IJRCS
International Journal of Robotics and Control Systems
Vol. 4, No. 4, 2024, pp. 1656-1693
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs
ASCEE

A Systematic Review of the Use of Technology in Educational Assessment Practices: Lesson Learned and Direction for Future Studies

Heri Retnawati [a,e,1*], Elena Kardanova [b,2], Sumaryanto [c,3], Lantip Diat Prasojo [d,4], Jailani [a,5], Elly Arliani [a,6], Kana Hidayati [a,7], Mathilda Susanti [a,8], Himmawati Puji Lestari [a,9], Ezi Apino [e,10], Ibnu Rafi [a,11], Munaya Nikma Rosyada [a,12], Rugaya Tuanaya [e,13], Septinda Rima Dewanti [f,g,14], Rimajon Sotlikova [h,15], Gulzhaina Kuralbayevna Kassymova [i,j,16]

[a] Department of Mathematics Education, Universitas Negeri Yogyakarta, Sleman, Indonesia
[b] Center for Psychometrics and Measurement in Education, HSE University, Moscow, Russia
[c] Department of Sport Sciences, Universitas Negeri Yogyakarta, Sleman, Indonesia
[d] Department of Educational Administration, Universitas Negeri Yogyakarta, Sleman, Indonesia
[e] Department of Educational Research and Evaluation, Universitas Negeri Yogyakarta, Sleman, Indonesia
[f] School of Early Childhood and Inclusive Education, Queensland University of Technology, Brisbane, Australia
[g] Department of Guidance and Counseling, Universitas Negeri Yogyakarta, Sleman, Indonesia
[h] Master of Arts in TESOL Program, Webster University in Tashkent, Tashkent, Uzbekistan
[i] Institute of Pedagogy and Psychology, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[j] Educational Leadership Program, Queen Margaret University, Edinburgh, United Kingdom
[1] heri_retnawati@uny.ac.id; [2] ekardanova@hse.ru; [3] sumaryanto@uny.ac.id; [4] lantip@uny.ac.id; [5] jailani@uny.ac.id;
[6] arlianielly@uny.ac.id; [7] kana@uny.ac.id; [8] mathilda@uny.ac.id; [9] himmawati@uny.ac.id,
[10] eziapino.2021@student.uny.ac.id; [11] ibnu.rafi@hotmail.com; [12] munayanikma38@gmail.com;
[13] rinanaya74@yahoo.com; [14] septindarima.dewanti@hdr.qut.edu.au; [15] rima.sotlikova@gmail.com;
[16] g.kassymova@abaiuniversity.edu.kz
* Corresponding Author

ARTICLE INFO

Article history
Received August 15, 2024
Revised September 21, 2024
Accepted October 09, 2024

Keywords
Use of Technology;
Test Administration;
Giving Feedback;
Test Scoring;
Test Item Generation;
Response Time Recording;
Educational Assessment;
Systematic Review

ABSTRACT

Previous studies have demonstrated that technology helps achieve learning outcomes. However, many studies focus on just one aspect of technology's role in educational assessment practices, leaving a gap in studies that examine how various aspects affect the use of technology in assessments. Hence, through a systematic work, we analyzed the extent and manner in which technology is integrated into educational assessments and how education level, domain of learning, and region may affect the use of technology. We reviewed empirical studies from two major databases (i.e., Scopus and ERIC) and a national journal whose focus and scope are on educational measurement and assessment, following PRISMA guidelines for systematic reviews. The findings of the present study are directed towards emphasizing the roles of technology in educational assessment practices and how these roles are adapted to varying educational contexts such as the level of education, the three domains of learning (i.e., cognitive, psychomotor, and affective), and the setting in which the assessment was conducted. These findings not only highlight the current roles of technology in educational assessment but also provide a roadmap for future research aimed at optimizing the integration of technology across diverse educational contexts.

1. Introduction

Educational assessment is an integral and essential part of the learning process as it provides information on student learning progress and achievement [1]-[5]. The information obtained from educational assessments allows teachers to evaluate the effectiveness of the learning methods they have implemented in the classroom and make sound data-driven decisions to improve learning. The objective of assessment is thus not only limited to identifying students' level of competence, but also to identify students' strengths and weaknesses that enable the design of learning methods or strategies in accordance with characteristics and needs of students. The emergence of digital technologies such as computers, mobile devices, the internet, and various applications has led to new opportunities to improve and change the way educational assessments are carried out. Technology has offered the potential to improve the efficiency, accuracy [6], and attractiveness of assessments and create quality assessments through ensuring quality of test or questionnaire items [7]-[12], as well as to make it easier to collect, analyze, and store assessment data and even to report assessment results [13]-[15].

Although it has been suggested that the use of technology in education or the learning process is not mandatory as it is not always helpful [16], [17], it has been believed that the appropriate use of technology can support the planning, implementation and evaluation of learning and support the achievement of various educational objectives. Extensive studies suggest that the extent to which technology is appropriately used tends to be relative to the type of educational technology that corresponds to the primary focus of that type of technology. Previous studies [18]-[24] have demonstrated the use of technology in the learning process, starting from planning, implementation, and assessment of learning. It has been shown that the use of technology in the learning process has been able to assist teachers in accomplishing their work related to organizing the learning process more easily, effectively, and efficiently [23], [25], [26].

Various types of educational assessments are available for teachers to use. Based on purpose, time of administration, construction, and design, assessments are divided into three types: diagnostic assessment, formative assessment, and summative assessment [27]. As the name implies, diagnostic assessment focuses on diagnosing the difficulties encountered by students including their strengths and weaknesses in learning a learning content or in mastering a competency that has been set in the learning objectives [27]-[29]. The results that teachers obtain from diagnostic assessments can thus provide more detailed information on student characteristics so that they can design learning more specialized and personalized according to the needs of each student [4], [27], [29]. As for formative assessment, which based on the purpose of assessment is also referred to as assessment for learning, it provides information related to the extent to which students understand the learning, in which direction students will learn and how they will get there [27]. The implementation of assessment for learning is integrated with learning activities, where the results obtained from the assessment are used for learning improvement and evaluation of the learning process that teachers have facilitated and learners have engaged in. Teachers who are able to utilize the results of assessment for learning are expected to maintain students' learning pace, provide adequate assistance for students who experience more challenges in achieving learning objectives, and improve the quality of learning. The existence of technology that has offered various features, such as Kahoot! and GeoGebra, is possible to assist teachers in conducting formative assessments.

The utilization of technology in formative assessment can be carried out in various ways to increase the effectiveness and efficiency of the assessment process. An instance of such technology utilization is the use of online platforms or applications that allow teachers to digitally administer questionnaire, quizzes, tests, or assignments [30]-[32]. The use of these technologies allows students to access and take the assessments through electronic devices such as computers, laptops, or smartphones [32]. Thus, teachers can conveniently monitor students' learning progress in real-time and provide quick and specific feedback. Technology also enables the use of various interactive and engaging assessment formats, such as gamification, simulation, or the use of multimedia [33]-[35]. This may increase student motivation and engagement in the assessment process. In addition, technology can assist in collecting and analyzing assessment data automatically [36], [37], allowing teachers to immediately identify areas that need special attention and adjust learning strategies according to student's needs. The same thing related to the potential utilization of technology also holds true for summative assessments. This assessment, carried out at the end of a unit of learning, provides information on the extent to which students have achieved the learning objectives. Summative assessments on a small scale vary in terms of their use, ranging from assessing critical thinking skills [38], evaluating the learning practices that teachers facilitate for their students, to making adjustments to better support student learning [1], [2], [39], [40]. While at a broader level, summative assessments can be conducted to obtain information that can be used for the purposes of mapping learner achievements across schools and regions.

A number of studies conducted through a comprehensive systematic review approach, such as [41], have investigated the role of technology in learning assessment practices. It has been provided evidence of the diverse impact of technology in assessment that points to the benefits offered by such technology such as increased student engagement and accommodating concerns about potential cheating in test-taking and the increased workload that teachers face in carrying out different types of assessments. Not only emphasizing the potential of using technology in assessment, other studies, such as [42], have also highlighted the importance of effective integration of technology and support from educational institutions for the use of technology in assessment practices. Mangaroska et al. [43] provide a comprehensive overview of assessment in technology-rich environments, focusing on diverse situations such as adaptive learning, data-driven approaches, and game-based learning. In addition, Lin [44] has explored different assessment tools for science, technology, engineering, and mathematics (STEM) education contexts. It is clear that extensive studies have been conducted focusing on the implementation of technology-supported educational assessment practices.

Again, given the large number of existing studies that aim to utilize technology in a variety of educational assessment practices, we have identified a number of studies that were conducted using systematic or non-systematic review approaches with the aim of examining the various uses of technology in assessment. Through a non-systematic review, a variety of possible uses of technology in educational assessment practices in early childhood education were identified, leading to a number of recommendations on the design, development and implementation of assessments that could possibly promote student learning [27]. Various roles offered by technology in supporting educational assessment practices in early childhood education level include developing test or non-test items; providing support for validity evidence, reliability estimates, and item characteristics of measurement instruments through psychometric evaluation; administering tests; scoring tests; and test reporting and interpretation. In a systematic review of studies conducted in more than 30 countries, it has been reported that technology has been used in educational assessment practices that make it possible to improve learning and reduce teacher workloads that are not directly related to facilitating student learning [41]. These potentials are obtained when technology is used in activities or matters related to the assessment process, starting from before, during, and after the assessment. The studies on the role of technology in assessment included in the review by Chen and colleagues [41] focused more on the elementary to high school level. The technology used in the assessment practices in the studies they [41] have reviewed was used for various purposes, such as to administer tests, conduct formative assessments or assessments for learning that enable better learning, conduct test scoring, and report test results and their interpretation.

In a recently published study, Madland and colleagues [45] have reported the findings of their literature review on the use of technology in assessment practices in higher education. One of the focuses of their findings is how technology has played a role in supporting formative and summative assessment practices in higher educatio n with an emphasis on the benefits of and strategies for using such technology. Taking into account previous studies that have uncovered the use of technology in educational assessment practices through systematic or non-systematic review approaches, it still remains an opportunity for further exploration on the use of technology to support educational assessment practices especially by considering aspects that have received less extensive attention and exploration. The present study which uses the systematic review method is thus intended to investigate the use or role of technology in educational assessment practices by taking into account aspects that may be related to or influence the use of technology. We should recognize that technology comes in many forms. Based on the review of previous studies mentioned earlier, in this study, technology in educational assessment practices can include computer-based testing, learning management systems, data analysis and reporting tools, online survey tools, and interactive quiz tools. The systematic review method allowed us to obtain literature relevant to the role of technology in educational assessment practices based on a detailed and comprehensive search strategy and synthesize the findings of empirical studies presented in the literature [46]. The following are the research questions (RQs) that drive us to achieve the objective of the present study.

1. RQ1: To what extent has technology been used to support educational assessment practices?
    - RQ1a: What technologies are used in educational assessment practices?
    - RQ1b: To what extent are these technologies used in educational assessment practices?
2. RQ2: How does the role of technology in educational assessment practices relate to the level of education?
3. RQ3: How does the role of technology in educational assessment practices relate to the three domains in Bloom's taxonomy of learning (i.e., cognitive, psychomotor, and affective) that become the focus of assessment?
4. RQ4: How does the role of technology in educational assessment practices relate to the region in which the assessment took place?

Through providing answers to these research questions, this study is expected to contribute to the body of knowledge on how technology has been instrumental in supporting educational assessment practices. How these technologies are used to play certain roles in educational assessment practices brings lessons to be learned. In addition, the study also provides a glimpse of trends in the role of technology use in assessment practices across three aspects, namely the level of education, the three domains of learning, and the region by continent where assessment practices are conducted. Exploring the role of technology across different levels of education makes it possible to identify which roles of technology should be maximized in assessment practices while adapting to the characteristics of students at each level of education. Findings from the exploration of the role of technology across domains allow for further investigation into the reasons behind the use of technology that emphasizes one domain over another and which or what technologies should be used for particular assessment practices in particular domain. Meanwhile, the exploration of technology roles across different regions will provide an opportunity to further investigate the possible reasons why one technology role is more dominant than another in a particular region. The identification of the role of technology in supporting assessment practices across different educational levels, domains, and regions suggests an understudied area that can be explored by future studies.

2. Method

The systematic review conducted in this current study aims to investigate the role of technology in supporting assessment practices in education and to investigate how this role relates to the level of education, the three domains of learning as proposed by Bloom (i.e., cognitive, psychomotor, and affective), and the region in which educational assessment is carried out. This investigation was carried out by reviewing journal articles that reported the results of empirical studies. The findings of this review are expected to contribute to the provision of lessons learned by educators, policy makers, and researchers in the use of technology to support assessment practices and to provide direction for future studies in an area that has received little attention. In this section, we describe our strategy for searching, collecting, and selecting the articles that would ultimately be carefully reviewed and synthesized in this study to answer the research questions we raised in this study. All of these processes are carried out according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 [47].

2.1. Search Strategy

We performed a literature search of journal articles automatically by combining terms and keywords using two databases (i.e., Scopus and Education Resources Information Center (ERIC)) and manually in a national journal focusing on educational measurement and assessment. We used the Scopus database because it is the largest existing database that provides literature from a wide range of disciplines, including social sciences and education, offers enriched and comprehensive metadata records of the literature, and provides quality control assurance of the literature it includes [48], [49]. We also used ERIC because it is a database that provides a vast collection of literature primarily in education and the literature it includes is quality controlled through peer review [50], [51]. The terms and keywords we used to search the literature were selected based on the key types of educational assessment and the terms frequently associated with technology that we identified based on our experience so far from reading a large body of scientific literature in educational assessment and technology. We conducted a literature search on these two databases in April 2024. The search yielded 757 documents from the Scopus database, 1191 documents from the ERIC database, and 152 documents from the national journal database focusing on educational measurement and assessment. The results we obtained from this literature search are presented in the identification stage of the PRISMA flow diagram (see Fig. 1). Through initial screening, three duplicate documents were found from ERIC and one document was withdrawn from Scopus. This process resulted in 2096 documents ready for further screening. Details about the keywords or terms we used in the literature search including the Boolean operators we used to combine them, and the filtering aspects applied are presented in Table 1.

Furthermore, we determined the eligibility and exclusion criteria for documents that have passed the initial screening (see Table 2). The eligible studies must be journal articles in English published from 2020 to 2024 to ensure that the present review derives findings based on up-to-date literature. Additionally, we required full access to journal articles for an in-depth review of the article's contents. We ensure the scientific validity of the articles by selecting those that have undergone peer review before publication.

Table 1. Queries used on searching process

Database	Keywords Search and Other Applied Filters
Scopus	(TITLE-ABS-KEY((technology OR computer OR mobile OR digital) AND (based OR aided OR assisted) AND (summative OR formative OR feedback) AND (assessment OR evaluation)) AND PUBYEAR > 2018 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCTYPE,"ar")) AND (LIMIT-TO (LANGUAGE,"English")) AND (LIMIT-TO (SRCTYPE,"j")) AND (LIMIT-TO (PUBSTAGE,"final")) AND (LIMIT-TO (OA,"all")) AND (LIMIT-TO (SUBJAREA,"SOCI")))
ERIC	(e-assessment OR assess* OR evaluat* OR measur*) AND (technolog* OR web OR digital* OR computer OR laptop OR smartphone OR mobile OR online) (computer OR technology OR digital OR mobile OR smartphone OR web OR online) AND (based OR aided OR assisted) AND (assessment OR formative OR summative OR authentic)

2.2. Selection Process

The next process in this review is to select literature in such a way that the literature selected is the most relevant to the objectives of the current study. This process belongs to the screening stage in the PRISMA flow diagram (see Fig. 1). We conducted this process through data management that was assisted by Zotero reference manager software. We utilized the group libraries feature to determine the eligibility of each document based on predetermined criteria. We imported the database with digital object identifier (DOI) and performed the selection process in the software simultaneously. The selection process was done manually in three stages, namely reading the title, reading the abstract, then reading the content of each article. Eligibility of documents refers to the criteria that have been agreed upon, such as literature type, published time, language, availability, and stage of review (see Table 2). This process was conducted by three authors to increase judgment reliability. Although it was time-consuming, this process was carried out by three authors in the same place and time by scrutinizing the articles and judging whether they could be selected for further processing. The three authors discussed any discrepancy in judgment that arose to reach consensus on the selection of articles. The first stage resulted in 142 articles from Scopus, 228 articles from ERIC, and 25 articles from national journal, or a total of 395 articles ready for the next process. The next process is reading the abstract and full text. From this process, 226 articles were excluded from the document content reading process. The selection process yielded 169 articles to proceed in the coding process. PRISMA flow diagram for the selection of literature included in this study presented in Fig. 1.

