



AMERICAN JOURNAL OF INTERDISCIPLINARY RESEARCH AND INNOVATION (AJIRI)

ISSN: 2833-2237 (ONLINE)

VOLUME 3 ISSUE 3 (2024)

**PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA**

Examining Accessibility, Credibility, and Accountability in Digital Assessment: A Systematic Literature Review

Richard Mulenga^{1*}, Allan Mwenya²,

Article Information

Received: July 01, 2024

Accepted: August 05, 2024

Published: August 08, 2024

Keywords

Digital Assessments, Assistive Technologies, Algorithmic Bias, Post COVID-19 Era

ABSTRACT

This paper examines the accessibility, credibility, and accountability of digital assessments through systematic literature reviews. Although the extant literature is replete with studies on educational digital assessments, it seems that, so far, no systematic review study has focused on re-examining the accessibility, credibility, and accountability of digital assessments in the pre-and post-COVID-19 period. The growing ubiquity of digital assessments in academic and professional contexts, especially in the post-COVID-19 era, makes it necessary to conduct this systematic review. The main finding of this study is that, despite the growing ubiquity of digital assessments post the COVID-19 crisis, digital assessments seem to be deficient in employing assistive technologies. Additionally, the sudden migration to digital assessments pose challenges of maintaining the same standards of validity and reliability commensurate with traditional in-person assessments. Therefore, going forward, we recommend extensive integration of assistive technologies in academic and professional digital assessments to enhance accessibility. Additionally, digital assessing authorities should also establish mechanisms for detecting cheating and plagiarism in digital assessments. This study underscores the overarching need for holistic approaches that balance technological innovation with ethical imperatives in digital assessments. This study contributes to the 21st-century understanding of the complex dynamic digital assessment landscapes.

INTRODUCTION

This study examines the accessibility, credibility, and accountability of digital assessments through systematic literature reviews for the period 2000-2023. The rapid integration of digital technologies in assessment practices has revolutionized global education and professional paradigms. It is irrefutable that the COVID-19 pandemic accelerated the adoption of digital platforms across various sectors, including education and professional assessments. Digital assessments have transformed the landscape of both professional and academic environments in this digital era, bringing a range of complexities and challenges. This paper re-examines the accessibility, credibility, and accountability of digital assessments for the period 2000 to 2023. The growing ubiquity of digital assessments in academic and professional contexts demands that their accessibility, credibility, and accountability be carefully examined (Usher, 2018; Dempere *et al.*, 2023). Although digital assessments have many benefits, including increased flexibility and efficiency, there are concerns that they could potentially exacerbate already-existing inequalities and produce biased results (Thompson, 2020; Dempere *et al.*, 2023). Digital assessments, which include computer-based exams, online tests, and e-portfolios, have many benefits. These benefits include ease of administration and scoring, flexibility in handling a variety of content, and the possibility of gathering extensive data. However, aside from the advantages, there are critical factors that digital assessors need to keep in mind: Accountability,

credibility, and accessibility (Lang & Etzkowitz, 2018; Slepankova, 2021; Lee & Kim, 2021).

Although the extant literature is replete with studies on educational digital assessments during and post-COVID-19 (Smith & Jones, 2020; Li & Lalani, 2020; Barbour *et al.*, 2020; Lee & Kim, 2021) as well as reviews on the effects of artificial intelligence (AI) on digital assessment (See, for example; Slepankova, 2021; Susnjak, 2022; Dempere *et al.*, 2023; Smith, 2023), the literature review seems to indicate that so far, no systematic review study has focused on the digital assessments' accessibility, credibility and accountability (validity) in both educational and professional settings with special focus on the Post-COVID-19 period. This review factors in the proliferation of online assessments in the backdrop of the use of AI (Pickard *et al.*, 2017; Fulmer *et al.*, 2018; Xiao *et al.*, 2020). The significance of undertaking this study is threefold. First, given the increasing prevalence of digital assessments in educational and professional settings post-COVID-19 pandemic, universal accessibility is of paramount importance. As more traditional assessments migrate to digital assessments, authorities must ensure that digital assessments are accessible to all cohorts of individuals or populations. This cohort must include individuals with various disabilities or impairments (Smith & Jones, 2020; Dempere *et al.*, 2023). Additionally, digital assessments should accommodate different learning styles to promote inclusivity (Brown *et al.*, 2018; Smith & Jones, 2020). Second, with the proliferation of online resources, it is vital that an examination of the

¹ ZCAS University, Dedan Kamathi Road, Box 35243, Lusaka, Zambia

² Examinations Council of Zambia (ECZ), Box 504432, Lusaka Zambia

* Corresponding author's e-mail: richardmulenga2@gmail.com

reliability and accountability of digital assessments is re-examined to ensure that digital assessments accurately measure the intended outcomes (Johnson & Wang, 2019; Susnjak, 2022). Lee and Kim (2021) succinctly assert that continuous research on digital assessment must be done to assess factors such as (assessment) validity, reliability, and fairness to ensure that the credibility of digital assessments in both educational and professional settings is maintained. The third reason for conducting this study borders on accountability. Accountability is essential in guaranteeing the integrity and fairness of digital assessments. This implies that educational and professional assessing entities must establish clear guidelines and procedures for developing, administering, and evaluating digital assessments to uphold accountability standards. In addition, mechanisms for detecting cheating and plagiarism in digital assessments must be implemented to ensure that academic and professional integrity is always upheld (García & López, 2020; Susnjak, 2022; Martínez & Pérez, 2022).

