unesco
Global Education Monitoring Report

Background paper prepared for
the Global Education Monitoring Report

Technology in education

USE OF TECHNOLOGY-BASED ASSESSMENTS: A SYSTEMATIC REVIEW

DANDAN CHEN

2023

ABSTRACT

Technology-driven shifts have created opportunities to improve efficiency and quality of assessments. Meanwhile, they may have exacerbated underlying socioeconomic issues in relation to educational equity. The increased implementation of technology-based assessments during the COVID-19 pandemic compounds the concern about the digital divide, as digital access, connectivity, and coping strategies vary across the globe. This systematic review was intended to answer how the use of technology-based assessments has affected the education system's functioning, compared to traditional assessments that do not employ any technology solution. It covered 34 countries from 34 full-text sources in English published in 2018-2022. A total of 12 themes emerged corresponding to six hypotheses about technology-based assessments. In summary, when compared with traditional paper-based exams, mixed evidence was found when testing assumptions about technology-based assessments' roles in cheating reduction, learning boost, monitoring support, instructional improvement, and non-teaching workload reduction. Strong supporting evidence was found when testing assumptions about technology-based assessments' higher measurement precision, easier interpretation, higher learner engagement, and more interaction with others at the learning level, in addition to smoother communication with parents at the educating level. Limited but positive evidence at the management level suggested that technology-based assessments are more cost-effective and time-efficient.

TABLE OF CONTENT

1.Introduction

Assessment, which involves the direct evaluation of students' learning outcomes, is integral to other aspects of education, including pedagogy and curriculum. No discussion of technology as a learning tool is complete without addressing assessment (Burns, 2021). Globally, technology has dramatically transformed educational assessments, automating test administration and improving test precision and fairness. For instance, many countries have replaced the paper-based mode of assessment with the computer-based mode. The Programme for International Student Assessment (PISA), the largest international large-scale assessment in the world, established in 2000, was mostly switched from the paper-based mode to the computer-based mode in 2015. Also, technology-driven adaptive designs have been widely implemented in assessments. The Programme for the International Assessment of Adult Competencies (PIAAC), one of the first international large-scale assessments, implemented a multistage adaptive testing (MST) design in 2012 for around 40 countries (Yamamoto et al., 2018).

Given the fundamental roles assessments can play, there has naturally been discussion over the utility of technology-based assessments. On the one hand, technology-driven shifts have created opportunities to improve efficiency and quality of assessments (Bennett, 2002). On the other, they may have also exacerbated underlying socioeconomic issues related to educational equity. For instance, the digital divide, defined as a social inequity in access to and frequency and ability of using technology across regions and populations, could have widened the achievement gap (Ercikan et al., 2018). Lately, during the pandemic, the international community has ramped up efforts to move education online (OECD, 2021), accelerating the integration of technology in assessments of and for learning. This increased implementation compounds existing concerns about the digital divide across the globe (e.g., Fisher et al., 2022). Today it is particularly exigent to examine the utilization of technology-based assessment, as funders want to know the impact of considerable undertakings coming into national and international technology-based assessment programs.