Table 2. Eligibility and exclusion criteria

Criterion	Eligibility	Exclusion
Literature type	Journal article	Conference proceeding, book, systematic review articles
Published time	Since 2020 to 2024	Before 2020
Language	English	Non-English
Availability	Full-text available	Restricted access
Stage of review	Peer-reviewed only	

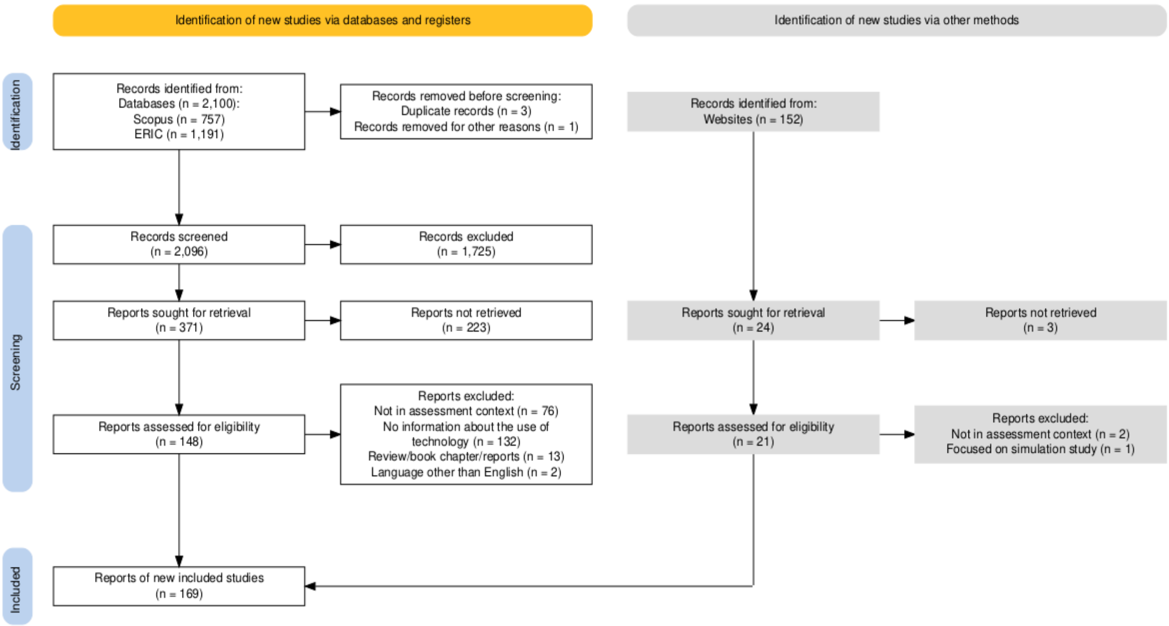

Fig. 1. PRISMA flow diagram for the selection of literature included in this study

2.3. Coding and Analysis

Coding was carried out on a licensed ATLAS.ti software. A total of three authors performed the coding process using the same license. The data that needed to be coded from each article included the level of education at which the technology-based assessment was held, the name of technology used for the assessment, the construct (i.e., skill, ability, concept, or competency) that becomes the focus of an assessment practice, the country in which the assessment is conducted, and the role of the technology in educational assessment practices. The assigned code related to the role of technology also reflects how technology has been used to support the assessment practice. We coded education levels into four categories by considering the possibility of a technology to be used in supporting educational assessment practices: kindergarten, elementary school, junior high school, senior high school, and higher education. We coded the role of technology use in assessment practice reported by journal articles using the possible roles offered by technology in assessment practice as identified in the study conducted by Neumann and colleagues [27]. As mentioned earlier, the technology roles in educational assessment practices that they have identified include providing support in item development, psychometric validation, test administration, test scoring, and test reporting and interpretation. However, we also opened up the possibility that there may be other roles for technology than those mentioned in [27]. Accordingly, in the present study, we used a deductive (top-down) approach [52] in coding the roles of technology to support assessment practices reported by the journal articles given that the coding was based on pre-defined codes, categories, or themes; and we are still open to new codes that may emerge.

We should acknowledge that the coding process was challenging as the three coders were required to detect codes contained in the literature that demonstrated the focus of this study. We frequently found a journal article that discussed more than one level of education, the name of the technology used, and the role of technology in assessment practice. Coders would certainly be expected to be more sensitive in grasping the codes that align best with the focus of the assessment practice that a journal article reports on. Although the three coders are reasonably familiar with educational assessment theory and practice, cross-validation of the codes that each coder identified was conducted to ensure accuracy and consistency. This cross-validation has detected a small number of discrepancies between a word, phrase, sentence, or paragraph and the code assigned to it. This cross-validation along with discussion among coders to reach consensus was done to support the consistency in the coding process among the three coders, instead of using a quantitative approach through inter-coder reliability using certain statistical measure such as Cohen's kappa as suggested by Krippendorff [53] and McHugh [54]. In order to assist coders in coding the role of technology, all coders were allowed to thoroughly read the literature and were facilitated to align their perceptions through discussion on a variety of possible activities or practices categorized into the five types of assessment practices as demonstrated in [27]. When a coder identified an activity or a practice that had little relevance to the five types of assessment practices, the three coders discussed to reach consensus on the possibility of introducing a new code related to the role of technology in educational assessment practices. This strategy also facilitated all coders in coding and making adjustments to the new codes across all journal articles.

After the coding was completed, the next step was to review the codes, categorize the codes based on themes that align with the research questions in this study, and generate visualizations. Since the coding demonstrated such a wide variety of countries, we opted to categorize the countries into their respective continents. The same was applied to the coding results of the constructs that become the focus of assessment practice. We categorized the constructs into three domains in Bloom's taxonomy of learning, namely cognitive, psychomotor, and affective. The three coders discussed until consensus was reached in assigning each construct to the appropriate domain. We visualized the findings on the role of technology in assessment practices and how it relates to levels of education, the three domains (i.e., cognitive, psychomotor, and affective) that are the focus of assessment, and the regions by continent where assessment took place in Sankey diagrams, respectively. Sankey diagrams enable us to observe the trend of the flow from one node in one layer to another node in another layer, where the flow has a certain thickness that indicates the magnitude or proportion of codes in a node to its corresponding node [55]. Through Sankey diagrams, we managed to provide a more detailed snapshot of trends regarding which technology roles received more and which received less attention based on education level, domains of the constructs focused on in assessment practices, and continent. Overall, the results obtained from the extraction and coding of the studies included in this study were analyzed and synthesized through a narrative approach by aligning themes that corresponded with the research questions of this study.

3. Results and Discussion

In the following section, the findings regarding the role of technology in educational assessment practices are described and discussed according to the research questions. By examining 169 journal articles reporting empirical studies, we first describe how technology has been used to support educational assessment practices. Afterwards, we report how the level of education, the three domains of learning, and the regions where the previous studies were conducted relate to the use of technology in educational assessment practices. Fig. 2 presents the identified roles of technology in educational assessment, which are test administration, giving feedback, test reporting and interpretation, response time recording, test item generation, test scoring, and psychometric evaluation. There is information about the frequency of each role of technology used in empirical studies. Upon further identification, several evaluations of psychometric properties of measurement instrument activities appear more than once, which depicted in Fig. 2.

Table 3 presents the technologies that previous studies have used to perform specific roles in educational assessment practices. The technologies listed include computer-based applications, online-based applications, computer software, and physical tools. References are included in the listed technology to facilitate easier access for researchers and readers who seek a more comprehensive understanding of the findings of the present study.

3.1. The Roles of Technology in Educational Assessment Practices

There have been extensive studies exploring the use of technology in the learning process, where technology is not merely used to support learning activities in the classroom to achieve effective learning, but also used in assessment. The findings of our study demonstrate that the existence of technology with all the features it offers has made it possible to conduct quality assessments and perform specific roles in supporting educational assessment practices as Fig. 2 has shown. Quality assessment here relates to the use of measurement instruments in assessments that are valid, reliable, and satisfy certain psychometric properties and assessments that are useful to improve the quality of learning. Our study has indicated that technology in educational assessment is not necessarily about changing the way a test or non-test measurement instrument is administered to students from paper-and-pencil mode to computerized administration mode. While it is undeniable that most of the studies we reviewed focused on the role of technology to administer tests [56]-[69], we also found a large proportion on the role of technology to provide feedback to students [70]-[84], perform test scoring [57], [85]-[98], and to evaluate the psychometric properties of a measurement instrument [99]-[115].

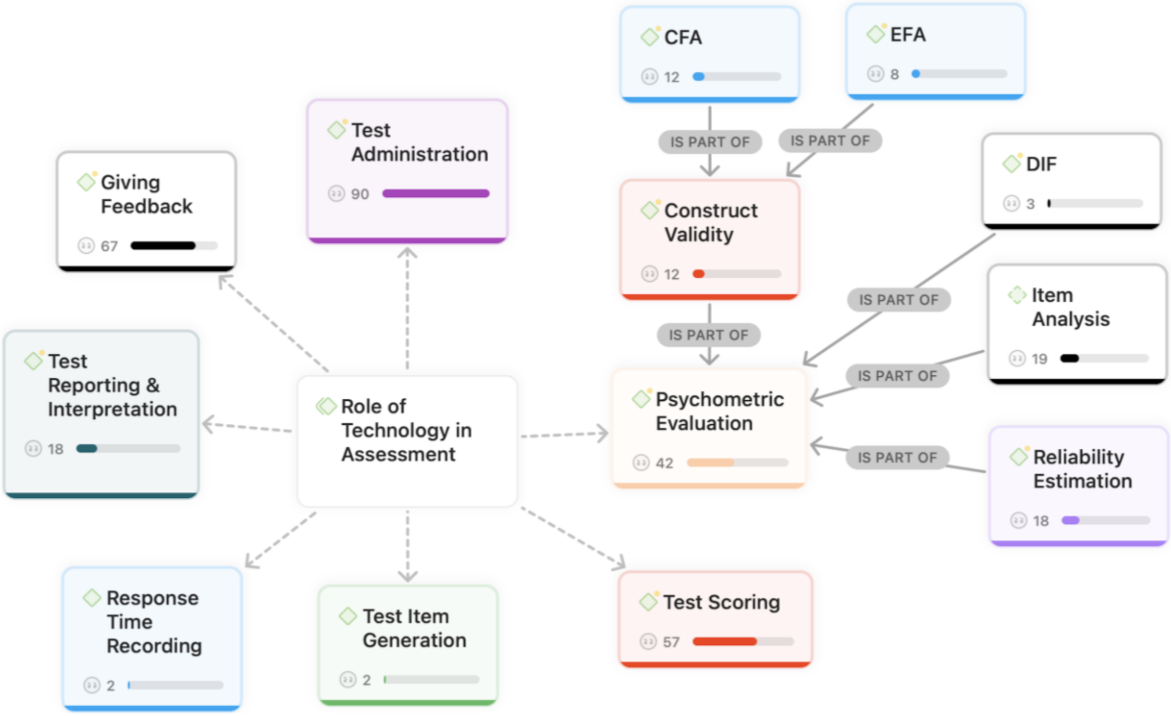

Fig. 2. The identified roles of technology in educational assessment

Besides, it has also been identified that the existence of technology can provide conveniences in automatically generating test items [183], [206], recording the time that students take to respond to each test item [86], [98], and even in reporting test results and their interpretation [123], [127], [134], [138], [156], [159], [170]-[172]. Given that technology contributes to providing quality assurance for measurement instruments through investigating psychometric properties, we also further explored the psychometric properties that previous studies investigated using specific technologies. Psychometric properties identified include reliability estimates, analysis the characteristics (statistics or parameters) of items, differential item functioning (DIF) which indicates measurement bias in test items, and construct validity. The use of technology to explore psychometric properties in the form of providing construct validity evidence in previous studies was performed through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

Table 3. Technologies used to support educational assessment practices

Role of technology	Technology Used in Educational Assessment Practices for Corresponding Role
Test administration	QuizOne [116], CleverTesting [117], State of Texas Assessment of Academic Readiness (STAAR) [89], TestVision [118], Wondershare-Supreritem [119], Wix.com [59], Articulate Rise 360 [120], Online test developed by Yureka Education Center (YEC) [66], Sero! [120], online formative assessment tool (OFAT) [63], *Programa en Li ́nea de Apoyo a las Tareas Academicas* (PLATA) [121], ILIAS software [122], Mini-CEX WebApp [123], JACK (e-assessment system) [124], OLM-SRL [125], MOCCA-College [126], XUEYIKU (mobile application) [127], Zoombinis [128], computer-delivered ELA [129], simulated-based task [57], Miro [120], Microsoft Visio [130], GraLeV [131], e-Mat testing [60], COPTEFL [132], computer-based assessment (CBA) tool [86], CAFF [133], Socrative [83], [85], [134], Moodle [72], [78], [130], [135], [136]-[138], Mobius Assessment or STACK [87], [139], Microsoft Teams [65], [140], [141], Kahoot! [79], [84], [142], Blackboard [93], [140], [141], [143], [144], WRITER [145], Mulberry [146], android-based gamification app [147], CEREC software [148], computer-based simulation [57], Concerto [149], CrossQuestion game [61], Dewis [80], e-assessment system (EAS) [150], e-exam platform [151], e-portfolio [59], Edmodo [152], ELLA-Math CBA system [153], IHMC CMap tool [58], Inspera [154], Lectora online [144], MCAT [97], Monster P.I. [155], OpenCT [98], PhysTHOTS-CAT [156], Pocket Money [157], Q-Global [158], SEAKMAP [159], SelfAssess-plugin for Moodle [160], SELweb's ER assessment [100], smartphone app [64], technology-based assessment [161], Tryout application [74], and web-based dynamic assessment [68].
Giving feedback	CleverTesting [117], Pigaiwang [82], OJ system [162], PLATA [121], CPR Tutor [163], OLM-SRL [125], XUEYIKU (mobile application) [127], Automated Writing Evaluation [75], Zoombinis [128], Microsoft Word [73], Panopto [70], CAFF [133], CoFee [164], AI [149], [165]-[167], eDia [77], [168]-[170], Socrative [83], [85], [134], Mobius Assessment or STACK [87], [139], Machine Learning [57], [58], [76], [96], [162], [171], Kahoot! [79], [84], [142], WRITER Tool [145], Mulberry [146], Automarker [172], CEREC software [148], Computer-assisted knowledge graph analysis [173], Dewis [80], E-assessment system [150], Game-based assessment [174]-[176], Grammarly [82], Lectora online [144], Neural network [177], NLP tools [177], Pe(er)fectly Skilled [178], SelfAssess-plugin for Moodle [160], Tryout Application [74], Viewbrics [179], Virtual reality [180], Wikipedia corpus [171], and Zoom chat application [181].
Test scoring	QuizCbot (conversational chatbot) [92], STAAR online platform [89], speech assessment for Moodle (SAM) [88], TestVision [118], Python [182], [183], OJ system [162], CPR Tutor [163], CBA Tool [86], CoFee [164], BLSTM [184], EvoGrader [185], artificial intelligence (AI) [165]-[167], [186], EnglishCentral [88], eDia [77], [168]-[170], Socrative [83], [85], [134], Moodle [72], [78], [130], [135]-[138], Kahoot! [79], [84], [142], GAMET [96], [187], Blackboard: LMS [93], [140], [141], [143], [144], R program [91], [94], [95], [110], [114], [188]-[192], Android-based gamification [147], ANNOTA platform [193], ARTE [96], Automarker [172], BILOG-MG [62], [66], [67], Computer-assisted knowledge graph analysis [173], Custom GPT (GPT-4) [165], Dewis [80], E-assessment system [150], ELLA-Math CBA system [153], GPT-3 [186], Lectora online [144], MCAT [97], OpenCT [98], Optical identify (OID) [175], Quest [91], [104], TAACO [96], video-based communication assessment (VCA) [90], and writeAlizer [187].
Psychometric evaluation	SmartPLS [64], Mplus [188], Winsteps [100], [194]-[197], SPSS [101]-[103], [105], [106], [111], [168], [188], [197]-[202], LISREL [99], [105], [108], [109], [111]-[113], [198], AMOS [107], [194], [199], [202], R program [91], [94], [95], [110], [114], [188]-[191], [192], Anates [28], BILOG-MG [62], [66], [67], Conquest [60], [101], [203], IRTPRO [204], Jamovi [168], PARSCALE [205], Quest [91], [104], Stata [203], and XCalibre program [184].
Test reporting and interpretation	QuizCbot (conversational chatbot) [92], Mini-CEX WebApp [123], CBA tool [86], eDia [77], [168]-[170], Blackboard: LMS [93], [140], [141], [143], [144], Dewis [80], PhysTHOTS-CAT [156], SEAKMAP [159], Tryout application [74], and video-based communication assessment (VCA) [90].
Response time recording	CBA tool [86]
Test item generation	Python [183] and MATrix LABoratory (MATLAB) with the Symbolic Math Toolbox extension [206]

3.1.1. Technology in Educational Assessment Practices for Test Administration

Various technologies with certain features have made it easier to administer tests to students. Administering tests using technology includes activities to register or generate items, create item banks and databases, select items from available item banks, import or export items, and manage test design on the technology used [117], [207]. The technology used can be classified into several types according to the purpose of measurement, the characteristics of the measurement instrument used, and the measurement target. These types of technologies include learning management systems (LMS), computer-based testing software, online survey platforms, game-based assessments, and mobile applications. The review also found some specialized testing platforms explicitly developed for administering tests. In addition to organizing the learning process, learning management systems (LMS) offer features for organizing assessments throughout and at the end of learning. A study by Köroglu [152] demonstrated the use of Edmodo, one of the LMS platforms, to support a digital formative assessment. Similarly, Dlab et al. [130] utilized Moodle LMS to conduct online tests through self-assessment and theory exams. The majority of tests organized through the LMS are held online, such as writing tests using Blackboard [143], completing case reports from clinical cases on Blackboard collaborate [140], and administering questions online [138]. Apart from the LMS, there is a distinct website for administering tests, namely Q-Global. One study used the website to administer a test to measure students' English and vocabulary skills [158].

The review found that computer-based test (CBT) development is adapted into several types and uses various terms. Studies conducted by Suhardi [67], Sahin and Gelbal [97], and Goodwin et al. [155] used the term computerized adaptive testing (CAT) to administer the examination. In addition to CAT, studies by Yao [69], Nagy and Korom [168], Renes et al. [118] administered computer-based assessments. The assessment content presented varies, ranging from English language skills to reading comprehension. Tests or tasks within computer-based assessments contain more than just textual content; they may also incorporate visual elements such as images or videos [154]. Notably, computer-based assessment offers a unique experience for students by equipping various question types presented. For example, study by Pásztor et al. [77] developed an online figurative test through the eDia system where students had to work on tasks related to generalization, discrimination, cross classification, recognizing relationships, differentiating relationships, and system construction. Another study by Simon et al. [169] used eDia interactive diagnostic testing environment to administer a test that aims to measure the visual literacy framework of fourth grade students. Some examples of test types are completing maps with available symbols, image-assembly tasks for rearrangement of several puzzle pieces into meaningful shapes (outline/silhouette), creating two-dimensional spatial representations from available images, recognition and differentiation tasks, visual interpretation and recognition of connections (attaching meaning to abstract images), and modality change.

The utilization of online survey platforms for test implementation is considered more practical due to its presentation in the form of a quiz. Some examples of quiz-type tests are reported by studies from Alharbi and Meccawy [85], Guzman-Orth et al. [129], and Phoophuangpairoj and Pipattarasakul [96]. The quiz-type test uses multiple-choice questions. Quizzes offer the advantage of being a time-efficient assessment tool while also effectively measuring students' pre-existing knowledge before the learning process begins. Technology offers a game-based assessment, which enhances the engagement and excitement of the test experience. One application or website that can be discovered in this review is Kahoot!. Tests through Kahoot! can be done by students on their cellphones/tablets [84] or collaboratively [176]. Kahoot! can be used as a summative assessment in online learning [79] or a formative assessment tool [142], [176]. Numerous studies have innovatively designed their testing platforms to cater to specific testing requirements. For instance, Marlianawati et al. [64] and Kusairi [74] have developed smartphone-accessible test applications. Furthermore, various e-assessment systems have been developed to evaluate students' competencies in mathematics [60], assess music theory mastery [151], and assess students' proficiency in medical skills [150].