This study adopts a systematic literature review methodological approach of the extant literature on the themes of accessibility, credibility, and accountability factoring in the post-COVID-19 era. It is envisaged that going forward, this study will help improve the accessibility of digital assessments for diverse populations, enhance the integrity of digital assessments through robust design and implementation practices and promote fair and ethical digital assessments.

Research Questions

The following research questions guided our study:

- (i) How do current digital assessment practices impact the accessibility of assessments for diverse populations, especially in the post-COVID-19 era?
- (ii) What factors contribute to or undermine the credibility of digital assessments?
- (iii) How can accountability mechanisms be implemented to ensure fairness and ethical use of digital assessments in general and in the post-COVID-19 world in particular?

LITERATURE REVIEW

Systematic Review on the Accessibility of Digital Assessments for Diverse Populations

In this section, we systematically review how the integration of digital technologies or current digital assessment practices impact accessibility for diverse populations, including individuals with disabilities and non-native speakers of English, and how modern technological developments, such as the use of artificial intelligence impact accessibility, credibility and accountability in 21st-century digital assessments. The systematic review followed the order of research questions.

Review on the Accessibility of Digital Assessments for Diverse Populations

Digital technology integration offers both opportunities

and challenges for modern educational and professional assessment practices, especially when it comes to diverse populations, including people with disabilities. This literature review of relatively recent literature explores these challenges and proposes strategies for addressing them.

Smith *et al.* (2020) conducted a study on the review of accessibility considerations in educational digital assessments. They found that many digital assessments rely mainly on text-based formats, which systematically excludes individuals with visual impairments who rely on screen readers or enlarged fonts. Additionally, the study revealed that extant digital assessments seem to under-employ assistive technologies. This implies that most current digital assessment strategies still hinder access for individuals with various disabilities and needs.

Lang and Etzkowitz (2018) analyzed the impact of online testing on higher education. The study revealed that design limitations such as complex graphics, animations or time-sensitive elements had the potential to disadvantage students with cognitive differences. Moreover, Lang and Etzkowitz (2018) assert that in terms of language dependence, traditional digital assessments can be challenging for non-native English speakers because they are frequently provided in a single language. Idiom, vocabulary, and sentence structure nuances can cause misunderstandings and make it more difficult to evaluate students' knowledge accurately. Non-native speakers may struggle with the reading comprehension demands of digital assessments, leading to misinterpretations and inaccurate responses (Bachman & Farrand, 2010; Smith *et al.*, 2020). Thompson (2020) contended that the lack of awareness of accessibility guidelines and best practices by educators and assessment developers was one of the significant barriers to accessing digital assessments.

The literature cites limited testing accommodations as another significant barrier to accessibility. For instance, Smith *et al.* (2020) argued that current digital assessment practices may not offer adequate accommodations for non-native speakers. This includes absence or limited extended testing time or translated versions of the assessments. These barriers may have detrimental effects on diverse populations. These include increasing inequalities, among others. Smith *et al.* (2020) explain further that lack of accessibility may cause students' actual abilities to be understated, which would impede the effectiveness of learning outcomes. Lang and Etzkowitz (2018) assert that unfairly disadvantaged access can exacerbate the achievement gap between students who have disabilities and those who do not. Accessibility impediments may cause decreased motivation and engagement among students and professionals because students and professionals who face obstacles related to accessibility may feel frustrated, anxious, and therefore, may be less inclined to participate in digital assessments (Lang & Etzkowitz, 2018; Moss *et al.*, 2019; Thompson, 2020).

Digital assessments offer valuable benefits, but

accessibility concerns remain to be resolved. To mitigate the barriers to, and improve access to digital assessments, we recommend the following strategies gleaned from the literature:

Universal Design for Learning (UDL) Principles

Educational and professional digital assessors should effectively integrate UDL principles, provide a variety of channels for participation and representation (CAST, 2018; Ercikan & Jin, 2020). This could entail providing alternate response formats, such as voice recordings or interactive drawing tools on the screen.

Multiple Format Options

Offer multiple formats for the test or assessment material, such as text-to-speech, larger fonts, and closed captions for multimedia components (Kirkup, 2019; Smith *et al.*, 2020).

Augment Accessibility Features

Ensure compatibility with modern assistive technologies and incorporate features like screen reader compatibility and keyboard navigation options (CAST, 2018; Moss *et al.*, 2019).

Language Accommodations

Offer translated assessments or extend the assessment time for non-native speakers (Lang & Etzkowitz, 2018; Burgstahler, 2020; Ercikan & Jin, 2020).

Enhanced Training

Provide digital assessors with training on accessibility features and best practices for administering digital assessments to diverse populations (Lang & Etzkowitz, 2018; Smith *et al.*, 2020).

By implementing these recommendations by digital assessment authorities, we envision digital assessment will be more inclusive and accessible.

Review of the Factors Contributing to the Credibility of Digital Assessment

The literature review brought to the fore several factors that may contribute to the credibility of digital assessments. These include, among others, assessment design, test security, potential bias, student motivation and test anxiety (Wilson, 2009; Lang & Etzkowitz, 2018; Moss *et al.*, 2019; Burgstahler, 2020; Ercikan & Jin, 2020). However, this study focuses on the following factors:

Assessment Design

A prerequisite for credibility digital assessment is that assessment tasks should be directly related to the learning objectives they are meant to measure (Wilson, 2009; Moss *et al.*, 2019; Burgstahler (2020). According to Moss *et al.* (2019) and Burgstahler (2020), digital assessments ought to foster higher-order thinking and critical thinking and should minimize rote memorization.