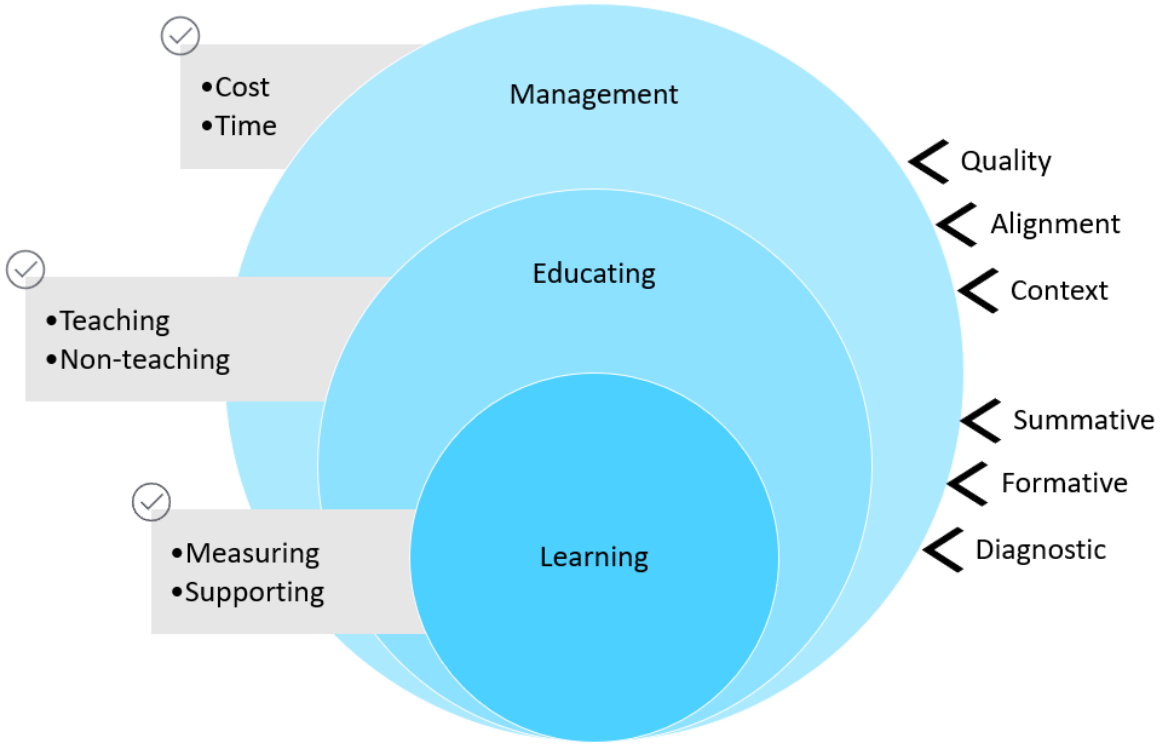
This paper was intended to synthesize evidence from a systematic review of the existing literature on *how the use of technology-based assessments has affected the education system's functioning*, compared to traditional assessments that do not employ any technology solution. It was expected to include a wide array of countries to improve geographical representation of the evidence. The reviewed literature covered peer-reviewed journal articles and grey literature (e.g., institutional reports, conference papers, preprints) from 2018-2022. Current trends, main challenges, and opportunities associated with implementing technology-based assessments were discussed in the end, in addition to gaps in the literature, implications for future research, and limitations of this study.

2.Conceptual Framework

A conceptual framework was constructed to guide the literature selection, as shown in Figure 1. It was adapted from four widely circulated frameworks: (1) Clarke's (2012) SABER framework about evaluating assessment programs

(i.e., three drivers: *enabling context*, *system alignment*, and *assessment quality*), (2) Care and Anderson's (2016) framework on the learning ecosystem (i.e., four interrelated circles: *formal education*, *community education*, *parent and family*, and *society at large*), (3) Lock's (2007) Iron triangle for management (i.e., three constraints: *quality*, *cost*, and *time*), and (4) Black and Wiliam's (1998) typology of assessments (i.e., three types: *summative*, *formative*, and *diagnostic* assessments).

Figure 1. *Guiding framework: Key aspects for consideration of a technology-based assessment.*



Note. Adapted from Black and Wiliam (1998), Clarke (2012), Care and Anderson (2016), and Lock (2007).

The conceptual framework presented in Figure 1 emphasizes the interrelated procedures in a technology-based assessment: learning, educating, and management. Following Care and Anderson (2016), learners, teachers, and system-level decisionmakers are the three major users in any learning ecosystem. At the learning level, the key users are learners. At the educating level, the key users are primarily teachers but can also include learners when highly interactive assessments direct learners to learn from each other via peer feedback. At the management level, the key users are school administrators and governmental agencies responsible for systemic policy design, in addition to teachers whose situation sheds light on the efficacy of the managerial mechanism and who in many occasions directly contribute to system-wide decision-making. Following Black and William (1998), technology-based assessments can take the form of summative, formative, and diagnostic assessments.

Given this framework, a few issues were considered in the learning, educating, and management processes in technology-based assessments. Following Clarke (2012), the quality of assessment matters, as reflected particularly in the learning process, to accurately measure learning outcomes and effectively support learning. Also, the alignment and context of the use of technology-based assessment matters. They could reflect, at the educating

level, in teachers' teaching and non-teaching activities across myriad geographical settings. They could materialize, at the management level, in the administrative cost and time of implementing technology-based assessments, following Lock (2007): While the quality of assessment matters, there is often an emphasis on striking a balance among quality, cost, and time, to avoid overly investing in one or two of the three but sacrificing the rest.

This conceptual framework was employed to answer the central question regarding the use of technology-based assessments. It was not intended to cover every essential element in this regard but to frame the scope of the study. By comparing technology-based assessments with the traditional, paper-based assessments, six hypotheses were scrutinized: At the learning level, (1) learners' learning achievement is better measured by technology-based assessments, and (2) learners are better supported in the learning process by technology-based assessments. At the educating level, it was hypothesized that (3) teachers are better prepared for teaching by technology-based assessments, and (4) they are better supported in non-teaching activities by technology-based assessments. At the management level, it was hypothesized that (5) technology-based assessments are more cost-effective, and (6) technology-based assessments are more time-efficient in educational changes.