3.1.2. Technology in Educational Assessment Practices for Providing Feedback

One of the main advantages of using technology in assessment is its ability to provide real-time and effective feedback. Feedback comes in various forms, such as comments and scores [164], automated messages that match students' answers [208], interactive dialogue between lecturers and students through apps [135], and direct messages that provide verbal instructions and explanations [161]. In addition, technology also enables consistent and objective feedback through typographic evaluation [165] and instant feedback with color coding for correct or incorrect answers [80]. Other forms of feedback include the use of emoticons to add visual elements [179], explainer videos to clarify material [70], [174], and audio feedback to correct critical errors [163]. The combination of these different forms of feedback allows for a more comprehensive assessment and supports a more effective learning process.

Based on the implementation technique, feedback was provided through several different techniques to ensure maximum effectiveness and engagement. Firstly, feedback is delivered directly after each question, allowing participants to receive a response immediately after answering. This technique provides an opportunity to correct or deepen their understanding immediately [74], [77], [80], [129], [137], [161]. Second, feedback is also delivered at the end of the answer [70], [72], [133], [144], [170]. Providing feedback at the end of an answer can provide an overall picture of participants' performance after completing a set of questions, helping them evaluate their overall results and identify areas for improvement. Third, feedback is provided in a real-time or synchronous format, enabling direct interaction between the participant and the feedback giver [72], [87], [139], [172], [181]. This technique facilitates direct dialogue, allowing participants to ask questions and get instant clarification regarding the feedback provided. With the combination of these feedback techniques, it is expected that participants can obtain more comprehensive and useful information for their learning. Some interesting aspects of innovative and effective feedback include the use of color coding to mark correct or incorrect answers [80], providing feedback immediately after the student answers, and providing instructional assistance in the form of guiding questions if answers are incorrect [77]. In addition, there are feedback dialogue features between lecturers and students, and the use of virtual characters that can read out instructions and model tasks [135], [161]. Systems such as prepCheck and typography evaluator provide consistent, objective, specific and accurate feedback, often based on pre-defined criteria [148], [165].

Technologies used in education often involve different types of feedback to improve student learning outcomes. For example, machine learning is used to help evaluate concept maps [58], while automated writing evaluation and criterion automated corrective feedback provide formative assessment feedback to improve students' writing performance by dividing feedback into categories such as grammatical errors, word usage errors, and errors in writing mechanics [71], [75]. Formative video feedback offers video feedback after students complete the online course [70]. OFAT and Socrative quiz provide various feedback options, including instant feedback, picture clues, and explanation [63], [85]. Computerized formative assessment and multimedia feedback systems provide automated scoring, feedback for writing scientific arguments, and feedback that depends on students' cognitive resources [76], [133]. Feedback can also be provided instantly through feedback windows or through e-assessment systems with dynamic color-coded feedback that distinguishes correct and incorrect answers [74], [80]. Interactive feedback provides step-by-step guidance to solve problems or tasks, while more in-depth feedback such as personalized feedback and targeted content-based feedback helps students improve performance and knowledge elaboration [87], [173]. Kahoot! interactive exercise feedback assessment and feedback with emoticons and text are some examples of methods used to provide feedback in various formats [79], [179].

Some technologies such as feedback boards, immediate feedback, and feedback with constructive support provide feedback as soon as the student completes the task and help the student understand the mistakes and improvements needed [77], [84]. Feedback through video explanation is also used for categories of material that students have not understood [74]. Technologies such as PLATA and typography evaluator provide specific feedback as well as additional content for course reviews, with a focus on consistency and objectivity [121], [165]. On the other hand, technologies such as Mulberry provide feedback when students make coding errors, while examiner feedback offers in-depth explanations of errors and improvements [146], [150]. Feedback from external sources after presentations and interactive concept mapping tools also provide judgements that match the aspects listed in the computer system [180], [209]. Finally, software such as CPR Tutor and online quiz feedback provide audio and automated feedback to improve student performance in various contexts [136], [163]. Kusairi et al. [174], Barana et al. [87], and Barana and Conte [139] also emphasize the importance of interactive feedback that helps students understand problem-solving strategies by providing step-by-step solution and immediate feedback after each step or task.

3.1.3. Technology in Education Assessment Practices for Test Scoring

As identified through this literature review, one of the key roles of technology in assessments is test scoring. Technology enables teachers to automatically and rapidly grade students' work, saving time, and facilitating the scoring process. Teachers can input scoring rubrics [78], [81], [185], or establish scoring criteria for each test [163], [186], [209]. Teachers can assign scores to individual questions or tasks for quiz-based tests and interactive tasks [98], [147]. This feature can increase flexibility in scoring for objective and subjective tests. Additionally, automated scoring with technology helps maintain consistency and objectivity in grading. In addition to the scoring process, technology provides test results to educators and students directly. Some of these technological tools are include Socrative [83], [85], [170], eDia system [170], Moodle LMS [130], CALT [207], and STAAR [89]. Test results may be presented as a total score, as well as individual scores for each test item, question, or task. In more detail, technology in the scoring process is capable of providing information about correct and incorrect answers [80], [96], [144]. Moreover, the use of GAMET in the scoring process ensures the identification of any errors made by students, such as grammatical, misspelling, and typographical errors [96]. A study by Lee et al. [159] used a deep learning model (artificial intelligence/AI) to transform data into a final score, which was then disseminated to students to monitor their learning progress. The previously mentioned features benefit educators and learners as a formative assessment process, where test results can be used as a quick and continuous improvement.

Upon further examination, the scoring process using technology is often integrated with providing feedback. For example, a study by Mizumoto and Eguchi [186] demonstrates that automated essay scoring with GPT can provide automated scoring and feedback on student writing quality results. Technological tools such as machine learning [167] and Athena [164] can also provide scores and feedback. Furthermore, another study by Qian and Lehman [146] utilized Mulberry to assess students based on predetermined expected outcomes. If the students' answer aligns with the output, then the answer is considered correct. Conversely, Mulberry can promptly offer feedback if the answer is incorrect, allowing students to continue attempting until the correct answer is achieved.

3.1.4. Technology in Education Assessment Practices for Evaluating Psychometric Properties of Measurement Instruments

Technology plays an important role in evaluating the psychometric properties of measurement instruments in educational assessment. In general, the evaluation of psychometric properties of measurement instruments includes providing evidence of validity, estimating reliability, analyzing items, and detecting item bias. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) have become two widely used approaches to provide evidence of the validity of measurement instruments. Our present study reports on the large number of studies that employed SPSS to perform the EFA procedure, some of which are Berlian et al. [102], Daryono et al. [103], and Ergül and Tasar [198]. Meanwhile, some studies have used LISREL [99], [105], [111] and SmartPLS [64], [115], [210] to perform the CFA procedure to provide validity evidence of measurement instruments. Moreover, other applications such as AMOS also seem to be the preferred choice for researchers to conduct CFA procedure [194], [199], [202]. It has also been identified that MPlus can be used as an alternative to perform the CFA procedure [188]. Although SPSS is likely to be most commonly used for EFA, it is also used by some researchers to estimate the reliability of measurement instruments [102], [105]. Besides SPPS, Stata is also an option for other studies [203] to estimate the reliability of measurement instruments.

For the purpose of analyzing the item characteristics of measurement instruments, previous studies generally used the classical test theory and modern test theory approaches. Previous studies analyzing item characteristics under the classical test theory approach used technologies such as Anates [28] and R program [190]. However, this study reveals that the use of technology is necessary when researchers choose a modern test theory approach to analyze item quality. Some studies preferred the Rasch model for the purpose by utilizing technological software such as Quest [91], [104], Winsteps [195]-[197], Conquest [60], [101], and Jamovi [207]. Other studies conducted item analysis based on item response theory by utilizing software such as the R program [95], [110], [191], BILOG-MG [62], [66], [67], IRTPRO [204], PARSCALE [205], and XCalibre [184]. Some previously mentioned studies used different packages from the R program, such as 'psych', 'G-DINA', and 'mirt'. The 'mirt' package also enables users to conduct multidimensional IRT parameter estimation using maximum-likelihood methods, supporting both exploratory and confirmatory models [211].

A quality measurement instrument should also be unbiased and not show preference towards any particular gender, race, ethnicity, or culture. This principle underlies the basis for the development of item bias detection methods in educational assessment. Some studies have employed item bias detection by applying the differential item function (DIF) method using Winstep [100], [197]. The DIF analysis with the aid of certain technologies allows it to be extended to the analysis of bias at the test level in the form of test differential functioning (DTF) across various aspects, such as gender and ethnicity [100]. Recent studies have also demonstrated that other researchers utilize open-source software like R program to analyze DIF in measurement instruments, as demonstrated by Sumin et al. [114]. They were using regression logistics technique to detect DIF by the gender, to be specific.

3.1.5. Technology in Education Assessment Practices for Test Reporting and Interpretation

Technology's role in assessment extends beyond scoring, which is used for test reporting and interpretation. This role encompasses two activities: reporting and interpreting test results. Previous studies have identified variations in the information provided in test result reports. The information delivered on the test result report depends on the test makers' needs and the variables being measured. A study by Auphan et al. [86] presents a report on student reading performance that includes accuracy (weighted scores) and speed (response times). Accuracy refers to the score of students answering correctly, while speed indicates the time taken to complete the test. Li et al. [171] also explained that the test result contains details of scores, time taken, and the number of completed tasks, which helps teachers identify areas where students may be struggling. Furthermore, Lubrick and Wellington [93] highlighted that technology can quickly provide test result reports in raw scores to students. In addition to individual reports, the overall results can also be presented, as demonstrated by Barczak et al. [172] and Kusairi [74]. These studies emphasized the importance of including overall student scores and the time taken on the report. Reporting time taken by students aligns with the other role of technology in assessments, which is response time recording.

In addition to the information mentioned earlier, some previous studies have included test result reports along with feedback for each test item or question. This kind of format is typically designed for students as test takers. It provides individual feedback on each question and information on both correct and incorrect answers [134]. Students given information about the correction feedback can identify concepts they have not mastered properly and make mistakes in them. Similarly, Kaiss et al. [92] conducted a study that included a PDF format report containing answers to quiz questions and explanations, allowing students to evaluate their work after taking the test. As for the assessor, such as a teacher, the system can provide an assessment report detailing feedback related to six different inductive processes, which can be downloaded by the teacher [77]. Along with the comprehensive test result report, the technology also streamlines the process for teachers to handle raw score data, which can be separately communicated to students [93].

Not all technology tools provide both roles (reporting and interpreting) simultaneously. While most studies focus on the reporting aspect, several explore using technology to interpret test results in assessment. A study by Varga et al. [170] utilized the eDia system that provides detailed results about each student's score, a spider web diagram presenting a students' performance alongside the class average, while also offering personalized feedback. Additionally, the eDia system also provides supplementary documents that help interpret the results. These supporting documents can assist teachers in explaining student performance through test results. Another study by Yuan et al. [127]

used application-based assessment to generate statistical grades and reports for resident students, helping doctors avoid potential errors while saving time.

3.1.6. Technology in Education Assessment Practices for Response Time Recording

Previous studies have shown that assessments can be administered to students using technology, such as a computer, where this practice has come to be known as computer-based assessment. The use of computers that have a number of features and facilities and other technologies that can be integrated with computers for test administration allows teachers to record the time it takes students to respond to test items and complete the test. We have found two studies, namely [86] and [98], that utilize technology for this role. Both studies were conducted under a computer-based assessment environment with different measurement focuses, in which one focusing on the reading skills of elementary and junior high school students [86] and the other on the scientific problem solving skills of senior high school students [98].

Under the computer-based assessment environment that aimed to investigate reading ability (word reading and reading comprehension) of elementary and junior high school students through three-option forced-choice questions, the study [86] utilized technology to automatically record the time it took students to respond to the questions immediately after the all response options appeared. The utilization of technology in this case makes it possible to assess students' reading ability further based on the behavior demonstrated by the student as indicated by the relationship between the accuracy of the responses provided and the speed in responding to a question. Based on the relationship between these two aspects, there are four possible pieces of information that can be obtained about guessing behavior in providing a response to a question and the efficiency of cognitive processes or ways of reading that can affect students' reading ability. While utilizing technology to record response times provides more in-depth information about students' reading ability, using technology for this role raises an issue. The issue relates to the fact that the response time data obtained not only shows the time it takes students to comprehend the reading but also something else that is not the main target, namely the time students spend reading all the options. In light of the results obtained regarding the use of technology to automatically record response times, Auphan and colleagues [86] suggested displaying optimal response times to students to prevent students from responding quickly by guessing.

Under a computer-based assessment environment, Zhao et al. [98] have utilized technology to record the time it took students to solve complex scientific problems through experiments that required them to demonstrate competence in scientific inquiry, scientific reasoning, and scientific explanation. The use of response time recording in assessment practices involving scientific scenario-based tasks allows for information to be obtained about where students struggle more in using certain strategies to control variables and design experiments that lead them to success in solving scientific problems.

3.1.7. Technology in Education Assessment Practices for Generating Test Items

The learning process carried out in a distance mode, as what has happened during the COVID-19 pandemic era, has raised various challenges, including in terms of conducting assessments, both for formative and summative purposes. One of these challenges concerns the great potential for students to cheat on other students' work given that the tests and test items administered to students are the same. A number of possible actions can be undertaken by teachers to overcome this challenge, such as making tests more personalized or individualized and varying test items and distractor positioning when multiple-choice test items are employed. Given the knowledge, competencies, and skills that teachers have, this way could also be a challenge for teachers. Beyond the challenges in educational assessment practice that occurred during the COVID-19 pandemic, assessment practice can also be challenged when it is implemented on a wide scale or in an adaptive testing environment that requires a large number of test items and multiple test forms such that the test forms should be equivalent in terms of the characteristics of the items composing the test form. The presence of technology with its advanced features allows these challenges to be addressed, that is through the utilization of automatic item generation (AIG).

Of the 169 empirical studies we included in this review, we identified two, namely [183] and [206], focusing on the role of technology to automatically generate test items that allows for large-scale assessment while ensuring test security. The idea of maximizing the potential of technology to perform this role is based on the need to involve a large number of test items in assessments with a large number of test takers, while the time, cost, and effort to meet this need are limited. Technology that enables this role has been used for assessment purposes through computer-based assessment at the senior high school level with the focus lies on Turkish literature [183] and at the higher education level with the focus lies on undergraduate mathematics (i.e., linear algebra, differential calculus, and integral calculus) [206]. The former uses scripts written in the Python programming language to generate test items, while the latter uses the computational system MATrix LABoratory (MATLAB) with the Symbolic Math Toolbox extension. Furthermore, it is also worth reporting that the type of test items generated through such technology is five-option multiple choice.

The working procedures used under both technologies to automatically generate test items are basically similar, i.e., starting with determining the key characteristics, content, and elements of the test items to be generated based on the desired specifications or based on the parent item. In addition, it is also worth determining the specifications of the items to be generated by considering the knowledge, competencies, or skills that students are expected to demonstrate in solving the test items. Although the use of technology to automatically generate test items has been shown to offer a number of advantages such as saving the time teachers need to construct a large number of test items with diverse characteristics [206] and allowing for the generation of equivalent test items [183], maximizing the use of technology to generate test items also comes with its own challenges that need to be minimized or overcome. These challenges are mainly related to the need to ensure that the test items generated are appropriate to the level of thinking required by the students to solve the test items and the need for adequate skills in developing algorithms and a good command of certain programming languages and techniques [206]. Above all, the existence of technology that allows assessment practices to be carried out at a more advanced level such as generating test items automatically is expected to assist teachers in administering varied test items to each student through item pools that can be developed further [183]. In addition, it is also hoped that the potential of the technology for test item generation can be adapted and extended to the generating of complementary components of a test item, such as the graph of a function, as suggested by Viskotová and Hampel [206].

3.2. The Role of Technology in Educational Assessment Practices and How it Relates to the Education Levels

Besides reporting on the roles of technology in educational assessment practices, it is also worth investigating how these roles relate to levels of education. Investigating these relations is interesting as it allows us to look deeper into the trends of the roles of technology in educational assessment practices at every level of education from kindergarten to higher education. Fig. 3 describes the roles of technology in educational assessment practices at the levels of kindergarten, elementary school, junior high school, senior high school, and higher education (i.e., university, institute, or college). It has been demonstrated that at the kindergarten level, technology contributes a very limited proportion to educational assessment practices, where it was used only for administering tests [100] and evaluating the psychometric properties of measurement instruments [100], [108].

At the elementary school level, the development of technology, which offers a wide range of features, has the advantage of being used in a wide range of educational assessment practices. Of the seven types of technology roles in educational assessment practices, six can be found at the elementary school level, albeit in small numbers. It has been revealed that at the elementary school level, technologies with their features have been used by previous studies to report test results and their interpretation [77], [84], [86], [170], record response time [86], provide feedback [77], [84], [128], [170], perform test scoring [77], [86], [89], [94], [170], investigate the psychometric properties of a measurement instrument [112], [168], [194], and administer tests to students [68], [77], [84], [86], [89], [128], [131], [155], [168]-[170]. Among the six types of technology roles utilized in educational assessment practices at the elementary school level, the results of our review have placed test administration as the most widely used role by previous studies.