Item Quality

Ambiguous or poorly designed items can lead to misinterpretations and unreliable scores. Therefore, well-constructed items with clear instructions, unambiguous language, and appropriate difficulty levels are crucial in maintaining the quality and credibility of digital assessments (Wilson, 2009; Lang & Etzkowitz, 2018; Moss *et al.*, 2019). Digital assessments can benefit from incorporating various item formats like multiple-choice, open-ended questions, and simulations. This allows for a more comprehensive evaluation of student knowledge and skills (Lang & Etzkowitz, 2018; Ercikan & Jin, 2020).

Test Security

Test security encompasses authentication, item banking and test administration.

Authentication

Implementing secure login protocols with multi-factor authentication helps prevent unauthorized access and protects the integrity of the assessment (Lang & Etzkowitz, 2018; Burgstahler, 2020).

Item Banking

Storing assessment items in secure databases with controlled access minimizes the risk of question exposure beforehand, ensuring the assessment measures current knowledge.

Test Administration

Clear instructions, a well-monitored testing environment, and protocols to address potential cheating attempts are essential for maintaining test security and fair assessment (Moss *et al.*, 2019; Burgstahler, 2020; Ercikan & Jin, 2020)

Potential for Bias

Potential bias is normally discussed on three fronts namely algorithmic bias, language bias and accessibility bias.

Algorithmic Bias

Digital assessments may inadvertently embed bias in scoring algorithms or item selection. Algorithms trained on biased data can perpetuate those biases, leading to unfair disadvantages for certain populations (Lang & Etzkowitz, 2018; Burgstahler, 2020).

Language Bias

Assessments delivered solely in one language can disadvantage non-native speakers. Nuances in vocabulary and sentence structure can lead to misinterpretations (Smith *et al.*, 2020).

Accessibility Bias

Digital assessments that lack accessibility features, such as text-to-speech or screen reader compatibility, can disadvantage individuals with disabilities (Norris & Lopoez, 2016; Smith *et al.*, 2020).

To be considered credible, digital assessments must have a strong base of thoughtfully created tasks, safe administration procedures, and a critical eye for potential biases. By factoring in these considerations, assessors and digital assessment developers can guarantee that digital assessments function as dependable and trustworthy indicators of digital assessments.

Review on Accountability: Fairness & Ethical Use of Digital Assessments

While it is irrefutable that digital assessments are becoming more popular and pronounced in various industries including but not limited to healthcare, education, and industry, questions about equity and moral application remain unanswered. Extant literature asserts that fostering an environment of trust, fairness, and ethical data practices in digital assessments promotes accountability of digital assessments (Elish & Elish, 2018; Pascu *et al.*, 2019; Epstein *et al.*, 2019; Gebru *et al.*, 2020; Smith *et al.*, 2020). The following are some proposed strategies that can ensure accountability, fairness and ethical use of digital assessments gleaned from the literature:

Clear Guidelines and Policies

Development Standards

Potential bias in the algorithms that drive digital assessments presents a serious challenge. Biased training data can result in discriminatory outcomes that disproportionately affect specific demographic groups. Establishing clear guidelines for the design, development, and implementation of digital assessments promotes consistency and reduces the risk of bias or unfair practices. Examples include guidelines on item quality, accessibility features, and test security protocols (Lang & Etzkowitz, 2018; Epstein *et al.*, 2019; Pascu *et al.*, 2019; Gebru *et al.*, 2020).

Data Privacy Policies

Transparent data privacy policies outlining how, for instance, student data is collected, stored, and used during digital assessments are essential for building trust and ensuring ethical data practices (Lang & Etzkowitz, 2018; Pascu *et al.*, 2019; Anselimus, 2023). Epstein *et al.* (2019) contend that digital developers and organisations using these assessments should be held accountable for any biases and unforeseen outcomes. Because of this, the development process must therefore be transparent, allowing independent audits and revealing the data used to train algorithms.

Guidelines for Accessibility

Putting accessibility standards into practice guarantees that every digital assessment taker including students, regardless of ability, has an equal chance to demonstrate their knowledge and skills (Elish & Elish; Smith *et al.*, 2020; Anselimus, 2023).

Promoting Transparency in the Assessment Process

Human Oversight

Human oversight remains crucial alongside algorithmic assessments. Elish and Elish (2018) underscore the importance of human review processes to identify and rectify potential biases in algorithmic decision-making. Additionally, Gebru *et al.* (2020) and Smith *et al.* (2020) advise that appeal mechanisms should be established to allow individuals to contest assessment results perceived as unfair or inaccurate.

Score Reporting

Providing clear and detailed score reports with explanations for performance and areas for improvement allows digital assessment takers to understand their strengths and weaknesses (Pascu *et al.*, 2019; Gebru *et al.*, 2020). This fosters a growth mindset and promotes ownership of the learning curves.

Appeal processes

Establishing clear facilitate for students to appeal scores or contest potential assessment irregularities facilitates fairness and upholds student or professional rights. (Norris & Lopoez, 2016; Epstein *et al.*, 2019).

Transparency in item Selection and Scoring

Educators and assessment developers should strive for transparency in how assessment items are selected and scored, particularly with algorithmic assessments, to mitigate concerns about bias (Lang & Etzkowitz, 2018; Burgstahler, 2020).

Oversight Structures for Continuous Improvement

Independent Audits

Regular independent audits of digital assessments by external experts or 'digital assessment auditors can help identify potential biases, security vulnerabilities, and areas for improvement in the overall design and implementation of digital assessment processes (Pascu *et al.*, 2019; Ercikan & Jin, 2020).

