, vol. 55, no. 1, p. 229–236, Aug 2010.

[26] M. Ward, L. Gruppen, and G. Regehr, “Measuring self-assessment: current state of the art,” *Advances in health sciences education*, vol. 7, pp. 63–80, 2002.

[27] L. Cheniti Belcadhi, “Personalized feedback for self assessment in lifelong learning environments based on semantic web,” *Computers in Human Behavior*, vol. 55, p. 562–570, Feb 2016.

[28] E. Panadero and M. Alqassab, “An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading,” *Assessment & Evaluation in Higher Education*, vol. 44, no. 8, p. 1253–1278, Nov 2019.

[29] J. Lu and Z. Zhang, “Understanding the effectiveness of online peer assessment: A path model,” *Journal of Educational Computing Research*, vol. 46, no. 3, pp. 313–333, 2012.

[30] L. Li, “The role of anonymity in peer assessment,” *Assessment &* Evaluation *in Higher Education*, vol. 42, no. 4, p. 645–656, May 2017.

[31] S. Kalyuga, “Enhancing instructional efficiency of interactive e-learning environments: A cognitive load perspective,” *Educ Psychol Rev*, p. 13, 2007.

[32] J. Liu, “The assessment agent system: design, development, and evaluation,” *Educational Technology Research and Development*, vol. 61, no. 2, p. 197–215, Apr 2013.

[33] E. Gouli, K. Papanikolaou, and M. Grigoriadou, “Personalizing assessment in adaptive educational hypermedia systems,” in *Adaptive Hypermedia and Adaptive Web-Based Systems: Second International Conference, AH 2002 Malaga, Spain, May 29–31, 2002 Proceedings 2´* . Springer, 2002, pp. 153–163.

[34] F. Lazarinis, S. Green, and E. Pearson, “Creating personalized assessments based on learner knowledge and objectives in a hypermedia

web testing application," *Computers & Education*, vol. 55, no. 4, p. 1732–1743, Dec 2010.

[35] J.-J. Navarro and C. V. Mourgues-Codern, "Dynamic assessment and computerized adaptive tests in reading processes," *Journal of Cognitive Education and Psychology*, vol. 17, no. 1, p. 70–96, 2018.

[36] T. Lendyuk, S. Rippa, and S. Sachenko, "Simulation of computer adaptive learning and improved algorithm of pyramidal testing," in *2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*. Berlin, Germany: IEEE, Sep 2013, p. 764–769. [Online].

[37] E. Guzman and R. Conejo, "Self-assessment in a feasible, adaptive webbased testing system," *IEEE Transactions on Education*, vol. 48, no. 4, p. 688–695, 2005.

[38] E. Gvozdenko and D. Chambers, "Beyond test accuracy: Benefits of measuring response time in computerised testing," *Australasian Journal of Educational Technology*, vol. 23, no. 4, Oct 2007. [Online]. Available: http://ajet.org.au/index.php/AJET/article/view/1251

[39] M. Kosinski and J. Rust, "The development of concerto: An opensourceonline adaptive testing platform," *Paper presenetd at the International Association for Computerized Adaptive Testing, Pacific Grove, CA*, 2011.

[40] L. Al-Rajhi, R. Salama, and S. Gamalel-Din, "Personalized intelligent assessment model for measuring initial students' abilities," in *Proceedings of the 2014 Workshop on Interaction Design in Educational Environments - IDEE '14*. Albacete, Spain: ACM Press, 2014

[41] R. Conejo, E. Guzman, J.-L. P´ erez-De-La-Cruz, and E. Mill´ an, "In-´ troducing adaptive assistance in adaptive testing," *Journal of Artificial Intelligence in Education*, p. 3, 2005.

[42] Y. Choi and C. McClenen, "Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks," *Applied Sciences*, vol. 10, no. 22, p. 8196, Nov 2020.

[43] W. J. Van der Linden, *Handbook of item response theory: Three volume set*. CRC Press, 2018.

[44] J.-C. Falmagne, D. Albert, C. Doble, D. Eppstein, and X. Hu, Knowledge *spaces: Applications in education*. Springer Science & Business Media, 2013.

[45] J.-P. Doignon and J.-C. Falmagne, "Knowledge spaces and learning spaces," *the California Digital Library*, 2016.

[46] R. D. Arasasingham, M. Taagepera, F. Potter, and S. Lonjers, "Using knowledge space theory to assess student understanding of stoichiometry," *Journal of Chemical Education*, vol. 81, no. 10, p. 1517, Oct 2004.