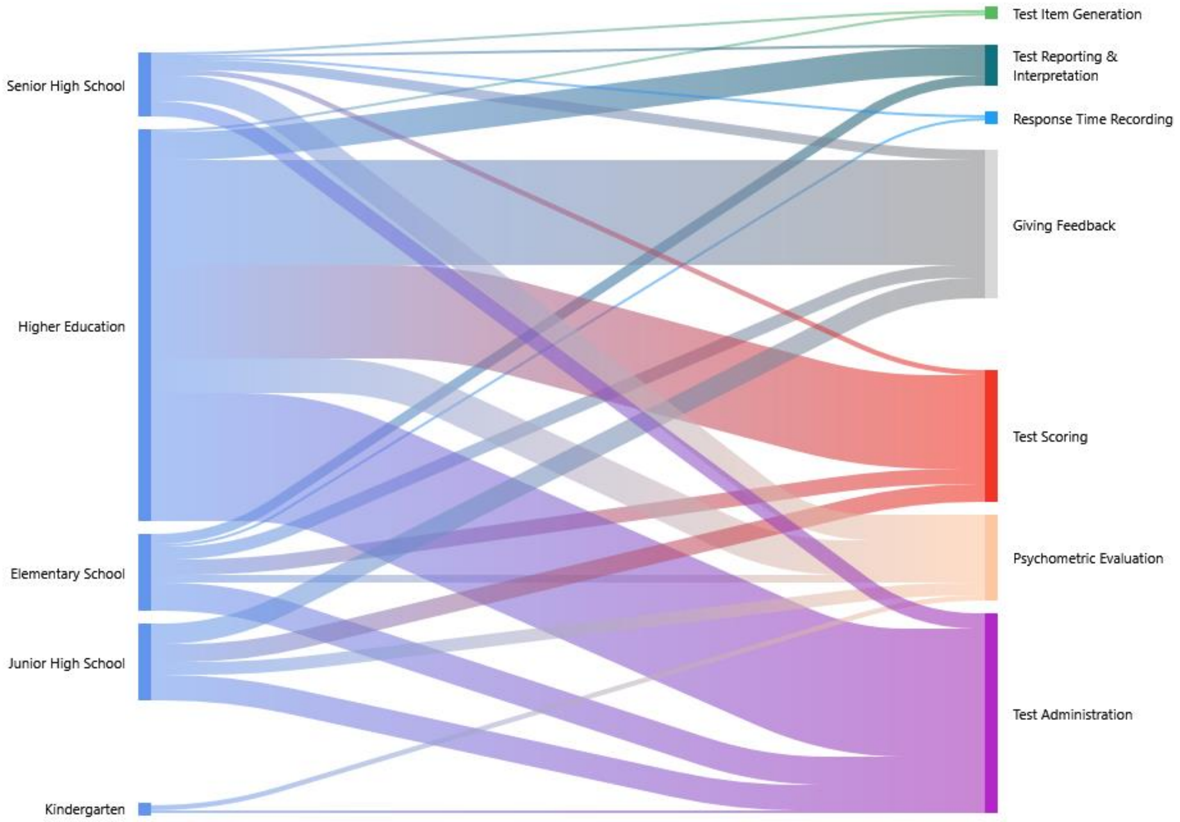

Fig. 3. The role of technology in educational assessment practices and how it relates to the levels of education

At the junior high school level, technology has provided various features that allow teachers or practitioners and researchers in the field of educational measurement and assessment to provide feedback to students [76], [79], [87], [128], [139], [179], [209], conduct test scoring [60], [79], [87], [89], [94], [153], [209], evaluate the psychometric properties of an instrument [60], [101], [107], [196], [203], and administer tests [60], [79], [87], [89], [128], [139], [155], [157]. The four roles share almost equal proportions, where the role of technology to administer tests still being the dominant one. At the senior high school level, although the quantity of studies that utilize technology for specific assessment practices is less than at the elementary school and junior high school levels, technology at this level has been used to the greatest extent given its potential. This is demonstrated by the presence of technology being used for all possible roles in educational assessment practices. In addition, while previous studies focusing on elementary school and junior high schools positioned technology mostly for administering tests, studies focusing on the senior high school level were more likely to focus more on the role of technology to evaluate psychometric properties of measurement instruments for specific constructs [62], [99], [103], [105], [106], [109], [110], [111], [192], [197].

Among the five levels of education that the present study focused on, we found that previous studies have used technology to support educational assessment practices more in higher education than in the other four levels of education. Technology at this level is also used to its fullest extent to support educational assessment practices in almost all identified roles. Similar to what we found at the elementary and junior high school levels regarding the dominant role of technology, the dominant role of technology for administering tests was again identified at the higher education level [56], [57], [59], [65], [69], [78], [80], [85], [88], [93], [96], [97], [121], [125], [142], [143], [149], [152], [159], [193] although the dominance was less pronounced than for providing feedback [63], [70]-[73], [78], [80]-[83], [85], [93], [121], [150], [160], [166], [180]. Furthermore, of the six roles that technology takes in supporting assessment practices in higher education, we have learned that it is mostly used for administering tests, providing feedback, and scoring student work on a test [57], [78], [80], [81], [83], [85], [88], [90], [92], [93], [96], [97], [130], [150], [162], [166], [193], [207]. In general, the results of our investigation placed the role of technology for test administration, providing feedback, and test scoring as the three most frequently found in previous studies by education level, respectively. By education level, the findings of our study indicate that empirical studies focused on the use of technology for test item generation and response time recording are very scarce across all education levels. We unfortunately found no studies at the higher education level that focused on the role of technology to record response time.

3.3. The Role of Technology in Educational Assessment Practices and How it Relates to the Three Domains in Bloom's Taxonomy of Learning (i.e., Cognitive, Psychomotor, and Affective) that Become the Focus of Assessment

Within the cognitive domain, a majority of studies have used technology for test administration (55 of 148) [58]-[63], [65], [66], [72], [74], [77], [80], [83], [86], [97], [119], [157], [168] and giving feedback (34 of 148) [58], [63], [70], [72], [74], [76], [77], [80], [83], [87], [93], [121], [128], [129], [133], [137], [139], [167]. However, there is still limited study that focuses on the use of technology for test item generation and response time recording. In contrast to the cognitive domain, technology has primarily been utilized to deliver feedback in the psychomotor domain [71], [73], [75], [81], [82], [150], [163], [171], [173], [176], [180], followed by test administrations [57], [64], [125], [132], [143], [152], [158], [169], [176]. In the psychomotor domain, no study was found that used technology for response time recording and test item generation. For the affective domain, we find that most previous studies for psychometric evaluation [99]-[102], [107], [111]-[115], [188], [194], [195], [200], [210] and test administration [63], [68], [72], [85], [89], [100], [135], [155] use technology in educational assessment practices. Notably, technology has not been employed for response time recording, test item generation, test reporting, and interpretation in assessments targeting the affective domain.

A total of 218 assessment activities utilizing technology were identified, with test administration emerging as the predominant technology-enhanced role employed for measuring the cognitive domain (73 out of 218) [79], [118], [120], [122], [124], [126], [131], [134], [136], [147], [153], [156], [161], as opposed to the psychomotor [143], [150], [152], [158], [169], [176] and affective [63], [68], [72], [85], [89], [100], [135], [155] domains. The next most common technology role is giving feedback. This role is most often used for the assessment in the cognitive domain. The role of technology for psychometric evaluation is almost equally used for the affective and cognitive domains. Educational assessment practices such as reporting and interpreting test results using technology have been used in the cognitive [74], [77], [80], [86], [93], [134], [138], [156], [159], [171] and psychomotor [81], [171] domains but have not been used in the affective domain. The role of assessment as response time recording and test item generation has not been used in the affective and psychomotor domains, but rather has been used in cognitive domain [86], [98], [183], [206]. The role of technology in educational assessment practices and how it relates to the three domains of learning shown in Fig. 4.

3.4. The Role of Technology in Educational Assessment Practices and How it Relates to the Region by Continent in Which the Assessment Took Place

In addition to investigating the roles of technology in educational assessment practices with respect to the level of education and the domain on which measurement is focused, it is also worth exploring the connections between the roles of technology and where assessment takes place. The places where the assessments are conducted are initially identified by country. However, given that our investigation has identified many countries where assessments are conducted with technological support, for the purposes of this study, we categorized the countries into continents covering the corresponding countries. Our findings regarding the relationship between the roles of technology in assessment practices and where assessments are conducted by continent are presented in Fig. 5. Of the 169 empirical studies we reviewed, Fig. 5 indicates that educational assessment practices using technology are mostly conducted in Asia, followed by Europe, the Americas, Australia and lastly Africa. Meanwhile, when these technology roles were based on the continent where the assessment took place, we identified the three most used technology roles to be test administration, providing feedback, and test scoring, respectively. By continent, existing studies suggest support for the heavy dominance of the use of technology to administer tests, and at the same time demonstrate that the use of technology for test reporting and interpretation, generating test items, and recording response time has been understudied.

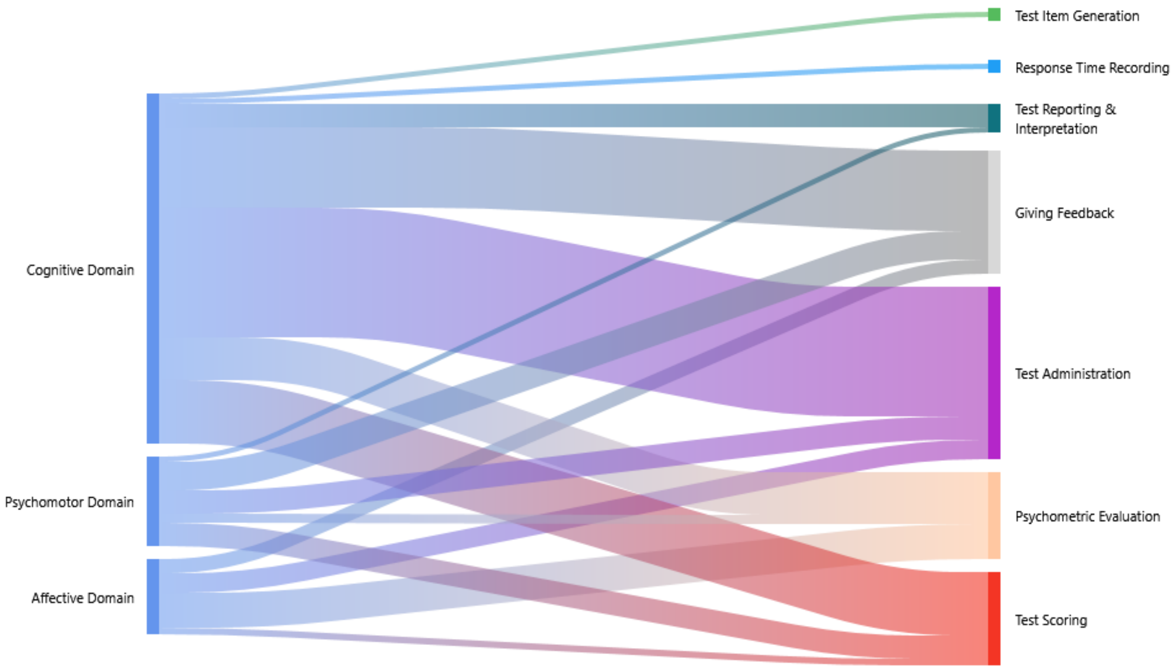

Fig. 4. The role of technology in educational assessment practices and how it relates to the three domains of learning

In Asia and Europe, the existence of technology with its various features is utilized to do seven things related to educational assessment practices, namely to administer tests, provide feedback to students or improve learning, test scoring, investigate the psychometric properties of measurement instruments, report test results and their interpretation, record the time it takes students to respond to test items, and generate test items. The role of technology to administer tests remains the dominant one used by empirical studies conducted in Asia [56], [59], [61], [65]-[69], [98], [119], [125], [132], [145], [149], [152], [175] and Europe [58], [63], [77], [79], [80], [86], [87], [116], [130], [139], [154], [160], [169], [193], [207]. The dominance of the role of technology for test administration in Asia that is so obvious compared to other roles of technology is not found in Europe. Although it remains the role that has received the most attention by studies conducted in Europe, the difference between the number of studies focusing on the use of technology for test administration and those for feedback provision is only one. This indicates that European studies place technology for both roles almost proportionally. While there have been studies in Asia and Europe that have paid attention to the use of technologies for test item generation and response time recording [86], [183], [206], both roles offered by the technology remain understudied.

Studies focusing on assessment practices conducted in the Americas utilized technology for five purposes in order of frequency: test administration [57], [93], [120], [121], [123], [126], [128], [129], [153], [155], [158], [161], [176], test scoring [57], [81], [89], [90], [93], [94], [153], [167], [177], [185], feedback provision [73], [76], [81], [93], [121], [128], [129], [161], [167], [176], test reporting and interpretation [81], [90], [93], [123], and psychometric evaluation [203]. It turns out that studies in the Americas are still paying more attention to the potential provided by technology for administering tests and very few studies have focused on evaluating the psychometric properties of measurement instruments using technology. In addition, we found no studies in the Americas devoted to exploring technologies for response time recording and test item generation. Similarly, studies in Africa focused more on using technology to administer tests [117], [142], conduct test scoring [92], provide feedback [117], and report test results and interpretations [92]; in which the role of technology for test administration was found to be the most prevalent in African studies. However, none of the African studies included in the present study focused on the educational assessment practices of response time recording, test item generation, and psychometric evaluation using existing technologies. Furthermore, while the studies we reviewed showed a predominance of using technology to administer tests in Asia, Europe, the Americas, and Africa, we found that in Australia, studies focused more on using technology to provide feedback [70], [136], [166]. In addition, technology was also used for test scoring [136], [166] and test administration [136], [157], with both sharing the same frequency.

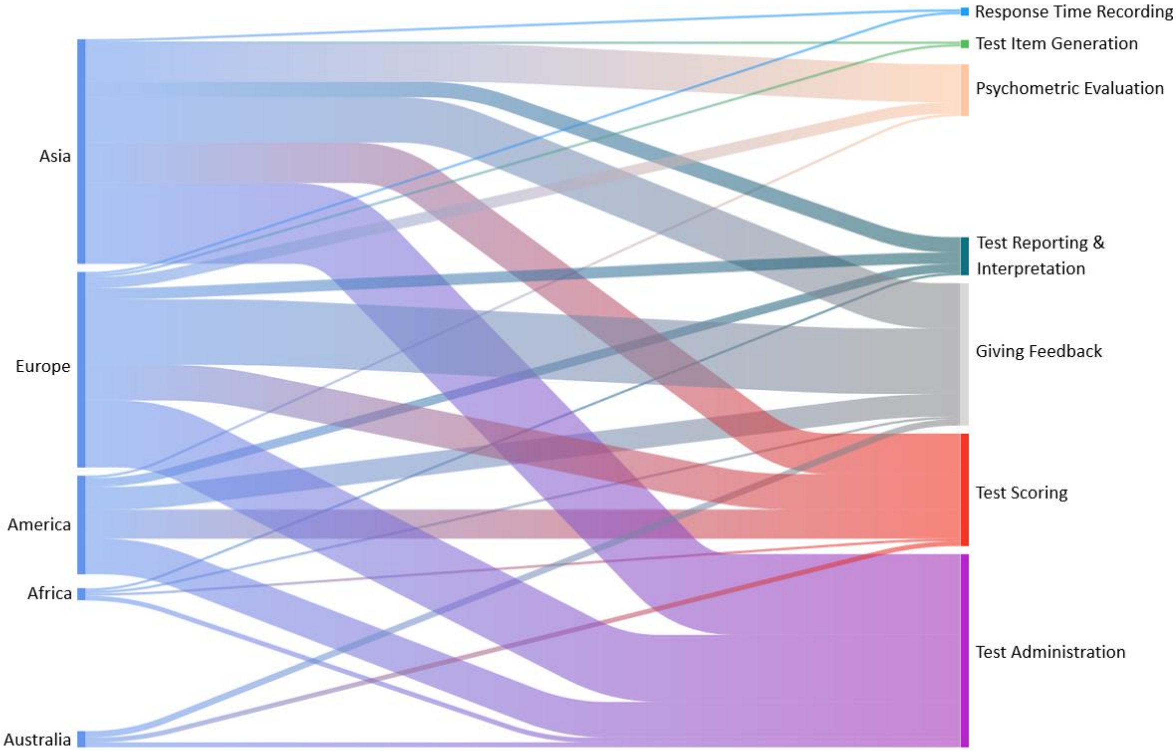

Fig. 5. The role of technology in educational assessment practices and how it relates to with the region by continent where the assessment was conducted

3.5. General Discussion: Lesson Learned from Previous Studies So Far

Given the development of technology that supports assessment practices to a greater extent than simply changing the mode of test administration, we reviewed existing studies to investigate further the extent to which technology has been used for various purposes related to educational assessment practices. Motivated by the study [27] which has demonstrated a number of roles that technology offers in supporting assessment practices in education, we reviewed a number of empirical studies to further investigate the extent to which the roles of technology are utilized considering the level of education, the domain of focus in the assessment, and the region where the assessment takes place and to identify other possible roles offered by technology in educational assessment practices.

Our study that was conducted through a systematic review approach found two other roles that technology provides to be utilized in educational assessment practices in addition to the roles we identified in [27], namely response time recording and test item generation. These two roles, however, still need to be explored more in future studies as out of the 169 studies we reviewed, only two studies each were devoted to these two roles. Meanwhile, the three roles of technology that are most widely used in assessment practices are test administration, feedback provision, and test scoring, respectively.

In addition to obtaining findings on the roles that various technologies provide in supporting educational assessment practices, our study has also identified the names of technologies that can be used to serve these roles. Through this identification, we came across a number of technology names that can serve more than one role, making it possible to carry out more diverse assessment practices with a single particular technology. Furthermore, when it comes to the question of which technologies should be used to ensure improvements in learning outcomes, although the focus of this study is on educational assessment, the answer would be the technologies listed that can serve the role of providing automated, interactive, or immediate feedback. By relating the role of technology in assessment practices to educational levels and learning domains, it was noted that previous studies have paid much attention to the role of technology for administering tests in higher education and for administering tests developed to measure student competencies in the cognitive domain. At last, geographically, the use of technology for test administration is prevalent in Asia, Europe, the Americas, and Africa, while the use of technology for providing feedback is most widespread in Australia.

Technology integration in assessment for administering tests has gained popularity, as it is used for assessment in higher education to kindergarten. This shows the adaptability of various forms of technology across different subjects. The selection of the type of technology considers the test form, the domain being measured, and the necessary features to support the test format [27], [212]. Technology-assisted assessment is implemented through diverse modes, such as computer-based tests, game-based tests, and online-based assessments. Most technology-assisted tests assess the cognitive domain, followed by the psychomotor and affective domains. Cognitive domain assessments predominantly utilize subjective test formats such as multiple-choice questions embedded in quizzes. Many applications and websites offer quiz features with multiple-choice question formats. In addition, technology enables teachers to design examinations that assess the process rather than just the result of students' responses. Employing tests as a formative assessment can provide sufficient information to evaluate students' understanding throughout the learning process. Some studies employed different practice in formative assessment, such as self-assessment [111], [112], [193], [213] and peer-assessment [166], [214]. Moreover, different practices in formative assessment could encourage students to have an active role in the learning process [215].

While technology has received immense attention for its role in administering tests, it has also received tremendous recognition for its role in providing feedback on the student's learning and the learning that the teacher facilitates. Several applications or websites as technological tools can facilitate objective and personalized feedback. Through the personalized feedback, teacher help the students to understand their strength and identify areas for their improvement [216], [217]. Moreover, technology allows teachers or administrators to provide interactive feedback to their students, enabling students to confirm the comments directly with the feedback that teachers or administrators have provided. Tests organized using technology to measure the cognitive domain require various features to ensure comprehensive feedback for students. For instance, students should receive information on correct and incorrect answers along with some corrections upon completing a test. This feature benefits students as they can immediately correct the mistakes that they made and prevent the reinforcement of incorrect knowledge. Our present study demonstrates that technology predominantly delivers feedback to students in higher education level, except in kindergarten. These findings suggest an opportunity to create assessment activities that provide feedback in kindergarten with the support of technology.