Stakeholder Involvement

Including educators, students, subject matter experts, and professional digital accessibility specialists in the development and review of digital assessments fosters diverse perspectives and helps identify potential issues early on (Thompson, 2020; Smith *et al.*, 2020).

Data-Driven Decision-Making

Regularly analyzing assessment data to identify potential biases, unintended consequences, and areas for improvement allows for continuous enhancement of digital assessment practices and ensures they remain fair and ethical (Pascu *et al.*, 2019; Ercikan & Jin, 2020).

By implementing a well-defined oversight structure, transparent practices, and clear guidelines, digital

assessment stakeholders can guarantee the responsible utilization of digital assessments. In the end, this benefits every learner by creating an environment of trust, justice, and moral data practices (Ercikan & Jin,2020; Society

for Information Technology & Teacher Education,2023; Haleem & Ditsa, 2024).

Table 1 shows a summary of knowledge gaps identified from the literature

Table 1: Summary of knowledge gaps identified from the Literature Review

Area	Knowledge Gap
Accessibility	There is a dearth of data regarding the long-term effects of inaccessible digital assessments including the lack of effective use of assistive technologies post COVID-19 era (Moss <i>et al.</i> , 2019; Ercikan & Jin,2020; Smith <i>et al.</i> , 2020)).
Credibility	Need for further investigation into the effectiveness of different accessibility features in mitigating bias for diverse populations (Liu & Zhao, 2018; CAST, 2018)
Accountability	-Feasibility and effectiveness of student involvement in identifying and mitigating potential biases in digital assessments (Bennett, 2011; Lang & Etkowitz, 2018). -Development of clear and consistent ethical guidelines for the use of algorithmic scoring and item selection in digital assessments (Popham, 2009)
Long-term impact of digital assessments on accessibility practices	Limited research explores how digital assessments evolve to cater to diverse needs over time (Stufflebeam,2000; Lang & Etkowitz, 2018)
Effective strategies for mitigating algorithmic bias	Current research identifies potential bias in algorithms, but further exploration is needed on effective mitigation strategies (Kirkup, 2019; Burgstahler,2020)
The impact of digital assessments on non-cognitive skills	Most research focuses on content knowledge; more studies are needed to understand how digital assessments measure non-cognitive skills like critical thinking and collaboration (Brown, 2019; Lane <i>et al.</i> , 2021)
The role of artificial intelligence (AI) in promoting accessibility and reducing bias	Emerging AI technologies in the post-COVID 19 era might offer solutions, but research is needed to explore their ethical implications and effectiveness (CAST, 2018; Burgstahler,2020; Lane <i>et al.</i> , 2021; Martínez & Pérez, 2022)
Standardized best practices for data security in digital assessments	While the importance of data security is established, clear and comprehensive best practices for different digital assessment contexts are needed (Koehler & Mishra,2008; Kirkup, 2019; UNESCO, 2021).

Source: Authors' elaboration on various Literature

METHODOLOGY

Search Strategy

A comprehensive search was conducted using academic databases like Educational Resources Information Center (ERIC), Google Scholar, Wiley, Web of Science, Research Gate and PsychINFO , among others. Search terms included “digital assessment,” “accessibility,” “credibility,” “assessment design”, “non-native speakers of the English language, “assistive technology,” “algorithmic bias,” “individuals with disabilities,” “data privacy,” “item quality”, “ethical use”, “accessibility features,” and “content validity.” Journal articles published within the last ten years (2000-2023) and focusing on the impact of digital assessments on accessibility were given priority.

Inclusion and Exclusion Criteria

Studies included in the systematic review focused on the accessibility of digital assessments for diverse populations employing empirical research methods such as surveys, experiments, and case studies published in peer-reviewed journals were included in the review. The studies that focused solely on the development or technical aspects of digital assessments, commentaries and opinion papers and were not published in peer-reviewed journals were

excluded. With regards to data analysis and synthesis, thematic analysis related to the integration of digital technologies in assessment practices in education paradigms, specifically related to the challenges of accessibility, credibility, and accountability of digital assessments, was done. The specific challenges faced by individuals with disabilities and non-native speakers in digital assessments were also factored in (Benard & Lan, 2015; Chapman, 2018; CAST, 2023)

Theoretical Underpinnings

Three (3) theoretical frameworks underpin the examination of accessibility, credibility, and accountability in digital assessment. The three theoretical frameworks are Universal Design for Learning (UDL), Technological Pedagogical Content Knowledge (TPACK), and the Socio-Technical Systems Theory.

Universal Design for Learning (UDL)

In the realm of digital evaluation, Universal Design for Learning (UDL) principles underscore the significance of offering various channels for representation, expression, and interaction to guarantee inclusivity for all students, including those with disabilities (CAST, 2018).

The UDL theory provides a framework for designing educational environments and assessment practices that accommodate the diverse needs of learners (Rose & Meyer, 2002). UDL encourages the design of instructional materials and activities that allow educational goals to be attainable by individuals with diverse differences in their abilities to see, hear, speak, move, read, write, understand English, organize and engage without having to adapt the curriculum repeatedly to meet special needs Lee and Kim (2021). The main thrust of UDL is to build flexibility that can be adjusted to fit all students' strengths and needs. UDL theory is anchored on four (4) core principles: multiple means of representation (the knowledge network), multiple means of expression (enlisting the brain's effective network), multiple means of engagement (enlisting the brain's strategic network) and multiple means of assessment (enlisting the brain's knowledge, effective, and strategic networks (Sherlock Center, 2009).