[47] C. Doble, J. Matayoshi, E. Cosyn, H. Uzun, and A. Karami, "A databased simulation study of reliability for an adaptive assessment based on knowledge space theory," *International Journal of Artificial Intelligence in Education*, vol. 29, no. 2, p. 258–282, May 2019.

[48] Y. Fang, Z. Ren, X. Hu, and A. C. Graesser, "A meta-analysis of the effectiveness of aleks on learning," *Educational Psychology*, vol. 39, no. 10, p. 1278–1292, Nov 2019.

[49] J. Pearl, "Bayesian networks," *the California Digital Library*, 2011.

[50] M. J. Culbertson, "Bayesian networks in educational assessment: The state of the field," *Applied Psychological Measurement*, vol. 40, no. 1, p. 3–21, Jan 2016.

[51] H. J. Henriksen and H. C. Barlebo, "Reflections on the use of bayesian belief networks for adaptive management," *Journal of Environmental Management*, vol. 88, no. 4, p. 1025–1036, Sep 2008.

[52] J. A. Collins, J. E. Greer, and S. X. Huang, "Adaptive assessment using granularity hierarchies and bayesian nets," *Lecture Notes in Computer Science*, vol. 1086, pp. 569–577, 1996.

[53] W. Xing, C. Li, G. Chen, X. Huang, J. Chao, J. Massicotte, and C. Xie, "Automatic assessment of students' engineering design performance using a bayesian network model," *Journal of Educational Computing Research*, vol. 59, no. 2, pp. 230–256, 2021.

[54] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modelling and UserAdapted Interaction*, vol. 4, no. 4, p. 253–278, 1995.

[55] Z. A. Pardos and N. T. Heffernan, "Using hmms and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset," *Journal of Machine Learning Research W & CP*, vol. 40, 2010.

[56] S. Schultz and T. Tabor, "Revisiting and extending the item difficulty effect model," in *In Proceedings of the 1st workshop on massive open online courses at the 16th annual conference on artificial intelligence in education*.Citeseer, 2013, p. 33–40.

[57] Q. Liu, S. Shen, Z. Huang, E. Chen, and Y. Zheng, "A survey of knowledge tracing," *arXiv preprint arXiv:2105.15106*, 2021.

[58] P. I. Pavlik Jr, H. Cen, and K. R. Koedinger, "Performance factors analysis–a new alternative to knowledge tracing." *Online Submission*, 2009.

[59] C. Maier, R. S. Baker, and S. Stalzer, "Challenges to applying performance factor analysis to existing learning systems," in *Proceedings of the 29th International Conference on Computers in Education*, 2021.

[60] J.-J. Vie and H. Kashima, "Knowledge tracing machines: Factorization machines for knowledge tracing," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, p. 750–757, Jul 2019.

[61] B. Cope and M. Kalantzis, "Big data comes to school: Implications for learning, assessment, and research," AERA Open, vol. 2, no. 2, p. 233285841664190, Apr 2016.

[62] Y. Chen and H.-H. Chang, "Psychometrics help learning: From assessment to learning," Applied Psychological Measurement, vol. 42, no. 1, p. 3–4, Jan 2018.

[63] D. Ifenthaler, Automated Essay Scoring Systems. Singapore: Springer Nature Singapore, 2022, p. 1–15.

[64] D. Kucak, V. Juricic, and G. DambiC, "Machine learning in education-a survey of current research trends." Annals of DAAAM & Proceedings, vol. 29, 2018.

[65] Z. Ke and V. Ng, "Automated essay scoring: A survey of the state of the art." in IJCAI, vol. 19, 2019, p. 6300–6308.

[66] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," Advances in neural information processing systems, vol. 28, 2015.

[67] M. Samarakou, P. Prentakis, D. Mitsoudis, D. Karolidis, and S. Athinaios, "Application of fuzzy logic for the assessment of engineering students," in 2017 IEEE global engineering education conference (EDUCON). IEEE, 2017, p. 646–650.

[68] Z. Jeremic, J. Jovanovi´ c, and D. Ga´ sevi˘ c, "Student modeling and´ assessment in intelligent tutoring of software patterns," Expert Systems with Applications, vol. 39, no. 1, p. 210–222, Jan 2012.

[69] R. Sripan and B. Suksawat, "Propose of fuzzy logic-based students' learning assessment," in ICCAS 2010. Gyeonggido: IEEE, Oct 2010, p. 414–417. [Online].