While Asia, Europe, the Americas, and Africa show the dominant role of technology for test administration, the present review indicates that assessment practices in Australia conducted with the support of technology are more emphasized to support student learning through the provision of meaningful or constructive feedback. The emphasis on the role of certain technologies in assessment practices in a country could possibly be influenced by the national curriculum. For instance, it may be reasonable to find that Australian studies emphasize the role of technology in providing feedback as the Australian curriculum places more emphasis on formative assessment or assessment for learning practices [218]-[220]. Generally speaking, policies in the national curriculum regarding which types of assessments to focus on, without neglecting other types of assessments, may influence the extent and manner in which technology should be used to support assessment practices.

Similar to the role of technology in providing feedback, test scoring is used for assessment from junior high school to higher education levels. While technology's role in educational assessment practices in kindergarten is still not that diverse, it is primarily utilized for the cognitive domain, reflecting the emphasis on the cognitive domain in test administration. This connection is evident in previous studies that have used technology for both roles simultaneously. Certain technologies offer comprehensive features for test administration and scoring student answers, streamlining the process and saving teachers and organizers time and effort by eliminating the manual correction of student responses. Such feature particularly beneficial for giving accurate and consistent results, crucial for ensuring the validity and reliability of the measurement instrument used in the assessment [221].

The assurance of the validity and reliability of measurement instruments is made possible through the presence of technology that can assist in conducting psychometric evaluations. The review we have performed has highlighted that the existence of technology with its advancements allows researchers to obtain quality instruments for use in educational assessment. The advancement of technology in assisting researchers to find quality instruments is demonstrated by the variety of technologies that can be used to provide evidence of construct validity through EFA and CFA, eliminate test or questionnaire items that contain bias through DIF analysis, and obtain test or questionnaire items according to the desired characteristics through item analysis. Through the present review, we have also identified technologies for evaluating psychometric properties of measurement instruments that have their own advantages and challenges. Among a number of technologies that can be used in psychometric evaluations, the R program with its powerful suite of packages has become the technology of choice for running psychometric analyses from simple to advanced ones. Although R can be fully used without requiring the payment of any license fees, using R can pose a challenge for those who are not comfortable with their coding or programming skills or for those who are more comfortable with technologies that offer a user-friendly graphical user interface (GUI). The use of technologies such as LISREL, SEM-PLS, Winsteps, BILOG-MG, IRTPRO, SPSS, and Jamovi for psychometric evaluation by previous studies indicates a high tendency to use user-friendly technologies even though the use of these technologies, except Jamovi, requires payment of license fees. To this end, regardless of the technology used for psychometric evaluation, as long as information regarding the psychometric properties of a measurement instrument can be conveniently obtained, it should not be a contentious issue.

From the review findings, some technologies are found to serve multiple roles in the assessment process within a study. This demonstrates that today's technology has many advanced features that can be used optimally. The technology is able to accommodate roles in organizing assessments. One example is Kahoot! [79], [84], [142] and Socrative [83], [85], [134]. Based on the review, both technologies can be used for test administration, giving feedback, and test scoring simultaneously. Additionally, other technologies, like eDia, can also support feedback provision, test scoring, test reporting and interpretation [77], [168]-[170]. This allows teachers and test administrators to work efficiently in familiarizing themselves with the application/software to be used and utilizing the application to administer the assessment.

3.5.1. Limitations and Directions for Future Studies

Although this study has been conducted through systematic procedures, we should acknowledge that the present study still has limitations which shall be reported. First, the findings of this study should not necessarily be generalized to a greater extent, especially when the roles of technology in supporting educational assessment practices are associated with the country or continent in which the assessment practices are conducted. The first limitation is driven by the fact that the databases used to obtain the literature reviewed in this study were only Scopus and ERIC and the type of literature selected was only journal articles reporting empirical studies. Several other databases such as Web of Science and PsycINFO can be considered for use in searching and retrieving literature when access to these databases is available and the type of literature in the form of conference proceedings papers can be used in addition to journal articles.

Second, while through this study we have been able to identify the various roles of technology in educational assessment practices and the technologies that fulfill these roles, we have not further explored the challenges involved in using specific technologies to fulfill specific roles in assessment practices. In addition, while some types of assessment practices conducted with the support of technology have been reported along with their challenges, others have not. While we acknowledge that the challenges that technology brings to supporting assessment or that arise as a consequence of implementing assessment practices in specific contexts with technology are not part of the main focus of the current study, we believe that an exploration of them could offer more insights and benefits in the development of educational assessment practices. The identification of these challenges allows for a deeper understanding of possible strategies to overcome these challenges and maximize the potential that specific technologies have to play a specific role in supporting educational assessment practices more fully than ever before.

In light the findings we have reported, as directions for future studies, it is important to focus on a deeper exploration of how technology can support educational assessment practices more effectively. Future studies could also include case studies of technology implementation in different educational institutions to provide insights into the specific challenges and advantages faced. In addition, a more in-depth study of the cultural, economic, and policy factors that might influence technology adoption in different regions would provide a more complete global understanding of how technology supports educational assessment practices. Evaluating the long-term impact of technology adoption, including the effect on student learning outcomes, is also an important area for study. Future studies focusing on the development and evaluation of new technologies that can support under-explored roles, such as response time recording and test item generation, would be valuable for improving the quality and efficiency of educational assessment in the future.

The third point, which is related to the previous one, concerns the use of artificial intelligence (AI) and large language models (LLMs) in particular, in educational measurement and assessment. Although we mentioned some tools using AI, the focus of the paper was broader and aimed at all the technologies used in assessment. But we cannot help but notice that the integration of AI in educational assessment has revolutionized assessment methods and practices, enabling new approaches in automated scoring, automated item generation, creating item content, generating task stimuli (such as images for example), personalized feedback to students, identifying areas of strength and weakness, and adapting to individual learning patterns to support better educational outcomes [222]-[226]. However, using AI in assessment comes with certain risks. LLMs are trained on diverse datasets, that equips the models with a broad understanding of language, that enables LLMs to generate coherent, contextually relevant responses to user prompts. But AI technologies reflect and can even reinforce biases in the training data, and if an LLM was trained with nonrepresentative datasets, it can hold bias towards a certain group which can result in unfair assessments [227]. Second, the use of AI in assessment can lead to lack of transparency, as students and educators may not fully understand the criteria by which they are being evaluated. As such, further research is needed to explore the potential detriments of advanced AI technologies in the area of educational assessment.

Finally, the integration of modern technologies into educational assessment brings with it a host of ethical considerations that educators and policymakers must navigate carefully. One of the central concerns is privacy, as digital platforms can collect and store vast amounts of personal data, which can include not just academic performance but also behavioral and biometric data, creating risks of misuse or breaches. Another concern is the question of fairness, as not all students may have equal access to the necessary devices and reliable internet connections needed for tech-based assessments, that can potentially disadvantage certain groups of students and exacerbate existing inequalities. Furthermore, the use of technologies in educational assessment, for example automated scoring systems, raises significant ethical challenges about the transparency and explainability of assessments. Moreover, the reliance on technology raises questions regarding the validity and reliability of the assessments, as technical issues can lead to incorrect assessment results, which can unfairly impact students' educational opportunities and outcomes. There is also the issue of guaranteeing that the technological tools used do not inadvertently favor or prejudice any group of students, for example, individuals with disabilities who may require special accommodation. Thus, while modern technologies can offer innovative approaches to assessment, they must be implemented with a careful consideration of ethical principles, striving to ensure fairness, privacy, and equality for all students. These ethical challenges require thoughtful approach and call for further studies.

4. Conclusion

This study focuses on exploring the extent to which technology has been used in educational assessment practices and how the use of technology varies based on various factors. We found that the integration of technology in educational assessment practices has a very positive impact on the efficiency and effectiveness of the assessment process. Based on the review of 169 journal articles, technology was found to facilitate test administration through digital platforms, mobile applications, and computer-based software, and improve the quality of assessment by providing real-time feedback and enabling more in-depth psychometric evaluation. Technology also plays a role in speeding up and accuracy of analyses of item validity, reliability and bias, and supports more comprehensive reporting of test results. Our study also found that technology plays a significant role at all levels of education, from early childhood education to higher education. At the elementary school, junior high school, and senior high school levels, technology is primarily used for test administration, whereas at the higher education level, technology is used more widely, including in psychometric evaluation and providing feedback. The study also identified that while technology offers potential for response time recording and test item generation, both roles remain underexplored. Those two roles present new possibilities for improving the efficiency and accuracy of assessment. In addition, technology is more widely used for test administration in Asia, Europe and the Americas, whereas in Australia, the focus of the study is more on providing feedback. These findings further confirm that the application of technology in educational assessment not only improves operational efficiency but also increases the accuracy and reliability of assessment results. This is an important step towards better educational assessment practices and addressing challenges such as cheating by automating test item generation. This study emphasizes the importance of continuing to develop and explore the role of technology to improve overall assessment practices.

Author Contribution: All authors contributed sufficiently to the main contributor to this paper. All authors read and approved the final paper.

Funding: This study was funded by the Directorate of Research and Community Service (Direktorat Riset dan Pengabdian kepada Masyarakat, DRPM), Universitas Negeri Yogyakarta (UNY) through the Research Group and Penugasan bagi Guru Besar dan Tenaga Dosen Struktural research schemes for the fiscal year 2024.

Acknowledgment: We have to express our appreciation to the three anonymous reviewers and the editor for providing feedback on this paper which we found encouraging and very constructive.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] U. Karimah, H. Retnawati, D. Hadiana, P. Pujiastuti, E. Yusron, "The characteristics of chemistry test items on nationally-standardized school examination in Yogyakarta City," *REID (Research and Evaluation in Education)*, vol. 7, no. 1, pp. 1-12, 2021, https://doi.org/10.21831/reid.v7i1.31297.

[2] I. Rafi, H. Retnawati, E. Apino, D. Hadiana, I. Lydiati, and M. N. Rosyada, "What might be frequently overlooked is actually still beneficial: Learning from post national-standardized school examination," *Pedagogical Research*, vol. 8, no. 1, pp. 1-15, 2023, https://doi.org/10.29333/pr/12657.

[3] H. Retnawati, B. Kartowagiran, J. Arlinwibowo, and E. Sulistyaningsih, "Why are the mathematics national examination items difficult and what is teachers' strategy to overcome it?," *International Journal of Instruction*, vol. 10, no. 3, pp. 257-276, 2017, https://doi.org/10.12973/iji.2017.10317a.

[4] I. A. M. S. Widiastuti, "Assessment and feedback practices in the EFL classroom," *REID (Research and Evaluation in Education)*, vol. 7, no. 1, pp. 13-22, 2021, https://doi.org/10.21831/reid.v7i1.37741.

[5] M. Zhou, "Commentary: Significance of assessment in learning: The role of educational assessment tools," *Science Insights Education Frontiers*, vol. 18, no. 2, pp. 2881-2883, 2023, https://doi.org/10.15354/sief.23.co215.

[6] R. Rukli and N. A. Atan, "Simulation of low-high method in adaptive testing," *REID (Research and Evaluation in Education)*, vol. 10, no. 1, pp. 35-49, 2024, https://doi.org/10.21831/reid.v10i1.66922.

[7] N. Kania, Y. S. Kusumah, J. A. Dahlan, E. Nurlaelah, F. Gürbüz, and E. Bonyah, "Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills," *REID (Research and Evaluation in Education)*, vol. 10, no. 1, pp. 64-79, 2024, https://doi.org/10.21831/reid.v10i1.71032.

[8] R. Setiawan, W. Wagiran, and Y. Alsamiri, "Construction of an instrument for evaluating the teaching process in higher education: Content and construct validity," *REID (Research and Evaluation in Education)*, vol. 10, no. 1, pp. 50-63, 2024, https://doi.org/10.21831/reid.v10i1.63483.

[9] L. W. K. Yim, C. Y. Lye, and P. W. Koh, "A psychometric evaluation of an item bank for an English reading comprehension tool using Rasch analysis," *REID (Research and Evaluation in Education)*, vol. 10, no. 1, pp. 18-34, 2024, https://doi.org/10.21831/reid.v10i1.65284.

[10] N. H. Assa'diyah and S. Hadi, "Developing student character assessment questionnaire on French subject in state high schools," *REID (Research and Evaluation in Education)*, vol. 7, no. 2, pp. 168-176, 2021, https://doi.org/10.21831/reid.v7i2.43196.

[11] M. A. Hidayah and F. A. Setiawati, "Developing a character assessment instrument based on the school culture," *REID (Research and Evaluation in Education)*, vol. 8, no. 2, pp. 90-99, 2022, https://doi.org/10.21831/reid.v8i2.46802.

[12] U. Faizah, D. Zuchdi, and Y. Alsamiri, "An authentic assessment model to assess kindergarten students' character," *REID (Research and Evaluation in Education)*, vol. 5, no. 2, pp. 103-119, 2019, https://doi.org/10.21831/reid.v5i2.24588.

[13] D. Hadiana, B. Hayat, and B. Tola, "Comparison of methods for detecting anomalous behavior on large-scale computer-based exams based on response time and responses," *REID (Research and Evaluation in Education)*, vol. 6, no. 2, pp. 87-97, 2020, https://doi.org/10.21831/reid.v6i2.31260.

[14] D. S. Ciptaningrum, N. H. P. S. Putro, N. K. Sari, and N. Hasanuddin, "Evaluation of learning process: Knowledge of ICT integration among pre-service English language teachers," *REID (Research and Evaluation in Education)*, vol. 7, no. 1, pp. 46-56, 2021, https://doi.org/10.21831/reid.v7i1.30521.

[15] I. Ismiyati, H. Retnawati, S. Suranto, H. Haryanto, M. Sholihah, and T. Tusyanah, "The readiness of prospective teachers based on online teaching competencies and learning activities in COVID-19 pandemic: A cluster analysis-based approach," *REID (Research and Evaluation in Education)*, vol. 8, no. 2, pp. 127-139, 2022, https://doi.org/10.21831/reid.v8i2.45576.

[16] B. H. See, S. Gorard, B. Lu, L. Dong, and N. Siddiqui, "Is technology always helpful?: A critical review of the impact on learning outcomes of education technology in supporting formative assessment in schools," *Research Papers in Education*, vol. 37, no. 6, pp. 1064-1096, 2022, https://doi.org/10.1080/02671522.2021.1907778.

[17] K. J. Carstens, J. M. Mallon, M. Bataineh, and A. Al-Bataineh, "Effects of technology on student learning," *Turkish Online Journal of Educational Technology*, vol. 20, no. 1, pp. 105–113, 2021, https://eric.ed.gov/?id=EJ1290791.

[18] H. Akram, A. H. Abdelrady, A. S. Al-Adwan, and M. Ramzan, "Teachers' perceptions of technology integration in teaching-learning practices: A systematic review," *Frontiers in Psychology*, vol. 13, pp. 1-9, 2022, https://doi.org/10.3389/fpsyg.2022.920317.

[19] J. Arlinwibowo, H. Retnawati, and B. Kartowagiran, "The impact of ICT utilization to improve the learning outcome: A meta-analysis," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 11, no. 2, pp. 522-531, 2022, https://doi.org/10.11591/ijere.v11i2.22112.

[20] Á. A. Jiménez Sierra, J. M. Ortega Iglesias, J. Cabero-Almenara, and A. Palacios-Rodríguez, "Development of the teacher's technological pedagogical content knowledge (TPACK) from the Lesson Study: A systematic review," *Frontiers in Education*, vol. 8, pp. 1–11, 2023, https://doi.org/10.3389/feduc.2023.1078913.

[21] Y. D. Kristanto, "Technology-enhanced pre-instructional peer assessment: Exploring students' perceptions in a statistical methods course," *REID (Research and Evaluation in Education)*, vol. 4, no. 2, pp. 105-116, 2018, https://doi.org/10.21831/reid.v4i2.20951.

[22] P. Sivananda and A. A. Aziz, "Utilizing technology to promote active learning: A systematic literature review," *International Journal of Academic Research in Progressive Education and Development*, vol. 10, no. 3, pp. 784-803, 2021, https://doi.org/10.6007/IJARPED/v10-i3/10815.

[23] S. Timotheou *et al.*, "Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review," *Education and Information Technologies*, vol. 28, pp. 6695-6726, 2023, https://doi.org/10.1007/s10639-022-11431-8.

[24] H. Setiawan and S. Phillipson, "The effectiveness of game-based science learning (GBSL) to improve students' academic achievement: A meta-analysis of current research from 2010 to 2017," *REID (Research and Evaluation in Education)*, vol. 5, no. 2, pp. 152-168, 2019, https://doi.org/10.21831/reid.v5i2.28073.

[25] C. A. Purdescu, "The quantification of the time saved by the professors through the introduction of the electronic evaluation," *International Conference of Management and Industrial Engineering*, vol. 11, pp. 175-182, 2023, https://doi.org/10.56177/11icmie2023.33.

[26] H. Retnawati, "Learning trajectory of item response theory course using multiple softwares," *Olympiads in Informatics*, vol. 11, pp. 123-142, 2017, https://doi.org/10.15388/ioi.2017.10.

[27] M. M. Neumann, J. L. Anthony, N. A. Erazo, and D. L. Neumann, "Assessment and technology: Mapping future directions in the early childhood classroom," *Frontiers in Education*, vol. 4, p. 116, 2019, https://doi.org/10.3389/feduc.2019.00116.

[28] E. Eliaumra, D. P. Samaela, and N. K. Muhdin, "Developing diagnostic test assessment to measure creative thinking skills of Biology preservice teacher students," *REID (Research and Evaluation in Education)*, vol. 8, no. 2, pp. 152-168, 2022, https://doi.org/10.21831/reid.v8i2.50885.

[29] T. Fan, J. Song, and Z. Guan, "Integrating diagnostic assessment into curriculum: A theoretical framework and teaching practices," *Language Testing in Asia*, vol. 11, no. 1, pp. 1-23, 2021, https://doi.org/10.1186/s40468-020-00117-y.

[30] C. N. Blundell, "Teacher use of digital technologies for school-based assessment: A scoping review," *Assessment in Education: Principles, Policy & Practice*, vol. 28, no. 3, pp. 279-300, 2021, https://doi.org/10.1080/0969594X.2021.1929828.

[31] N. H. C. Hashim and K. Osman, "Teaching and learning by using online application during movement control order," *International Journal of Academic Research in Progressive Education and Development*, vol. 10, no. 2, pp. 605-614, 2021, https://doi.org/10.6007/IJARPED/v10-i2/10143.