Technological Pedagogical Content Knowledge (TPACK)

In order to guide effective teaching and learning practices in digital environments, TPACK integrates technology, pedagogy, and content knowledge (Mishra & Koehler, 2006). To improve the validity and credibility of assessment practices in the context of digital assessment, TPACK emphasizes the intersection of pedagogical approaches, subject matter expertise, and technology tools (Koehler & Mishra, 2008). The TPACK framework describes how educators' understanding of educational technologies and Pedagogical Content (PCK) interact with one another to produce effective teaching with technology. It builds on Shulman's (1987, 1986) PCK descriptions. The cornerstone of effective technology-based education is known as TPACK, which necessitates an understanding of how concepts are represented using technology, pedagogical strategies that use technology

to teach content constructively, information about what makes concepts difficult or easy to learn and how technology can help address some of the issues that students face, information about the prior knowledge and epistemological theories of students, and information about how technologies can be used to build on existing knowledge to develop new epistemologies or strengthen old ones.

The Socio-Technical Systems (STS) Theory

This theory highlights how technological infrastructures, organisational structures, and human factors interact to shape mechanisms for accountability, accessibility, and credibility (Avgerou, 2003). The Socio-Technical Systems (STS) theory examines the interaction between social and technical elements within complex systems (Trist & Bamforth, 1951). STS theory is based on the premise that an organization combines social and technical parts and is open to its environment. Given that both the social and technical elements must work together to accomplish tasks, work systems produce both physical products and social and/or psychological outcomes. The key issue of this theory is that work must be planned in such a manner that the social and technical components interact to yield favourable or positive outcomes (Thach & Woodman, 1994). The three theories underpin this study. The UDL and TPACK theories help explain digital assessments in the academic or education sectors, while the STS theory helps explain digital assessments in the professional arena.

Conceptual Framework

Figure 1 illustrates how we envision the interaction of the variables in this study. The independent variables are accessibility, credibility, and Accountability. The moderating variables are technology type and assessment purpose. The learning or professional outcome(s) constitute the dependent variable.

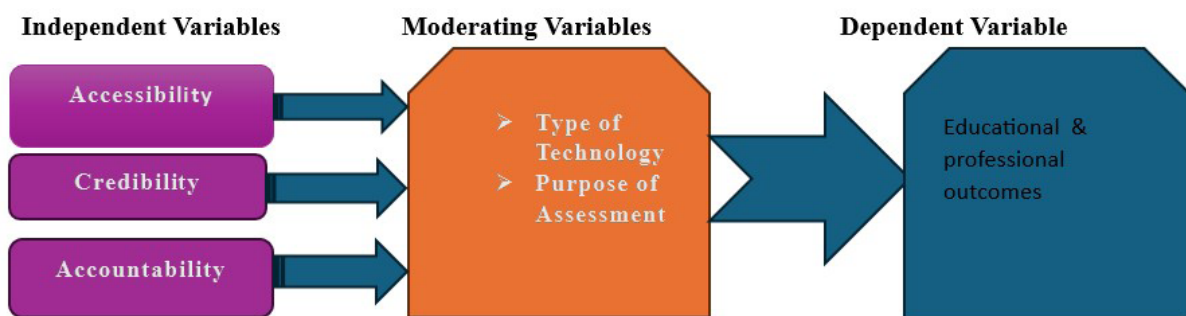


Figure 1: Conceptual Framework

Source: Authors' elaboration

Operational Definitions of Key Concepts

Accessibility refers to features for diverse learners, such as text-to-speech or closed captions, user-friendly interface designs, and culturally appropriate assessments (Edyburn *et al.*, 2010).

Popham (2009) explains that credibility in the context of digital assessments connotes algorithmic fairness in scoring, data security and privacy and assessment. It also means assessment design integrity in terms of question quality, validity and alignment with academic and

professional objectives, among others.

Regarding accountability, Stufflebeam (2000) asserts that it implies the roles and responsibilities of all stakeholders engaged in assessment procedures. It also refers to the transparency exhibited by educators, policymakers, and the roles of technology developers. Accountability also extends to ethical considerations such as data privacy and student well-being.

Type of technology as a moderating variable refers to the different digital platforms offering varying accessibility features and security protocols (Gregson, 2017). Depending on the assessment type, the level of accessibility, credibility, and accountability for assessment purposes varies. For instance, the objective/goal of formative assessment is different from that of summative assessment.

Learning outcomes imply that effective assessments should ultimately promote fair and accurate evaluation of student learning (Moss *et al.*, 2019).

DISCUSSIONS

Challenges Undermining the Integration of Digital Technologies into Digital Assessments

The integration of digital technologies into assessment procedures holds immense potential for enhancing efficiency and effectiveness. However, there are challenges that remain to be resolved. Three (3) major challenges that this systematic review unravels are:

Accessibility Issues

Ensuring accessibility for people with disabilities is one of the biggest obstacles to incorporating digital technologies into assessment practices. Digital platforms frequently do not have enough features to meet a range of needs, like screen readers for people with visual impairments (Yadav & Shavelson, 2017; Brown, 2019; Smith *et al.*, 2020). The increased reliance on online assessments has highlighted issues of digital divide and inclusivity. Accessibility in this context refers to the ease with which individuals can participate in assessments, considering factors such as internet access, digital literacy, and availability of appropriate devices. According to Li and Lalani (2020), the rapid transition to online learning and assessments has exposed significant disparities in access to technology and the Internet, which can hinder the participation of underprivileged groups. This digital divide can lead to unequal opportunities in professional development and certification processes, potentially marginalizing those without adequate resources.