[70] K. Chrysafiadi, C. Troussas, and M. Virvou, "Combination of fuzzy and cognitive theories for adaptive e-assessment," *Expert Systems with Applications*, vol. 161, p. 113614, Dec 2020.

[71] A. H. Nabizadeh, J. P. Leal, H. N. Rafsanjani, and R. R. Shah, "Learning path personalization and recommendation methods: A survey of the state-of-the-art," Expert Systems with Applications, vol. 159, p. 113596, 2020.

[72] A. Abyaa, M. Khalidi Idrissi, and S. Bennani, "Learner modelling: systematic review of the literature from the last 5 years," Educational Technology Research and Development, vol. 67, no. 5, p. 1105–1143, Oct 2019.

[73] P. A. Kirschner, "Stop propagating the learning styles myth," Computers & Education, vol. 106, pp. 166–171, 2017.

[74] A. Krouska, C. Troussas, and M. Virvou, "Computerized adaptive assessment using accumulative learning activities based on revised bloom's taxonomy," in Knowledge-Based Software Engineering: 2018: Proceedings of the 12th Joint Conference on Knowledge-Based Software Engineering (JCKBSE 2018) Corfu, Greece 12. Springer, 2019, pp. 252–258.

[75] S. Zhu, Q. Guo, and H. H. Yang, "Beyond the traditional: A systematic review of digital game-based assessment for students' knowledge, skills, and affections," Sustainability, vol. 15, no. 5, p. 4693, Mar 2023.

[76] J. Hautala, R. Heikkila, L. Nieminen, V. Rantanen, J.-M. Latvala,¨ and U. Richardson, "Identification of reading difficulties by a digital game-based assessment technology," Journal of Educational Computing Research, vol. 58, no. 5, pp. 1003–1028, 2020.

[77] C. Alonso-Fernandez, A. Calvo-Morata, M. Freire, I. Martınez-Ortiz, and B. Fernandez-Manj´ on, "Evidence-based evaluation of a serious´ game to increase bullying awareness," Interactive Learning Environments, vol. 31, no. 2, p. 644–654, Feb 2023.

[78] C. Saul and H.-D. Wuttke, "An adaptation model for personalized e-assessments," International Journal of Emerging Technologies in Learning (iJET), vol. 8, no. S2, p. 5, Jun 2013.

[79] D. A. Aljohany, R. Mohamed, and M. Saleh, "Assa: Adaptive e-learning smart students assessment model," International Journal of Advanced Computer Science and Applications, vol. 9, no. 7, 2018.

[80] Conejo R, Barros B, Bertoa MF (2019) Automated assessment of complex programming tasks using siette. IEEE Transactions on Learning Technologies 12(4):470–484.https://doi.org/10.1109/TLT.2018.2876249

[81] Magis D, Raˆıche G (2012) Random generation of response patterns under computerized adaptive testing with the r package catr. Journal of Statistical Software 48(8). https://doi.org/10.18637/jss.v048.i08,

[82] Feng M, Cui W, Wang S (2018) Adaptive learning goes to china. In: Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part II 19, Springer, pp 89–93

[83] Liu HI, Yang MN (2005) Qol guaranteed adaptation and personalization in e-learning systems. IEEE Transactions on Education 48(4):676–687

[84] Sodoké K, Raˆıche G, Nkambou R (2007) The adaptive and intelligent testing framework: Personfit. In: Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007). IEEE, p 715–717

[85] Salcedo P, Pinninghoff MA, Contreras R (2005) Computerized adaptive tests and item response theory on a distance education platform. In: Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach: First International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2005, Las Palmas, Canary Islands, Spain, June 15-18, 2005, Proceedings, Part II 1, Springer, pp 613–621

[86] Davino C, Fabbricatore R, Pacella D, et al (2020) Aleas: a tutoring system for teachingand assessing statistical knowledge. In: PSYCHOBIT

[87] Goss H (2022) Student learning outcomes assessment in higher education and in academic libraries: A review of the literature. The Journal of Academic Librarianship 48(2):102,485

[88] WEF, "The covid-19 pandemic has changed education forever. this is how," Apr 2020, accessed 31 january 2023. [Online]. Available: https://www.weforum.org/agenda/2020/04/coronavirus-educationglobal-covid19-online-digital-learning/

[89] Gardner J, O'Leary M, Yuan L (2021) Artificial intelligence in educational assessment: 'breakthrough? or buncombe and ballyhoo?'. Journal of Computer Assisted Learning 37(5):1207–1216. https://doi.org/10.1111/jcal.12577