[32] M. Muhardis, B. Tola, and H. Haribowo, "The respondent factors on the digital questionnaire responses," *REID (Research and Evaluation in Education)*, vol. 5, no. 2, pp. 144-151, 2019, https://doi.org/10.21831/reid.v5i2.26943.

[33] F. A. A. Maryo and E. Pujiastuti, "Gamification in EFL class using Quizizz as an assessment tool," *Proceedings of Digital Literacy in Education and Science*, vol. 3, pp. 75-80, 2022, https://doi.org/10.30595/pspfs.v3i.268.

[34] T. Hussein, M. Nat, H. F. Hasan, and A. Mahdi, "Development and evaluation of an online gamified assessment environment," *Proccedings of the 4th International Conference on Communication Engineering and Computer Science*, pp. 189–197, 2022, https://doi.org/10.24086/cocos2022/paper.744.

[35] P. Chakraborty, N. P. Kuruvatti and H. D. Schotten, "A novel serious game engineering based interactive visualization and evaluation platform for cellular technologies," *2017 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1-6, 2017, https://doi.org/10.1109/ISNCC.2017.8072012.

[36] A. F. C. D. Carmo, M. H. Shimabukuro, and E. H. D. Alcantara, "Using visual analytics techniques to evaluate the data quality in environmental datasets," *Boletim de Ciências Geodésicas*, vol. 22, no. 3, pp. 542-556, 2016, https://doi.org/10.1590/S1982-21702016000300031.

[37] P. Vittorini, S. Menini, and S. Tonelli, "An AI-based system for formative and summative assessment in data science courses," *International Journal of Artificial Intelligence in Education*, vol. 31, pp. 159-185, 2021, https://doi.org/10.1007/s40593-020-00230-2.

[38] N. H. Aswanti and W. Isnaeni, "Analysis of critical thinking skills, cognitive learning outcomes, and student activities in learning the human excretory system using an interactive flipbook," *REID (Research and Evaluation in Education)*, vol. 9, no. 1, pp. 37-48, 2023, https://doi.org/10.21831/reid.v9i1.53126.

[39] N. R. Hoover and L. M. Abrams, "Teachers' instructional use of summative student assessment data," *Applied Measurement in Education*, vol. 26, no. 3, pp. 219-231, 2013, https://doi.org/10.1080/08957347.2013.793187.

[40] M. Ulwatunnisa, H. Retnawati, M. Muhardis, and E. Yusron, "Revealing the characteristics of Indonesian language test used in the national-standardized school examinations," *REID (Research and Evaluation in Education)*, vol. 9, no. 2, pp. 210-222, 2023, https://doi.org/10.21831/reid.v9i2.31999.

[41] D. Chen, A. Jeng, S. Sun, and B. Kaptur, "Use of technology-based assessments: A systematic review covering over 30 countries," *Assessment in Education: Principles, Policy & Practice*, vol. 30, no. 5-6, pp. 396-428, 2023, https://doi.org/10.1080/0969594X.2023.2270181.

[42] O. T. Akintayo, C. A. Eden, O. O. Ayeni, and N. C. Onyebuchi, "Evaluating the impact of educational technology on learning outcomes in the higher education sector: A systematic review," *Open Access Research Journal of Multidisciplinary Studies*, vol. 7, no. 2, pp. 52-72, 2024, https://doi.org/10.53022/oarjms.2024.7.2.0026.

[43] K. Mangaroska, R. Tahir, M. Lorås and A. Mavroudi, "What do We Know about Learner Assessment in Technology-Rich Environments? A Systematic Review of Systematic Reviews," *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, pp. 16-20, 2018, https://doi.org/10.1109/ICALT.2018.00010.

[44] Q. Lin, Y. Yin, X. Tang, R. Hadad, and X. Zhai, "Assessing learning in technology-rich maker activities: A systematic review of empirical research," *Computers & Education*, vol. 157, p. 103944, 2020, https://doi.org/10.1016/j.compedu.2020.103944.

[45] C. Madland, V. Irvine, C. DeLuca, and O. Bulut, "Technology-integrated assessment: A literature review," *Open/Technology in Education, Society, and Scholarship Association Journal*, vol. 4, no. 1, pp. 1-48, 2024, https://doi.org/10.18357/otessaj.2024.4.1.57.

[46] L. S. Uman, "Systematic reviews and meta-analyses," *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, vol. 20, no. 1, pp. 57-59, 2011, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024725/.

[47] M. J. Page *et al*., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *bmj*, vol. 372, pp. 1-9, 2021, https://doi.org/10.1136/bmj.n71.

[48] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, and G. Lasa, "How-to conduct a systematic literature review: A quick guide for computer science research," *MethodsX*, vol. 9, pp. 1-12, 2022, https://doi.org/10.1016/j.mex.2022.101895.

[49] J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quantitative Science Studies*, vol. 1, no. 1, pp. 377-386, 2020, https://doi.org/10.1162/qss_a_00019.

[50] J. Strayer, "ERIC database alternatives and strategies for education researchers," *Reference Services Review*, vol. 36, no. 1, pp. 86-96, 2008, https://doi.org/10.1108/00907320810852050.

[51] T. Wright and S. Pullen, "Examining the literature: A bibliometric study of ESD journal articles in the Education Resources Information Center database," *Journal of Education for Sustainable Development*, vol. 1, no. 1, pp. 77–90, 2007, https://doi.org/10.1177/097340820700100114.

[52] T. Azungah, "Qualitative research: Deductive and inductive approaches to data analysis," *Qualitative Research Journal*, vol. 18, no. 4, pp. 383-400, 2018, https://doi.org/10.1108/qrj-d-18-00035.

[53] K. Krippendorff, "Reliability in content analysis: Some common misconceptions and recommendations," *Human Communication Research*, vol. 30, no. 3, pp. 411-433, 2004, https://doi.org/10.1111/j.1468-2958.2004.tb00738.x.

[54] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012, https://doi.org/10.11613/BM.2012.031.

[55] Z. Xu, B. Zhou, Z. Yang, X. Yuan, Y. Zhang, and Q. Lu, "NeatSankey: Sankey diagrams with improved readability based on node positioning and edge bundling," *Computers & Graphics*, vol. 113, pp. 10-20, 2023, https://doi.org/10.1016/j.cag.2023.05.001.

[56] A. H. Al-Maqbali and A. Al-Shamsi, "Assessment strategies in online learning environments during the COVID-19 pandemic in Oman," *Journal of University Teaching and Learning Practice*, vol. 20, no. 5, pp. 1-21, 2023, https://doi.org/10.53761/1.20.5.08.

[57] J. Andrews-Todd, J. Steinberg, M. Flor, and C. M. Forsyth, "Exploring automated classification approaches to advance the assessment of collaborative problem solving skills," *Journal of Intelligence*, vol. 10, pp. 1-24, 2022, https://doi.org/10.3390/jintelligence10030039.

[58] T. Bleckmann and G. Friege, "Concept maps for formative assessment: Creation and implementation of an automatic and intelligent evaluation method," *Knowledge Management & E-Learning*, vol. 15, no. 3, pp. 433-447, 2023, https://doi.org/10.34105/j.kmel.2023.15.025.

[59] Y.-H. Cheng, "E-portfolios in EFL writing: Benefits and challenges," *Language Education & Assessment*, vol. 5, no. 1, pp. 52-70, 2022, https://doi.org/10.29140/lea.v5n1.815.

[60] S. Chinjunthuk, P. Junpeng, and K. N. Tang, "Use of digital learning platform in diagnosing seventh grade students' mathematical ability levels," *Journal of Education and Learning*, vol. 11, no. 3, pp. 95-104, 2022, https://doi.org/10.5539/jel.v11n3p95.

[61] U. Durrani, O. Hujran, and A. S. Al-Adwan, "CrossQuestion game: A group-based assessment for gamified flipped classroom experience using the ARCS model," *Contemporary Educational Technology*, vol. 14, no. 2, pp. 1-15, 2022, https://doi.org/10.30935/cedtech/11568.

[62] F. Friyatmi, D. Mardapi, H. Haryanto, and E. Rahmi, "The development of computerized economics item banking for classroom and school-based assessment," *European Journal of Educational Research*, vol. 9, no. 1, pp. 293-303, 2020, https://doi.org/10.12973/eu-jer.9.1.293.

[63] D. González-Gómez, J. S. Jeong, and F. Cañada-Cañada, "Examining the effect of an online formative assessment tool (OFAT) of students' motivation and achievement for a university science education," *Journal of Baltic Science Education*, vol. 19, no. 3, pp. 401-414, 2020, https://doi.org/10.33225/jbse/20.19.401.

[64] H. Marlianawati, S. Suranto, and B. Kartowagiran, "Smartphone application for assessing teacher performance," *REID (Research and Evaluation in Education)*, vol. 9, no. 2, pp. 118-129, 2023, https://doi.org/10.21831/reid.v9i2.52384.

[65] S. Senel and H. C. Senel, "Use of take-home exam for remote assessment: A case study from Turkey," *Journal of Educational Technology and Online Learning*, vol. 4, no. 2, pp. 236-255, 2021, https://doi.org/10.31681/jetol.912965.

[66] E. J. Siwi, R. Anindyarini, and S. Nahar, "Item parameters of Yureka Education Center (YEC) English Proficiency Online Test (EPOT) instrument," *REID (Research and Evaluation in Education)*, vol. 6, no. 1, pp. 51-65, 2020, https://doi.org/10.21831/reid.v6i1.31013.

[67] I. Suhardi, "Alternative item selection strategies for improving test security in computerized adaptive testing of the algorithm," *REID (Research and Evaluation in Education)*, vol. 6, no. 1, pp. 32-40, 2020, https://doi.org/10.21831/reid.v6i1.30508.

[68] A. Tuluk and H. Yurdugul, "Design and development of a web based dynamic assessment system to increase students' learning effectiveness," *International Journal of Assessment Tools in Education*, vol. 7, no. 4, pp. 631-656, 2020, https://doi.org/10.21449/ijate.730454.

[69] D. Yao, "A comparative study of test takers' performance on computer-based test and paper-based test across different CEFR levels," *English Language Teaching*, vol. 13, no. 1, pp. 124-133, 2020, https://doi.org/10.5539/elt.v13n1p124.

[70] C. Cutting and K. Larkin, "The impact of weekly formative video feedback on pre-service teachers' experiences in online mathematics education," *Mathematics Teacher Education and Development*, vol. 23, no. 1, pp. 74-90, 2021, https://eric.ed.gov/?id=EJ1295257.

[71] G. T. L. Hoang, "Feedback precision and learners' responses: A study into ETS 'criterion' automated corrective feedback in EFL writing classrooms," *JALT CALL (Japan Association for Language Teaching Computer Assisted Language Learning) Journal*, vol. 18, no. 3, pp. 444–467, 2022, https://doi.org/10.29140/jaltcall.v18n3.775.

[72] J. S. Jeong, D. González-Gómez, and F. Yllana Prieto, "Sustainable and flipped STEM education: Formative assessment online interface for observing pre-service teachers' performance and motivation," *Education Sciences*, vol. 10, pp. 1-14, 2020, https://doi.org/10.3390/educsci10100283.

[73] S. Koltovskaia and S. Mahapatra, "Student engagement with computermediated teacher written corrective feedback: A case study," *JALT CALL (Japan Association for Language Teaching Computer Assisted Language Learning) Journal*, vol. 18, no. 2, pp. 286-315, 2022, https://doi.org/10.29140/jaltcall.v18n2.519.

[74] S. Kusairi, "A web-based formative feedback system development by utilizing isomorphic multiple choice items to support physics teaching and learning," *Journal of Technology and Science Education*, vol. 10, no. 1, pp. 117-126, 2020, https://doi.org/10.3926/jotse.781.

[75] M. Mohammadi, M. Zarrabi, and J. Kamali, "Formative assessment feedback to enhance the writing performance of Iranian IELTS candidates: Blending teacher and automated writing evaluation," *International Journal of Language Testing*, vol. 13, no. 1, pp. 206-224, 2023, https://doi.org/10.22034/ijlt.2022.364072.1201.

[76] M. Olivera-Aguilar, H.-S. Lee, A. Pallant, V. Belur, M. Mulholland, O. L. Liu, "Comparing the effect of contextualized versus generic automated feedback on students' scientific argumentation," *ETS Research Report Series*, vol. 22, no. 1, pp. 1-14, 2022, https://doi.org/10.1002/ets2.12344.

[77] A. Pásztor, A. Magyar, A. Pásztor-Kovács, and A. Rausch, "Online assessment and game-based development of inductive reasoning," *Journal of Intelligence*, vol. 10, no. 3, p. 59, 2022, https://doi.org/10.3390/jintelligence10030059.

[78] P. Senadheera and G. U. Kulasekara, "A formative assessment design suitable for online learning environments and its impact on students' learning," *Open Praxis*, vol. 13, no. 4, pp. 385-396, 2021, https://doi.org/10.55982/openpraxis.13.4.261.

[79] F. Toma, D. C. Diaconu, and C. M. Popescu, "The use of the Kahoot! learning platform as a type of formative assessment in the context of pre-university education during the COVID-19 pandemic period," *Education Sciences*, vol. 11, no. 10, p. 649, 2021, https://doi.org/10.3390/educsci11100649.

[80] I. Weir, R. Gwynllyw, and K. Henderson, "A case study in the e-assessment of statistics for nonspecialists," *Journal of University Teaching and Learning Practice*, vol. 18, no. 2, pp. 1–20, 2021, https://doi.org/10.53761/1.18.1.5.

[81] R. R. Wolf and A. B. Wolf, "Using AI to evaluate a competency-based online writing course in nursing," *Online Learning*, vol. 27, no. 3, pp. 41-69, 2023, https://doi.org/10.24059/olj.v27i3.3974.

[82] L. Xian, "The effectiveness of dynamic assessment in linguistic accuracy in EFL writing: An investigation assisted by online scoring systems," *Language Teaching Research Quarterly*, vol. 18, pp. 98–114, 2020, https://doi.org/10.32038/ltrq.2020.18.07.

[83] N. Yarahmadzehi and M. Goodarzi, "Investigating the role of formative mobile based assessment in vocabulary learning of pre-intermediate EFL learners in comparison with paper based assessment,"

Turkish Online Journal of Distance Education, vol. 21, no. 1, pp. 181-196, 2020, https://doi.org/10.17718/tojde.690390.

[84] S. S. Yilmaz, "Investigation of traditional and web 2.0 supported (a sample of kahoot) formative assessment and evaluation in science course," *Shanlax International Journal of Education*, vol. 11, no. S1, pp. 145-152, 2023, https://doi.org/10.34293/education.v11iS1-Oct.6689.

[85] A. S. Alharbi and Z. Meccawy, "Introducing socrative as a tool for formative assessment in saudi EFL classrooms," *Arab World English Journal*, vol. 11, no. 3, pp. 372-384, 2020, https://doi.org/10.24093/awej/vol11no3.23.

[86] P. Auphan, J. Ecalle, and A. Magnan, "The high potential of computer-based reading assessment," *Canadian Journal of Learning and Technology*, vol. 46, no. 1, pp. 1-23, 2020, https://doi.org/10.21432/cjlt27847.

[87] A. Barana, M. Marchisio, and M. Sacchet, "Interactive feedback for learning mathematics in a digital learning environment," *Education Sciences*, vol. 11, no. 6, p. 279, 2021, https://doi.org/10.3390/educsci11060279.

[88] P. Daniels, "Auto-scoring of student speech: Proprietary vs. open-source solutions," *TESL-EJ: The Electronic Journal for English as a Second Language*, vol. 26, no. 3, pp. 1-20, 2022, https://doi.org/10.55593/ej.26103int.

[89] H. Dasher and J. Pilgrim, "Paper vs. online assessments: A study of test-taking strategies for staar reading tests," *Texas Journal of Literacy Education*, vol. 9, no. 3, pp. 7-20, 2022, https://orcid.org/0000-0003-3407-8562.

[90] J. M. Faro *et al*., "Video-based communication assessment for weight management counseling training in medical residents: a mixed methods study," *BMC Medical Education*, vol. 22, no. 1, pp. 1–11, 2022, https://doi.org/10.1186/s12909-022-03984-6.

[91] I. Isnani, W. B. Utami, P. Susongko, and H. T. Lestiani, "Estimation of college students' ability on real analysis course using Rasch model," *REID (Research and Evaluation in Education)*, vol. 5, no. 2, pp. 95-102, 2019, https://doi.org/10.21831/reid.v5i2.20924.

[92] W. Kaiss, K. Mansouri, and F. Poirier, "Pre-evaluation with a personalized feedback conversational agent integrated in Moodle," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 6, pp. 177-189, 2023, https://doi.org/10.3991/ijet.v18i06.36783.

[93] M. Lubrick and B. Wellington, "Formative learning assessment with online quizzing: Comparing target performance grade and best performance grade approaches," *Journal of Learning and Teaching in Digital Age*, vol. 7, no. 2, pp. 297-306, 2022, https://doi.org/10.53850/joltida.1036295.

[94] S. H. Mercer and J. E. Cannon, "Validity of automated learning progress assessment in english written expression for students with learning difficulties," *Journal for Educational Research Online*, vol. 14, no. 1, pp. 39-60, 2022, https://doi.org/10.31244/jero.2022.01.03.

[95] L. G. Otaya, B. Kartowagiran, H. Retnawati, and S. S. Mustakim, "Estimating the ability of pre-service and in-service Teacher Profession Education (TPE) participants using item response theory," *REID (Research and Evaluation in Education)*, vol. 6, no. 2, pp. 160-173, 2020, https://doi.org/10.21831/reid.v6i2.36043.

[96] R. Phoophuangpairoj and P. Pipattarasakul, "Preliminary indicators of EFL essay writing for teachers' feedback using automatic text analysis," *International Journal of Educational Methodology*, vol. 8, no. 1, pp. 55-68, 2022, https://doi.org/10.12973/ijem.8.1.55.

[97] M. D. Sahin and S. Gelbal, "Development of a multidimensional computerized adaptive test based on the bifactor model," *International Journal of Assessment Tools in Education*, vol. 7, no. 3, pp. 323-342, 2020, https://doi.org/10.21449/ijate.707199.

[98] P. Zhao, C.-Y. Chang, Y. Shao, Z. Liu, H. Zhou, and J. Liu, "Utilization of process data in China: Exploring students' problem-solving strategies in computer-based science assessment featuring interactive tasks," *Journal of Baltic Science Education*, vol. 22, no. 5, pp. 929-944, 2023, https://doi.org/10.33225/jbse/23.22.929.