Language Barriers

When instructions and content are provided exclusively in English, non-native English speakers encounter considerable difficulties utilizing digital assessment tools. This may result in misconceptions and imprecise evaluations of their actual skills (Liu & Zhao, 2018; Eaton, 2019).

Technological Disparities

Since not everyone has equal access to technology and high-speed internet, socioeconomic disparities can make it more difficult to obtain digital assessment tools. Marginalized groups, such as people with disabilities and non-native speakers, may be disproportionately impacted by this (Lang & Etzkowitz, 2018; Kirkup, 2019; Gebru, *et al.*, 2020; Anselimus, 2023).

Credibility Issues

The credibility of online assessments raises deep concerns in the post-COVID-19, as it impacts the perceived validity and reliability of the results. This includes ensuring that the assessments accurately measure what they intend to, are free from biases, and are administered under conditions that uphold academic integrity. Research by Hodges *et al.* (2020) emphasizes that the sudden shift to online assessments poses challenges to maintaining the same standards of validity and reliability as traditional in-person professional and education assessments. Ensuring the credibility of these assessments requires rigorous design, secure administration, and robust methods to verify the identity of participants (Hodges *et al.*, 2020).

Psychometric Rigor

The validity and reliability of online assessments are crucial. Do they accurately measure the skills and knowledge they purport to assess, and can we trust their consistency across different administrations (Lang *et al.*, 2021)?

Content Bias

The content of online assessments should be free from bias that could unfairly disadvantage certain groups. Are the digital assessments culturally sensitive and inclusive (Cascio & Aguinis, 2018)?

Accountability Issues

Accountability in online assessments involves the mechanisms put in place to ensure that the organizations and individuals involved are responsible for maintaining standards, addressing misconduct, and continuously improving assessment practices. It also encompasses the transparency of the assessment processes and the ability of stakeholders to hold providers accountable for the quality and fairness of their assessments. The proliferation of online assessments in the post-COVID-19 era has led to increased scrutiny regarding how these assessments are monitored and regulated. As noted by Barbour *et al.* (2020), establishing clear guidelines and accountability measures is essential to protect the integrity of professional and educational assessments and to build trust among stakeholders.

Test Security

Online assessments are vulnerable to cheating and other forms of misconduct. Do the platforms have adequate safeguards to ensure the integrity of the testing process (Educational Testing Service, 2023)?

Data Privacy

The data collected through online assessments raises privacy concerns. How is the particular data set secured and used? Are there clear policies in place to protect the assessment taker information (International Association of Privacy Professionals, 2023)?

Recommendations or Strategies to Mitigate the Challenges

The corresponding strategies to mitigate the identified challenges include:

Universal Design for Learning (UDL)

UDL principles can improve accessibility by making digital assessment tools usable for the greatest number of people, including those with disabilities (Gregson, 2017; Burgstahler, 2020; UNESCO, 2021).

Multilingual Support

By adding multilingual support functions to digital assessment tools, non-native speakers' language barriers can be lessened. This entails offering translation tools in addition to providing instructions and content in several languages (Ercikan & Jin, 2020).

Equitable Access Initiatives

Organizations can carry out programs to help close the digital divide by giving people from underprivileged backgrounds access to gadgets, the internet, and training courses (Warschauer, 2019; Brown, 2019; National Center for Fair & Open Testing, 2023)

CONCLUSION

This paper examines the accessibility, credibility, and accountability of digital assessments through systematic literature reviews for the period 2000-2023. The growing ubiquity of digital assessments in academic and professional contexts, especially in the post-COVID-19 era makes it necessary that their accessibility, credibility, and accountability be re-examined. In response to the research question on how the current digital assessment practices impact the accessibility of assessments for diverse populations especially in the post-COVID-19 era, the study finds that the current digital assessment practices including the post-COVID 19 assessments, seem to be deficient in employing assistive technologies. With regards to the second research question on the factors that contribute to or undermine the credibility of digital assessments, the study finds that, overall, the sudden shift to online assessments posed challenges to maintaining the same standards of validity and reliability as traditional in-person exams or interviews. This implies that digital assessments still grapple with psychometric rigor and content bias issues. Finally, with respect to the third question on how accountability mechanisms can be implemented to ensure fairness and ethical use of digital assessments in the post-COVID-19 world, the integrity of professional and educational digital assessments may

be undermined because online assessments are vulnerable to cheating and other forms of misconduct. In addition, the data collected through online assessments raises privacy concerns. In view of the foregoing findings, we recommend intensive integration of assistive technologies in digital assessments to enhance accessibility. Digital assessing authorities should also establish clear guidelines and other mechanisms for detecting cheating and plagiarism in digital assessments. The credibility of online assessments in the post-COVID-19 era can be augmented by ensuring that the assessments accurately measure what they intend to, are free from biases, and are administered under conditions that uphold academic and professional integrity. This review contributes to a comprehensive understanding of the complex digital assessment landscapes' dynamics especially in the post COVID-19 era. By and large, it is imperative to tackle the identified challenges from a multi-sectoral perspective to ensure that future digital assessments remain fair, reliable, and trustworthy tools for evaluating skills and qualifications.

Acknowledgments

The authors thank the anonymous reviewers for their valuable, constructive criticisms of the earlier version of this manuscript. Their input significantly improved the quality of this study. All errors are ours.