3.Methodology

Following the suggestions from Card (2012) and Littell et al. (2008), the following procedure was utilized for this systematic review: Clarify key terms in the literature search, formulate the inclusion criteria and searching strategies, locate studies, screen, and select studies, assess the quality of the selected studies, and synthesize the evidence.

3.1.Key terms

Assessment in this study is broadly conceived as any evaluative activity to measure learning outcomes. Following Black and Wiliam (1998), assessment can serve any of the following purposes: obtain learners' prior knowledge before an educational activity (i.e., diagnostic assessment, such as the pre-course survey), understand learners' progress in the learning process (i.e., formative assessment, like the end-of-class quiz), or measure their ultimate achievement against certain benchmarks (i.e., summative purpose, like the final exam). *Technology-based assessment* here refers to the assessment that employs any digital tool.

3.2.Inclusion criteria

The studies included in this review met the following criteria:

(1) Having been released in English from January 1, 2018 up to May 31, 2022;
(2) Having been released as academic papers or institutional reports indexed with a full text in the selected databases;
(3) Having provided enough evidence (i.e., either first-hand or cited information) about technology-based assessments;
(4) Having focused on the assessment of learning outcomes in kindergarten up to secondary education;
(5) Having specified the geographical area where the evidence was collected;
(6) Having concentrated on the geographical areas outside the United States.

The literature from the last four years was covered. The end date May 31, 2022 was when this systematic review was initiated. The data included in these sources could be observational or experimental and either primary or secondary. The literature search was further narrowed down to focus on compulsory education, a priority area that significantly impacts individuals' lifelong learning and with which many developing countries were grappling. To improve geographical representation, criteria (5) and (6) were used to focus on context-specific evidence and cover as many countries as possible, after filtering out an extensive amount of the literature about the United States (where most internationally known technology-based assessments originated, including SAT, GRE, GMAT, and LSAT).

3.3. Literature selection

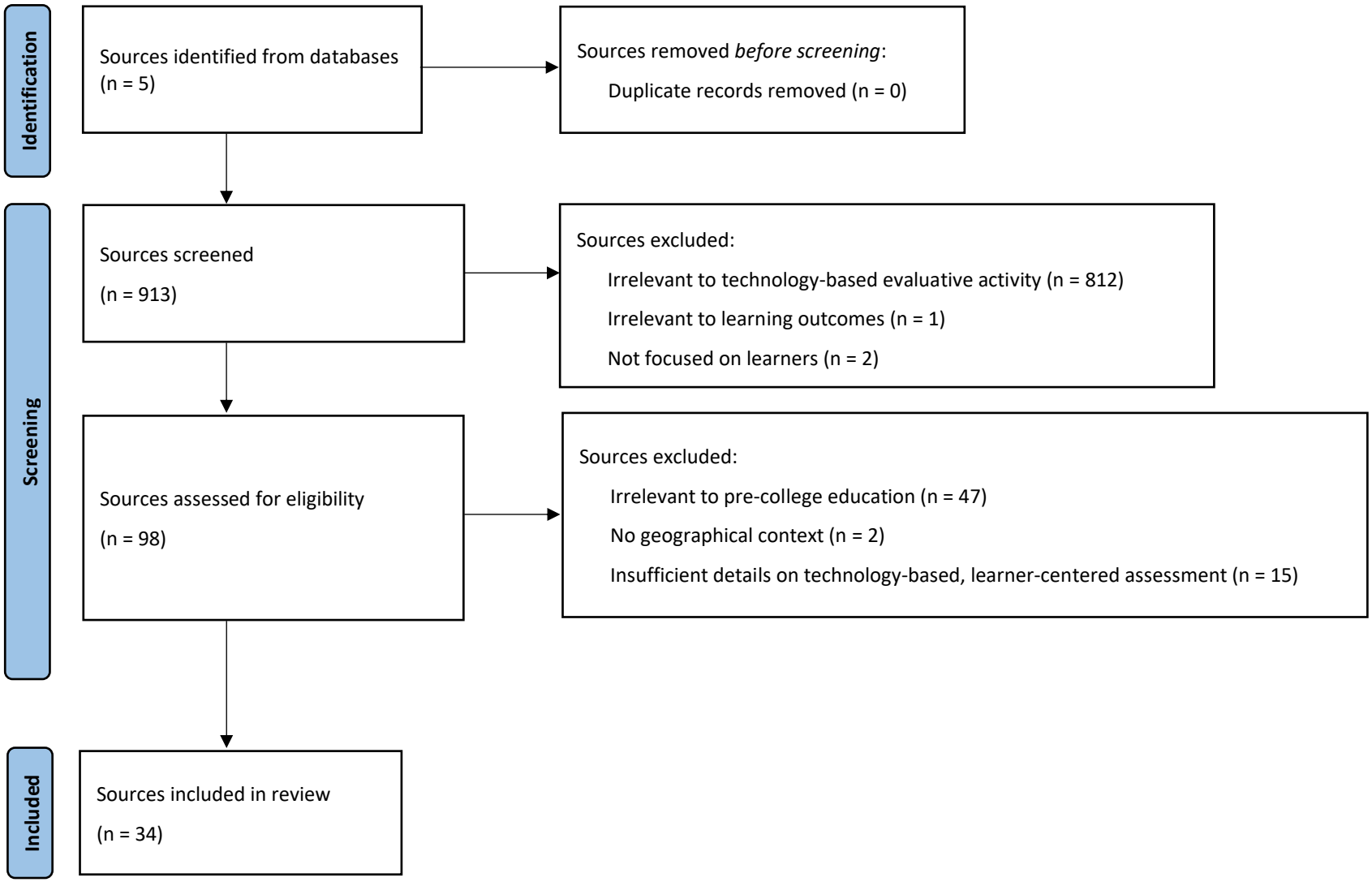
Inspired by Tong et al. (2007), the studies for further analysis were drawn from a selection of peer-reviewed journals and institutions' databases. Journal articles were searched from five research databases known for their comprehensive coverage of educational and psychometric research on assessments, namely ERIC, PsycINFO, PsycARTICLES, PAIS Index, and Sociological Abstracts. ERIC covers journals and magazines in vast areas of educational research. PsycINFO indexes scholarly articles in psychology, education, social science, business, and organizational behavior. PsycARTICLES supports the search for academic and practice literature in psychology and related disciplines. PAIS Index includes sources about political, economic, social, and cultural issues across countries. Sociological Abstracts provides international literature in sociology and related social and behavioral sciences disciplines. It was assumed that the sheer volume of sources from these databases would be large enough to reach saturation, for which more research studies from other databases would not add substantial new information (Morse, 1995).