[99] S. Abdullah, W. Warsiyah, and J. Ju'subaidi, "Developing a religiosity scale for Indonesian Muslim youth," *REID (Research and Evaluation in Education)*, vol. 9, no. 1, pp. 73-85, 2023, https://doi.org/10.21831/reid.v9i1.61201.

[100] B. A. Dunya, C. McKown, and E. Smith, "Psychometric properties and differential item functioning of a web-based assessment of children's emotion recognition skill," *Journal of Psychoeducational Assessment*, vol. 38, no. 5, pp. 627-641, 2020, https://doi.org/10.1177/0734282919881919.

[101] I. K. Amalina and T. Vidákovich, "An integrated STEM-based mathematical problem-solving test: Developing and reporting psychometric evidence," *Journal on Mathematics Education*, vol. 13, no. 4, pp. 587-604, 2022, https://doi.org/10.22342/jme.v13i4.pp587-604.

[102] M. Berlian, I. M. Mujtahid, R. Vebrianto, and M. Thahir, "Multiple intelligences instrument development: Identification system of multiple intelligences tutor," *REID (Research and Evaluation in Education)*, vol. 6, no. 2, pp. 119-129, 2020, https://doi.org/10.21831/reid.v6i2.35120.

[103] R. W. Daryono, V. L. Hariyanto, H. Usman, and S. Sutarto, "Factor analysis: Competency framework for measuring student achievements of architectural engineering education in Indonesia," *REID (Research and Evaluation in Education)*, vol. 6, no. 2, pp. 98-108, 2020, https://doi.org/10.21831/reid.v6i2.32743.

[104] H. H. Dewi, S. M. Damio, and S. Sukarno, "Item analysis of reading comprehension questions for English proficiency test using Rasch model," *REID (Research and Evaluation in Education)*, vol. 9, no. 1, pp. 24-36, 2023, https://doi.org/10.21831/reid.v9i1.53514.

[105] R. R. Fardhila and E. Istiyono, "An assessment instrument of mind map product to assess students' creative thinking skill," *REID (Research and Evaluation in Education)*, vol. 5, no. 1, pp. 41-53, 2019, https://doi.org/10.21831/reid.v5i1.22525.

[106] S. Farida and F. A. Setiawati, "Developing assessment instruments of debate practice in Indonesian Language learning," *REID (Research and Evaluation in Education)*, vol. 7, no. 2, pp. 145-155, 2021, https://doi.org/10.21831/reid.v7i2.43338.

[107] K. N. Fathiyah, A. Alsa, and D. Setiyawati, "Psychometric characteristic of positive affect scale within the academic setting," *REID (Research and Evaluation in Education)*, vol. 5, no. 2, pp. 120-129, 2019, https://doi.org/10.21831/reid.v5i2.25992.

[108] I. W. Gunartha, T. Sulaiman, S. P. Suardiman, and B. Kartowagiran, "Developing instruments for measuring the level of early childhood development," *REID (Research and Evaluation in Education)*, vol. 6, no. 1, pp. 1-9, 2020, https://doi.org/10.21831/reid.v6i1.21996.

[109] R. M. Lia, A. Rusilowati, and W. Isnaeni, "NGSS-oriented chemistry test instruments: Validity and reliability analysis with the Rasch model," *REID (Research and Evaluation in Education)*, vol. 6, no. 1, pp. 41-50, 2020, https://doi.org/10.21831/reid.v6i1.30112.

[110] M. Muchlisin, D. Mardapi, and F. A. Setiawati, "An analysis of Javanese language test characteristic using the Rasch model in R program," *REID (Research and Evaluation in Education)*, vol. 5, no. 1, pp. 61-74, 2019, https://doi.org/10.21831/reid.v5i1.23773.

[111] A. Nurrahman, S. Sukirno, D. S. Pratiwi, J. Iskandar, A. Rahim, and I. S. Rahmaini, "Developing student social attitude self-assessment instruments: A study in vocational high school," *REID (Research and Evaluation in Education)*, vol. 8, no. 1, pp. 1-12, 2022, https://doi.org/10.21831/reid.v8i1.45100.

[112] A. Setiawan, W. Cendana, M. Ayres, A. A. Yuldashev, and S. P. Setyawati, "Development and validation of a self-assessment-based instrument to measure elementary school students' attitudes in online learning," *REID (Research and Evaluation in Education)*, vol. 9, no. 2, pp. 184-197, 2023, https://doi.org/10.21831/reid.v9i2.52083.

[113] J. Subando, M. K. B. Wibowo, and F. Farkhani, "The development of measurement instruments of Sharia students' perceptions about Khilafah," *REID (Research and Evaluation in Education)*, vol. 9, no. 2, pp. 141-155, 2023, https://doi.org/10.21831/reid.v9i2.63966.

[114] S. Sumin, F. Sukmawati, and N. Nurdin, "Gender differential item functioning on the Kentucky Inventory of Mindfulness Skills instrument using logistic regression," *REID (Research and Evaluation in Education)*, vol. 8, no. 1, pp. 55-66, 2022, https://doi.org/10.21831/reid.v8i1.50809.

[115] Y. P. Susani, G. R. Rahayu, Y. S. Prabandari, R. Sanusi, and H. Mardiwiyoto, "Developing an instrument to measure student's perception of the medical education curriculum from the perspective of Communities of Practice theory," *REID (Research and Evaluation in Education)*, vol. 6, no. 2, pp. 109-118, 2020, https://doi.org/10.21831/reid.v6i2.31500.

[116] F. Falcão, D. M. Pereira, N. Gonçalves, A. De Champlain, P. Costa, and J. M. Pêgo, "A suggestive approach for assessing item quality, usability and validity of automatic item generation," *Advances in Health Sciences Education*, vol. 28, no. 5, pp. 1441-1465, 2023, https://doi.org/10.1007/s10459-023-10225-y.

[117] 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, pp. 113-128, 2021, https://doi.org/10.3991/ijet.v16i18.23841.

[118] J. Renes, C. P. M. Van Der Vleuten, and C. F. Collares, "Utility of a multimodal computer-based assessment format for assessment with a higher degree of reliability and validity," *Medical Teacher*, vol. 45, no. 4, pp. 433-441, 2023, https://doi.org/10.1080/0142159X.2022.2137011.

[119] D. G. H. Divayana, I. G. Sudirtha, and I. K. Suartama, "Digital test instruments based on wondershare-superitem for supporting distance learning implementation of assessment course," *International Journal of Instruction*, vol. 14, no. 4, pp. 945-964, 2021, https://doi.org/10.29333/iji.2021.14454a.

[120] M. Zagaar and W. Chen, "Assessment for deeper understanding using concept maps: Lessons learned from flipped teaching of pharmacology," *Medical Science Educator*, vol. 32, no. 6, pp. 1289-1297, 2022, https://doi.org/10.1007/s40670-022-01653-3.

[121] M. Reina *et al*., "PLATA: Design of an online platform for chemistry undergraduate fully automated assignments," *Journal of Chemical Education*, vol. 101, no. 3, pp. 1024–1035, 2024, https://doi.org/10.1021/acs.jchemed.3c00962.

[122] Y. Karay, B. Reiss, and S. K. Schauber, "Progress testing anytime and anywhere – Does a mobile-learning approach enhance the utility of a large-scale formative assessment tool?," *Medical Teacher*, vol. 42, no. 10, pp. 1154-1162, 2020, https://doi.org/10.1080/0142159X.2020.1798910.

[123] J. Fuentes-Cimma *et al*., "Utility analysis of an adapted Mini-CEX WebApp for clinical practice assessment in physiotherapy undergraduate students," *Frontiers in Education*, vol. 8, pp. 1-10, 2023, https://doi.org/10.3389/feduc.2023.943709.

[124] C. Eitemüller, F. Trauten, M. Striewe, and M. Walpuski, "Digitalization of multistep chemistry exercises with automated formative feedback," *Journal of Science Education and Technology*, vol. 32, no. 3, pp. 453-467, 2023, https://doi.org/10.1007/s10956-023-10043-2.

[125] C.-Y. Chou and N.-B. Zou, "An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models," *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, pp. 1-27, 2020, https://doi.org/10.1186/s41239-020-00233-y.

[126] B. Seipel, P. C. Kennedy, S. E. Carlson, V. Clinton-Lisell, and M. L. Davison, "MOCCA-College: Preliminary Validity Evidence of a Cognitive Diagnostic Reading Comprehension Assessment," *Journal of Learning Disabilities*, vol. 56, no. 1, pp. 58-71, 2023, https://doi.org/10.1177/00222194221121340.

[127] W. Yuan *et al*., "Improving the resident assessment process: application of App-based e-training platform and lean thinking," *BMC Medical Education*, vol. 23, no. 1, pp. 1-9, 2023, https://doi.org/10.1186/s12909-023-04118-2.

[128] E. Rowe *et al*., "Interactive Assessments of CT (IACT): Digital Interactive Logic Puzzles to Assess Computational Thinking in Grades 3-8," *International Journal of Computer Science Education in Schools*, vol. 5, no. 2, pp. 28-73, 2021, https://doi.org/10.21585/ijcses.v5i1.149.

[129] D. Guzman-Orth, Y. Song, and J. R. Sparks, "Designing accessible formative assessment tasks to measure argumentation skills for english learners," *ETS Research Report Series*, vol. 2019, no. 1, pp. 1–15, 2019, https://doi.org/10.1002/ets2.12251.

[130] M. H. Dlab, S. Candrlic, and M. Pavlic, “Formative assessment activities to advance education: A case study,” *Journal of Information Technology Education: Innovations in Practice*, vol. 20, pp. 37–57, 2021, https://doi.org/10.28945/4758.

[131] S. Seifert and L. Paleczek, “Digitally assessing text comprehension in grades 3-4: Test development and validation,” *Electronic Journal of e-Learning*, vol. 19, no. 5, pp. 336-348, 2021, https://doi.org/10.34190/ejel.19.5.2467.

[132] C. Isler and B. Aydin, “Developing and validating a computerized oral proficiency test of english as a foreign language (COPTEFL),” *International Journal of Assessment Tools in Education*, vol. 8, no. 1, pp. 38-66, 2021, https://doi.org/10.21449/ijate.854678.

[133] A. Suryadi and S. Kusairi, “Developing computer-assisted formative feedback in the light of resource theory: A case on heat concept,” *Journal of Technology and Science Education*, vol. 11, no. 2, pp. 343–356, 2021, https://doi.org/10.3926/jotse.1100.

[134] F. Molin, C. Haelermans, S. Cabus, and W. Groot, “Do feedback strategies improve students’ learning gain?-Results of a randomized experiment using polling technology in physics classrooms,” *Computers & Education*, vol. 175, p. 104339, 2021, https://doi.org/10.1016/j.compedu.2021.104339.

[135] M. B. Ada, “Evaluation of a mobile web application for assessment feedback,” *Technology, Knowledge and Learning*, vol. 28, pp. 23-46, 2023, https://doi.org/10.1007/s10758-021-09575-6.

[136] S. H. P. W. Gamage, J. R. Ayres, M. B. Behrend, and E. J. Smith, “Optimising Moodle quizzes for online assessments,” *International Journal of STEM Education*, vol. 6, no. 1, p. 27, 2019, https://doi.org/10.1186/s40594-019-0181-4.

[137] I. López-Tocón, “Moodle quizzes as a continuous assessment in higher education: An exploratory approach in physical chemistry,” *Education Sciences*, vol. 11, pp. 1–12, 2021, https://doi.org/10.3390/educsci11090500.

[138] S. W. Widyaningsih, I. Yusuf, Z. K. Prasetyo, and E. Istiyono, “The development of the HOTS test of physics based on modern test theory: Question modeling through e-learning of Moodle LMS,” *International Journal of Instruction*, vol. 14, no. 4, pp. 51–68, 2021, https://doi.org/10.29333/iji.2021.1444a.

[139] A. Barana and M. M. Conte, “Promoting socioeconomic equity through automatic formative assessment,” *Journal on Mathematics Education*, vol. 15, no. 1, pp. 227–252, 2024, https://doi.org/10.22342/jme.v15i1.pp227-252.

[140] B. Duffy, R. Tully, and A. V. Stanton, “An online case-based teaching and assessment program on clinical history-taking skills and reasoning using simulated patients in response to the COVID-19 pandemic,” *BMC Medical Education*, vol. 23, no. 1, p. 4, 2023, https://doi.org/10.1186/s12909-022-03950-2.

[141] K. Khalaf, M. El-Kishawi, M. A. Moufti, and S. Al Kawas, “Introducing a comprehensive high-stake online exam to final-year dental students during the COVID-19 pandemic and evaluation of its effectiveness,” *Medical Education Online*, vol. 25, no. 1, pp. 1–10, 2020, https://doi.org/10.1080/10872981.2020.1826861.

[142] N. Mdlalose, S. Ramaila, and U. Ramnarain, “Using Kahoot! as a formative assessment tool in science teacher education,” *International Journal of Higher Education*, vol. 11, no. 2, pp. 43–51, 2022, https://doi.org/10.5430/ijhe.v11n2p43.

[143] K. N. A. Al-Mwzaiji and A. A. F. Alzubi, “Online self-evaluation: the EFL writing skills in focus,” *Asian-Pacific Journal of Second and Foreign Language Education*, vol. 7, no. 1, pp. 1–16, 2022, https://doi.org/10.1186/s40862-022-00135-8.

[144] W. J. A. J. Hendriks *et al*., “Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores,” *BMC Medical Education*, vol. 19, no. 1, pp. 1–11, 2019, https://doi.org/10.1186/s12909-019-1610-2.

[145] M. Muftah, F. A. Y. Al-Inbari, B. Q. Al-Wasy, and H. S. Mahdi, “The role of automated corrective feedback in improving EFL learners’ mastery of the writing aspects,” *Psycholinguistics*, vol. 34, no. 2, pp. 82-109, 2023, https://doi.org/10.31470/2309-1797-2023-34-2-82-109.

[146] Y. Qian and J. D. Lehman, “Using targeted feedback to address common student misconceptions in introductory programming: A data-driven approach,” *SAGE Open*, vol. 9, no. 4, pp. 1–20, 2019, https://doi.org/10.1177/2158244019885136.

[147] M. F. Areed, M. A. Amasha, R. A. Abougalala, S. Alkhalaf, and D. Khairy, “Developing gamification e-quizzes based on an android app: The impact of asynchronous form,” *Education and Information Technologies*, vol. 26, no. 4, pp. 4857–4878, 2021, https://doi.org/10.1007/s10639-021-10469-4.

[148] U. Schepke, M. E. Van Wulfften Palthe, E. W. Meisberger, W. Kerdijk, M. S. Cune, and B. Blok, “Digital assessment of a retentive full crown preparation—An evaluation of prepCheck in an undergraduate pre-clinical teaching environment,” *European Journal of Dental Education*, vol. 24, no. 3, pp. 407–424, 2020, https://doi.org/10.1111/eje.12516.

[149] A. Mizumoto, Y. Sasao, and S. A. Webb, “Developing and evaluating a computerized adaptive testing version of the Word Part Levels Test,” *Language Testing*, vol. 36, no. 1, pp. 101–123, 2019, https://doi.org/10.1177/0265532217725776.

[150] W.-H. Chuo *et al.*, “Evaluate the feasibility of the implementation of e-assessment in objective structured clinical examination (OSCE) in pharmacy education from the examiner’s perspectives,” *Education Sciences*, vol. 11, pp. 1–14, 2021, https://doi.org/10.3390/educsci11050194.

[151] H. Çeliktas and R. E. Demirbatir, “Effect of online quizzes on music theory achievement of freshman music teaching students,” *Journal of Education and Learning (EduLearn)*, vol. 16, no. 1, pp. 130–136, 2022, https://doi.org/10.11591/edulearn.v16i1.20379.

[152] Z. Ç. Köroglu, “Using digital formative assessment to evaluate EFL learners’ english speaking skills,” *GIST Education and Learning Research Journal*, vol. 22, pp. 103–123, 2021, https://doi.org/10.26817/16925777.1001.

[153] A. A. Lopez, D. Guzman-Orth, D. Zapata-Rivera, C. M. Forsyth, and C. Luce, “Examining the accuracy of a conversation-based assessment in interpreting english learners’ written responses,” *ETS Research Report Series*, vol. 2021, no. 1, pp. 1–15, 2021, https://doi.org/10.1002/ets2.12315.

[154] T. de Lange, A. Møystad, and G. Torgersen, “How can video-based assignments integrate practical and conceptual knowledge in summative assessment? Student experiences from a longitudinal experiment,” *British Educational Research Journal*, vol. 46, no. 6, pp. 1279–1299, 2020, https://doi.org/10.1002/berj.3632.

[155] A. P. Goodwin *et al*., “Monster, P.I.: Validation evidence for an assessment of adolescent language that assesses vocabulary knowledge, morphological knowledge, and syntactical awareness,” *Assessment for Effective Intervention*, vol. 47, no. 2, pp. 89–100, 2022, https://doi.org/10.1177/1534508420966383.

[156] E. Istiyono, W. S. B. Dwandaru, R. Setiawan, and I. Megawati, “Developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use,” *European Journal of Educational Research*, vol. 9, no. 1, pp. 91–101, 2020, https://doi.org/10.12973/eu-jer.9.1.91.

[157] S. Hopkins and R. O’Donovan, “Developing assessments for students with intellectual disability to support differentiation,” *Mathematics Teacher Education and Development*, vol. 23, no. 3, pp. 132–147, 2021, http://orcid.org/0000-0002-5826-0193.

[158] C. Conner, A. R. Henry, E. J. Solari, and M. C. Zajic, “Conducting oral and written language adapted tele-assessments with early elementary-age children with autism spectrum disorder,” *Autism & Developmental Language Impairments*, vol. 7, pp. 1–15, 2022, https://doi.org/10.1177/23969415221133268.

[159] C.-A. Lee, N.-F. Huang, J.-W. Tzeng, and P.-H. Tsai, “AI-based diagnostic assessment system: Integrated with knowledge map in MOOCs,” *IEEE Transactions on Learning Technologies*, vol. 16, no. 5, pp. 873–886, 2023, https://doi.org/10.1109/TLT.2023.3308338.

[160] S. Radović, N. Seidel, J. M. Haake, and R. Kasakowskij, “Analysing students’ self-assessment practice in a distance education environment: Student behaviour, accuracy, and task-related characteristics,” *Journal of Computer Assisted Learning*, vol. 40, no. 2, pp. 654–666, 2024, https://doi.org/10.1111/jcal.12907.