Declaration of Competing Interests

The authors declare that they have no financial or non-financial interests to declare.

REFERENCES

- Anselmus, S. M. (2023). Assistive technologies and participation of students with visual impairments in extra-curricular activities - What does the literature say? *American Journal of Interdisciplinary Research and Innovation*, 2(4), 67–73. <https://doi.org/10.54536/ajiri.v2i4.2155>
- Avgerou, C. (2003). The link between ICT and economic growth in the discourse of development. In S. D. D. Ciborra (Ed.), *From control to drift: The dynamics of corporate information infrastructures* (pp. 90–107). Oxford University Press.
- Bachman, B., & Farrand, L. (2010). Against algorithmic discrimination. *Annals of the New York Academy of Sciences*, 1366(1), 165-180.
- Barbour, M. K., LaBonte, R., Hodges, C., Moore, S., Lockee, B. B., Trust, T., ... & Kelly, K. (2020). Understanding Pandemic Pedagogy: Differences between Emergency Remote, Remote, and Online Teaching. *Journal of Educational Technology Systems*, 49(1), 9-36.
- Barnard, L., & Lan, W. Y. (2015). Digital assessment literacy: Supporting students' ability to use digital tools effectively. *The Journal of General Education*, 64(4), 218–234.
- Bennett, R. E. (2011). *Accountability in education: Meeting the challenge of measuring complex learning*. Information Age

- Publishing.
- Brown, A., Smith, B., & Johnson, C. (2018). Enhancing accessibility in digital assessments: Best practices for inclusive design. *Journal of Educational Technology*, 25(3), 321–335.
- Brown, S. (2019). Enhancing accessibility in digital assessments. *Journal of Educational Technology Systems*, 48(2), 232–245.
- Burgstahler, S. (2020). Universal design in assessment. In L. M. West & L. A. Lewis (Eds.), *Handbook of research on multidisciplinary approaches to disability and equity in digital environments* (pp. 198-213). IGI Global.
- CAST. (2018). Universal design for learning guidelines version 2.2. Retrieved from <http://udlguidelines.cast.org>
- Cascio, W. F., & Aguinis, H. (2018). Content validity in employment selection: Review, recommendations, and future research. *Research in Organizational Behavior*, 38(1), 209-241.
- Cascio, W. F., & Aguinis, H. (2018). Content validity in employment selection: Review, recommendations, and future research. *Research in Organizational Behavior*, 38(1), 209-241.
- Chapman, A. (2018). *Accessibility in online assessments: A guide for educators*. University of Southampton.
- Cook, A. M., & Polgar, J. M. (2015). *Assistive Technologies: Principles and Practice* (4th ed.). Mosby.
- Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education*, 8, Article 1206936. <https://doi.org/10.3389/feduc.2023.1206936>
- Eaton, J. S. (2019). Rethinking assessment in a digital world: The ethics of validity, bias, and fairness. *Educational Researcher*, 48(8), 566–577.
- Edyburn, D. L., Higgins, S., & Judson, E. (2010). *Universal design for learning: Theory and practice*. Association for Supervision and Curriculum Development.
- Edyburn, D. L. (2010). Assistive Technology and Mild Disabilities. In J. Sanchez (Ed.), *Technology-Supported Environments for Personalized Learning: Methods and Case Studies* (pp. 21–45). IGI Global.
- Educational Testing Service. (2023). Protecting the integrity of our tests. <https://www.ets.org/>
- Elish, M. C., & Elish, K. C. (2018). Rethinking algorithmic fairness for criminal justice. *Law and Contemporary Problems*, 81(2), 309–342.
- Epstein, D., Caliskan, A., & Bryson, J. (2019). Fairness in machine learning. *arXiv preprint arXiv:1908.09532*.
- Ercikan, K., & Jin, Y. (2020). Challenges and opportunities of multilingual assessment. *Educational Measurement: Issues and Practice*, 39(1), 53-61.
- Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: A randomized controlled trial. *JMIR Mental Health*, 5(e64). <https://doi.org/10.2196/mental.9782>
- García, M., & López, R. (2020). Ensuring accountability in digital assessments: Strategies for educational institutions. *International Journal of Educational Administration*, 12(2), 145-158.
- Gebru, T., Morgenstern, J., Vecchiola, C., Vaughan, J., Stapleton, M., Mitchell, M., & Borenstein, T. (2020). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 80-89.
- Gregson, J. (2017). Blockchain technology for education: Unpacking the hype. *EDUCAUSE Review*, 52(2), 14–20.
- Haleem, Y., & Ditsa, E. M. G. (2024). Assessing the relationship between technological factors and the implementation of human resource information system: A survey in the municipal, metropolitan, and district assemblies in the Upper West Region of Ghana. *American Journal of Interdisciplinary Research and Innovation*, 3(2), 7–29. <https://doi.org/10.54536/ajiri.v3i2.2594>
- Hodges, C., Moore, S., Locke, B., Trust, T., & Bond, A. (2020). The Difference Between Emergency Remote Teaching and Online Learning. *Educause Review*. Retrieved from <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>.
- International Association of Privacy Professionals. (2023). *Data privacy essentials*. <https://iapp.org/>
- Johnson, D., & Wang, L. (2019). Evaluating the credibility of digital assessments: A review of current practices. *Assessment and Evaluation in Higher Education*, 44(4), 567-582.
- Kirkup, G. (2019). Technology and inequality: Concentrating on digital divisions. *Information, Communication & Society*, 22(2), 161–175.
- Koehler, M. J., & Mishra, P. (2008). Introducing TPACK. In AACTE Committee on Innovation and Technology (Ed.), *Handbook of technological pedagogical content knowledge (TPCK) for educators* (pp. 3–29). Routledge.
- Lang, J. M., & Etzkowitz, H. (2018). The impact of online testing on higher education: A review of the literature. *Educational Assessment*, 23(3), 182-210.
- Lang, J. M., Wheatcroft, J., & Wheeler, A. R. (2021). Examining the psychometric properties of online professional assessments during a pandemic. *International Journal of Testing*, 11(2), 1-16. <https://doi.org/10.1080/1530508X.2021.1879223>
- Lane, S., & Buckingham, J. (2021). Algorithmic fairness in educational assessment: Challenges and recommendations. *Educational Researcher*, 50(2), 138–146. <https://doi.org/10.3102/0013189X21989342>
- Lee, S., & Kim, J. (2021). Validity and reliability of digital assessments: A meta-analysis. *Educational Research Review*, 36, 1-15.
- Li, C., & Lalani, F. (2020). The COVID-19 pandemic has changed education forever. This is how. *World Economic Forum*. Retrieved from <https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/>
- Liu, M., & Zhao, Y. (2018). Language-related differential item functioning on a computerized adaptive test

- for non-native English speakers. *Educational and Psychological Measurement*, 78(3), 418–436.
- Martínez, P., & Pérez, J. (2022). Addressing cheating and plagiarism in digital assessments: Strategies and considerations. *Journal of Academic Integrity*, 10(1), 78–92.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Moss, C., Shah, A., & Brookhart, S. M. (2019). Advancing formative assessment practices in the digital age. *Educational Assessment*, 24(2), 132–152.
- National Center for Fair & Open Testing (Fair Test). (2023). *Bias in testing*. <https://fairtest.org/racebias/>
- Norris, D. F., & Lopez, K. A. (2016). The digital divide: Fact and fiction. *Annual Review of Sociology*, 42(1), 743–767.
- Pascu, M., Mavroceidi, E., & Oliveira (2019). *Assessment essentials: Understanding student learning*. Pearson Education Limited.
- Pickard, M., Schuetzler, R., Valacich, J., & Wood, D. (2017). Next-generation accounting interviewing: A comparison of human and embodied conversational agents (ECAs) as interviewers. *SSRN Electronic Journal*, April, 1–21. <https://doi.org/10.2139/ssrn.2959693>
- Popham, W. J. (2009). *Modern educational assessment: Principles, practices, and applications*. Pearson Education.
- Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning*. Association for Supervision and Curriculum Development.
- Sherlock Center. (2009). *Rhode Island modified UDL educator checklist—Version 1.2*. Providence: Rhode Island College. <http://www.ric.edu/sherlockcenter/udl/udleducatorchecklist.pdf>
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Slepankova, M. (2021). *Possibilities of artificial intelligence in education: An assessment of the role of AI Chatbots as a communication medium in higher education* (dissertation). Available at: <http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-108427>
- Smith, A., Lang, C., & Balsler, M. (2020). A review of accessibility considerations in digital assessments.
- Smith, C. (2023). GPT-5 might make ChatGPT indistinguishable from a human. *BGR*. Available at: <https://bgr.com/tech/gpt-5-might-make-chatgpt-indistinguishable-from-a-human/>
- Society for Information Technology & Teacher Education. (2023). *ISTE standards for students, educators, and administrators*. <https://iste.org/standards/students>
- Springer Web Accessibility Initiative, WAI. (2018). *Web content accessibility guidelines (WCAG) overview*. World Wide Web Consortium.
- Stufflebeam, D. L. (2000). *The seven evaluation continents: A comprehensive guide to educational and program evaluation*. Kluwer Academic Publishers.
- Thach, L., & Woodman, R. (1994). Organizational change and information technology: Managing on the edge of cyberspace. *Organizational Dynamics*, 23, 30–46.
- Thompson, A., & Mishra, P. (2007–2008). Breaking news: TPACK becomes TPACK! *Journal of Computing in Teacher Education*, 24(2), 38–64.
- Thompson, C. (2020). Algorithmic bias and the digital divide in education: A call for critical research and design. *Learning, Media and Technology*, 45(3), 425–440. <https://doi.org/10.1007/s40593-021-00285-9>
- Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations*, 4(1), 3–38.
- Universal Design for Learning Center. (2023). *CAST*. <https://udlguidelines.cast.org/>
- UNESCO. (2021). *Global partnership strategy for early childhood, 2021-2030* [Report].
- Usher, F. (2018). Insights into students' experiences with e-assessment. *Computers & Education*, 116, 190–202
- Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, 20(4-5), 520–532.
- Warschauer, M. (2019). *Technology and social inclusion: Rethinking the digital divide*. MIT Press.
- Web Accessibility Initiative, WAI. (2018). *Web content accessibility guidelines (WCAG) overview*. World Wide Web Consortium. <https://www.w3.org/WAI/>
- Wilson, M. (2009). *The design of assessment instruments in education and training*. SAGE Publications Ltd.
- Xiao, Z., Zhou, M. X., Chen, W., Yang, H., & Chi, C. (2020). If I hear you correctly: Building and evaluating interview Chatbots with active listening skills. In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14).
- Yadav, A., & Shavelson, R. J. (2017). Digital technology and student cognitive development: The neuroscience of the university classroom. In R. J. Shavelson (Ed.), *Measuring cognitive development in the digital age* (pp. 43–63).