[161] M. K. Wolf and A. A. Lopez, "Developing a technology-based classroom assessment of academic reading skills for english language learners and teachers: Validity evidence for formative use," *Languages*, vol. 7, no. 2, pp. 1–22, 2022, https://doi.org/10.3390/languages7020071.

[162] J. R. Rico-Juan, V. M. Sánchez-Cartagena, J. J. Valero-Mas, and A. J. Gallego, "Identifying student profiles within online judge systems using explainable artificial intelligence," *IEEE Transactions on Learning Technologies*, vol. 16, no. 6, pp. 955–969, 2023, https://doi.org/10.1109/TLT.2023.3239110.

[163] D. Di Mitri, J. Schneider, and H. Drachsler, "Keep me in the loop: Real-time feedback with multimodal data," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 4, pp. 1093–1118, 2022, https://doi.org/10.1007/s40593-021-00281-z.

[164] J. P. Bernius, S. Krusche, and B. Bruegge, "Machine learning based feedback on textual student answers in large courses," *Computers and Education: Artificial Intelligence*, vol. 3, pp. 1–16, 2022, https://doi.org/10.1016/j.caeai.2022.100081.

[165] M. Almasre, "Development and evaluation of a custom GPT for the assessment of students' designs in a typography course," *Education Sciences*, vol. 14, no. 2, pp. 1–19, 2024, https://doi.org/10.3390/educsci14020148.

[166] A. Darvishi, H. Khosravi, S. Sadiq, and D. Gašević, "Incorporating AI and learning analytics to build trustworthy peer assessment systems," *British Journal of Educational Technology*, vol. 53, no. 4, pp. 844–875, 2022, https://doi.org/10.1111/bjet.13233.

[167] L. Kaldaras, N. R. Yoshida, and K. C. Haudek, "Rubric development for AI-enabled scoring of three-dimensional constructed-response assessment aligned to NGSS learning progression," *Frontiers in Education*, vol. 7, p. 983055, 2022, https://doi.org/10.3389/feduc.2022.983055.

[168] M. T. Nagy and E. Korom, "Measuring scientific reasoning of fourth graders: Validation of the Science-K inventory in paper-based and computer-based testing environments," *Journal of Baltic Science Education*, vol. 22, no. 6, pp. 1050–1062, 2023, https://doi.org/10.33225/jbse/23.22.1050.

[169] T. Simon, I. Biró, and A. Kárpáti, "Developmental assessment of visual communication skills in primary education," *Journal of Intelligence*, vol. 10, pp. 1–20, 2022, https://doi.org/10.3390/jintelligence10030045.

[170] S. Varga, A. Pásztor, and J. Stekács, "Online assessment of morphological awareness in grades 2-4: Its development and relation to reading comprehension," *Journal of Intelligence*, vol. 10, no. 47, pp. 1–19, 2022, https://doi.org/10.3390/jintelligence10030047.

[171] K.-C. Li, M. Chang, and K.-H. Wu, "Developing a task-based dialogue system for english language learning," *Education Sciences*, vol. 10, no. 11, pp. 1–20, 2020, https://doi.org/10.3390/educsci10110306.

[172] A. L. C. Barczak, A. Mathrani, B. Han, and N. H. Reyes, "Automated assessment system for programming courses: A case study for teaching data structures and algorithms," *Educational technology research and development*, vol. 71, no. 6, pp. 2365–2388, 2023, https://doi.org/10.1007/s11423-023-10277-2.

[173] L. Zheng, M. Long, B. Chen, and Y. Fan, "Promoting knowledge elaboration, socially shared regulation, and group performance in collaborative learning: an automated assessment and feedback approach based on knowledge graphs," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, pp. 1–20, 2023, https://doi.org/10.1186/s41239-023-00415-4.

[174] S. Kusairi, D. Anggita Puspita, A. Suryadi, and H. Suwono, "Physics formative feedback game: Utilization of isomorphic multiple-choice items to help students learn kinematics," *TEM Journal*, vol. 9, no. 4, pp. 1625–1632, 2020, https://doi.org/10.18421/TEM94-39.

[175] X. Lv, L. Li, L. Guo, T. He, and S. Liu, "Game-based formative assessment of analogical reasoning in preschool children: Support from the internet of things technology," *Sustainability*, vol. 14, no. 21, pp. 1–17, 2022, https://doi.org/10.3390/su142113830.

[176] E. Rushton and S. Corrigan, "Game-assisted assessment for broader adoption: Participatory design and game-based scaffolding," *Electronic Journal of e-Learning*, vol. 19, no. 2, pp. 71–87, 2021, https://doi.org/10.34190/ejel.19.2.2143.

[177] A. J. King, J. M. Kahn, E. B. Brant, G. F. Cooper, and D. L. Mowery, "Initial development of an automated platform for assessing trainee performance on case presentations," *ATS Scholar*, vol. 3, no. 4, pp. 548–560, 2022, https://doi.org/10.34197/ats-scholar.2022-0010OC.

[178] E. Rusman and R. Nadolski, "Pe(e)rfectly skilled: Underpinnings of an online formative assessment method for (inter)active and practice-based complex skills training in higher education (HE)," *International Journal of Mobile and Blended Learning*, vol. 15, no. 2, pp. 1–14, 2023, https://doi.org/10.4018/ijmbl.318646.

[179] R. J. Nadolski, H. G. K. Hummel, E. Rusman, and K. Ackermans, "Rubric formats for the formative assessment of oral presentation skills acquisition in secondary education," *Educational Technology Research and Development*, vol. 69, no. 5, pp. 2663–2682, 2021, https://doi.org/10.1007/s11423-021-10030-7.

[180] S. van Ginkel *et al*., "Fostering oral presentation competence through a virtual reality-based task for delivering feedback," *Computers & Education*, vol. 134, pp. 78–97, 2019, https://doi.org/10.1016/j.compedu.2019.02.006.

[181] W. A. P. Udeshinee, O. Knutsson, and S. Männikkö-Barbutiu, "Text chat-mediated dynamic assessment towards self-regulation in language learning," *International Journal of Mobile and Blended Learning*, vol. 16, no. 1, pp. 1–21, 2023, https://doi.org/10.4018/ijmbl.335067.

[182] A. Sayin, S. Bozdag, and M. J. Gierl, "Automatic item generation for non-verbal reasoning items," *International Journal of Assessment Tools in Education*, vol. 10, pp. 131–147, 2023, https://doi.org/10.21449/ijate.1359348.

[183] A. Sayin and M. J. Gierl, "Automatic item generation for online measurement and evaluation: Turkish literature items," *International Journal of Assessment Tools in Education*, vol. 10, no. 2, pp. 218–231, 2023, https://doi.org/10.21449/ijate.1249297.

[184] I. Uysal and N. Dogan, "Automated essay scoring effect on test equating errors in mixed-format test," *International Journal of Assessment Tools in Education*, vol. 8, no. 2, pp. 222–238, 2021, https://doi.org/10.21449/ijate.815961.

[185] M. Park, "Effects of simulation-based formative assessments on students' conceptions in physics," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 15, no. 7, pp. 1–18, 2019, https://doi.org/10.29333/ejmste/103586.

[186] A. Mizumoto and M. Eguchi, "Exploring the potential of using an AI language model for automated essay scoring," *Research Methods in Applied Linguistics*, vol. 2, no. 2, p. 100050, 2023, https://doi.org/10.1016/j.rmal.2023.100050.

[187] S. H. Mercer, J. E. Cannon, B. Squires, Y. Guo, and E. Pinco, "Accuracy of automated written expression curriculum-based measurement scoring," *Canadian Journal of School Psychology*, vol. 36, no. 4, pp. 304–317, 2021, https://doi.org/10.1177/0829573520987753.

[188] O. T. Aricak, A. Avcu, F. Topçu, and M. G. Tutlu, "Use of item response theory to validate cyberbullying sensibility scale for university students," *International Journal of Assessment Tools in Education*, vol. 7, no. 1, pp. 18–29, 2020, https://doi.org/10.21449/ijate.629584.

[189] N. Cikrikci *et al*., "Development of a computerized adaptive version of the Turkish driving licence exam," *International Journal of Assessment Tools in Education*, vol. 7, no. 4, pp. 570–587, 2020, https://doi.org/10.21449/ijate.716177.

[190] M. Koch, F. M. Spinath, S. Greiff, and N. Becker, "Development and validation of the open matrices item bank," *Journal of Intelligence*, vol. 10, pp. 1–10, 2022, https://doi.org/10.3390/jintelligence10030041.

[191] A. Mohammed, A. K. S. Dawood, T. Alghazali, Q. K. Kadhim, A. A. Sabti, and S. H. Sabit, "A cognitive diagnostic assessment study of the reading comprehension section of the preliminary english test (PET)," *International Journal of Language Testing*, vol. 13, pp. 1-20, 2023, https://doi.org/10.22034/ijlt.2022.362849.1195.

[192] J. Seifried, S. Brandt, K. Kögler, and A. Rausch, "The computer-based assessment of domain-specific problem-solving competence—A three-step scoring procedure," *Cogent Education*, vol. 7, no. 1, p. 1719571, 2020, https://doi.org/10.1080/2331186X.2020.1719571.

[193] M.-J. Agost, P. Company, M. Contero, and J. D. Camba, "CAD training for digital product quality: A formative approach with computer-based adaptable resources for self-assessment," *International Journal of Technology and Design Education*, vol. 32, no. 2, pp. 1393–1411, 2022, https://doi.org/10.1007/s10798-020-09651-5.

[194] U. C. K. Al Muniandy, H. Zulnaidi, and S. H. Halili, "Validity and reliability of the situational motivational scale (SIMS) instrument: Using Rasch model and AMOS," *Malaysian Online Journal of Educational Sciences*, vol. 11, no. 1, pp. 34-46, 2023, https://ajba.um.edu.my/index.php/MOJES/article/view/41265/15463.

[195] M. Saefi *et al*., "Validating of knowledge, attitudes, and practices questionnaire for prevention of COVID-19 infections among undergraduate students: A Rasch and factor analysis," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 16, no. 12, pp. 1-14, 2020, https://doi.org/10.29333/ejmste/9352.

[196] M. A. Samsudin, T. S. Chut, M. E. Ismail, and N. J. Ahmad, "A calibrated item bank for computerized adaptive testing in measuring science TIMSS performance," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 16, no. 7, pp. 1–15, 2020, https://doi.org/10.29333/ejmste/8259.

[197] S. Soeharto, "Development of a diagnostic assessment test to evaluate science misconceptions in terms of school grades: A Rasch measurement aproach," *Journal of Turkish Science Education*, vol. 18, no. 3, pp. 351-370, 2021, https://doi.org/10.36681/tused.2021.78.

[198] D. Y. Ergül and M. F. Tasar, "Development and validation of the teachers' digital competence scale (TDiCoS)," *Journal of Learning and Teaching in Digital Age*, vol. 8, no. 1, pp. 148–160, 2023, https://doi.org/10.53850/joltida.1204358.

[199] A. S. Konca, O. Baltaci, and O. F. Akbulut, "Problematic technology use scale for young children (PTUS-YC): Validity and reliability study," *International Journal of Assessment Tools in Education*, vol. 9, no. 2, pp. 267-289, 2022, https://doi.org/10.21449/ijate.888936.

[200] J. D. G. Quinto, "Development and validation of survey instrument on game-based learning approach (SIGBLA)," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 15, pp. 233–242, 2022, https://doi.org/10.3991/ijet.v17i15.33267.

[201] S. Sovey, K. Osman, and M. E. E. Mohd-Matore, "Exploratory and confirmatory factor analysis for disposition levels of computational thinking instrument among secondary school students," *European Journal of Educational Research*, vol. 11, no. 2, pp. 639-652, 2022, https://doi.org/10.12973/eu-jer.11.2.639.

[202] M. P. Virtic, A. D. Plessis, and A. Šorgo, "Development and validation of the 'mentoring for effective teaching practicum instrument," *Center for Educational Policy Studies Journal*, vol. 13, no. 3, pp. 233–260, 2023, https://doi.org/10.26529/cepsj.1315.

[203] A. Rachmatullah *et al*., "Development and validation of the middle grades computer science concept inventory (MG-CSCI) assessment," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 16, no. 5, pp. 1-24, 2020, https://doi.org/10.29333/ejmste/116600.

[204] M. Polat, "Comparison of performance measures obtained from foreign language tests according to item response theory vs classical test theory," *International Online Journal of Education and Teaching*, vol. 9, no. 1, pp. 471-485, 2022, https://iojet.org/index.php/IOJET/article/view/1583.

[205] C. Eckerly, Y. Jia, and P. Jewsbury, "Technology-enhanced items and model – data misfit (Research Report No. RR-22-11)," *ETS Research Report Series*, vol. 22, no. 1, pp. 1–16, 2022, https://doi.org/10.1002/ets2.12353.

[206] L. Viskotová and D. Hampel, "Increasing the efficiency of teacher's work: The case of undergraduate mathematics mid-term assessment," *Mathematics Teaching Research Journal*, vol. 14, no. 2, pp. 88–104, 2022, https://files.eric.ed.gov/fulltext/EJ1350208.pdf.

[207] P. Gawliczek, V. Krykun, N. Tarasenko, M. Tyshchenko, and O. Shapran, “Computer adaptive language testing according to NATO STANAG 6001 requirements,” *Advanced Education*, vol. 17, pp. 19–26, 2021, https://doi.org/10.20535/2410-8286.225018.

[208] L. Kuklick and M. A. Lindner, “Affective-motivational effects of performance feedback in computer-based assessment: Does error message complexity matter?,” *Contemporary Educational Psychology*, vol. 73, pp. 1–14, 2023, https://doi.org/10.1016/j.cedpsych.2022.102146.

[209] K. A. Kroeze, S. M. Van Den Berg, B. P. Veldkamp, and T. De Jong, “Automated assessment of and feedback on concept maps during inquiry learning,” *IEEE Transactions on Learning Technologies*, vol. 14, no. 4, pp. 460-473, 2021, https://doi.org/10.1109/TLT.2021.3103331.

[210] T.-C. Hsu, Y.-S. Chang, M.-S. Chen, I.-F. Tsai, and C.-Y. Yu, “A validity and reliability study of the formative model for the indicators of STEAM education creations,” *Education and Information Technologies*, vol. 28, no. 7, pp. 8855–8878, 2023, https://doi.org/10.1007/s10639-022-11412-x.

[211] R. P. Chalmers, “mirt: A multidimensional item response theory package for the R environment,” *Journal of Statistical Software*, vol. 48, no. 6, pp. 1–29, 2012, https://doi.org/10.18637/jss.v048.i06.

[212] B. Boitshwarelo, A. K. Reedy, and T. Billany, “Envisioning the use of online tests in assessing twenty-first century learning: A literature review,” *Research and Practice in Technology Enhanced Learning*, vol. 12, no. 1, p. 16, 2017, https://doi.org/10.1186/s41039-017-0055-7.

[213] A. Surya and A. Aman, “Developing formative authentic assessment instruments based on learning trajectory for elementary school,” *REID (Research and Evaluation in Education)*, vol. 2, no. 1, pp. 13–24, 2016, https://doi.org/10.21831/reid.v2i1.6540.

[214] T. K. Gunning *et al*., “Who engaged in the team-based assessment? Leveraging EdTech for a self and intra-team peer-assessment solution to free-riding,” *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, pp. 1–22, 2022, https://doi.org/10.1186/s41239-022-00340-y.

[215] P. Karaman, “The effect of formative assessment practices on student learning: A meta-analysis study,” *International Journal of Assessment Tools in Education*, vol. 8, no. 4, pp. 801–817, 2021, https://doi.org/10.21449/ijate.870300.

[216] O. Ndayizeye, “Discrepancies in assessing undergraduates’ pragmatics learning,” *REID (Research and Evaluation in Education)*, vol. 3, no. 2, pp. 133–143, 2017, https://doi.org/10.21831/reid.v3i2.14487.

[217] J. D. Elicker, N. L. McConnell, “Interactive learning in the classroom: Is student response method related to performance?,” *Teaching of Psychology*, vol. 38, no. 3, pp. 147-150, 2011, https://doi.org/10.1177/0098628311411789.

[218] J. Willis and V. Klenowski, “Classroom assessment practices and teacher learning: An Australian perspective,” *Teacher learning with classroom assessment*, pp. 19-37, 2018, https://doi.org/10.1007/978-981-10-9053-0_2.

[219] J. Moss, S. Godinho, and E. Chao, “Enacting the Australian Curriculum: Primary and secondary teachers’ approaches to integrating the curriculum,” *Australian Journal of Teacher Education*, vol. 44, no. 3, pp. 24–41, 2019, https://doi.org/10.14221/ajte.2018v44n3.2.

[220] F. M. van der Kleij, J. J. Cumming, and A. Looney, “Policy expectations and support for teacher formative assessment in Australian education reform,” *Assessment in Education: Principles, Policy & Practice*, vol. 25, no. 6, pp. 620–637, 2018, https://doi.org/10.1080/0969594X.2017.1374924.

[221] I. T. C. A. T. Publishers, “Guidelines for technology-based assessment,” *International Test Commission and the Association of Test Publishers*, 2022, https://www.intestcom.org/upload/media-library/tba-guidelines-final-2-23-2023-v4-167785144642TgY.pdf

[222] U. Bezirhan and M. von Davier, “Automated reading passage generation with OpenAI’s large language model,” *Computers and Education: Artificial Intelligence*, vol. 5, pp. 1-13, 2023, https://doi.org/10.1016/j.caeai.2023.100161.

[223] O. Bulut *et al*., “The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges,” *Computers and Society*, 2024, https://doi.org/10.48550/arXiv.2406.18900.

[224] A. Khademi, "Can ChatGPT and Bard generate aligned assessment items? A reliability analysis against human performance," *Journal of Applied Learning & Teaching*, vol. 6, no. 1, pp. 75–80, 2023, https://doi.org/10.37074/jalt.2023.6.1.28.

[225] J. Rudolph, S. Tan, and S. Tan, "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?," *Journal of Applied Learning & Teaching*, vol. 6, no. 1, pp. 342–263, 2023, https://doi.org/10.37074/jalt.2023.6.1.9.

[226] H. Zhang, H. Song, S. Li, M. Zhou, and D. Song, "A survey of controllable text generation using transformer-based pre-trained language models," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1–37, 2024, https://doi.org/10.1145/3617680.

[227] I. O. Gallegos *et al*., "Bias and fairness in large language models: A survey," *Computation and Language*, 2023, https://doi.org/10.48550/arXiv.2309.00770.