"Assessment" and "technology" were the two major search terms. Their related terms were included, namely "test," "digital," "computer-based," "adaptive," "game-based," and "online." The term "learner" was added to narrow down the search, with its related terms "examinee," "student," and "pupil." Given that educational assessments are intended to measure learners' achievement, the term "achievement" was added, with its related terms "competency" and "ability." Given our focus on how the technology-based assessment can support learners, teachers, and decisionmakers, the term "quality" was added, in addition to related terms, namely "accuracy," "reliability," and "validity." Additional search queries were employed to exclude studies that contain key words like "higher education," "college students," "undergraduate students," and "college faculty" because the education levels ranged from kindergarten to upper secondary, and to exclude sources with "United States," "US" because the United States was not considered in this study. Appendix 1 specifies the search queries.

3.4. Analysis

Key evidence from the selected sources that either supported or challenged each hypothesis was synthesized and reported. This analysis followed guidelines from the U.K. Department for International Development (DFID, 2014), which have been widely used for assessing evidence across diverse geographical contexts. Key takeaways were then

Figure 2. *Flowchart of the literature selection process.*



summarized in the Results section, based on the synthesized evidence. Lastly, a sensitivity analysis was completed to show how different selected sources contributed to the body of evidence for each hypothesis. There was one procedure in which six coders were involved in the coding of the relevance (1 for "relevant" and 0 for "irrelevant" for each of the six pre-specified hypotheses) and quality (1 for "hit" and 0 for "miss" for each of the 20 criteria in the grading rubric in Appendix 2) of each source. A total of 884 binary codes were created, in which 23% of the codes were sampled to compute the inter-coder reliability. This reliability coefficient, in the form of percent agreement, was 0.96, which was high enough to demonstrate consistency in coding across coders for this systematic review.

Meanwhile, three major steps were taken to evaluate the quality of evidence. First, each selected source was evaluated for its overall quality, following the benchmarks listed in Table 1. In this step, a binary code was created to describe the quality of an article in each of 20 aspects in the evaluative protocol in Appendix 2, and a numeric score was obtained by summing these binary codes. A source was rated as "High" in quality when its numeric score was above 15, "Moderate" when this score ranged from 10 to 15, and "Low" when this score was below 10. Putting it into perspective, given that the full score is 20, sources rated with "High" met over 75% of the minimum requirements for a rigorous study, sources rated with "Moderate" met 50%-75% of the minimum requirements for a rigorous study, and sources rated with "Low" did not meet 50% of the minimum requirements for a rigorous study. Corresponding to each hypothesis, Appendix 3 lists each selected source covered in the review, including the complete reference, the quality rating, and the geographical context.

Table 1. *Assessing a source: Quality ratings, benchmarks, and interpretations.*

Quality rating	Total score	Interpretation
High	>15	Meet over 75% of the minimum requirements.
Moderate	10-15	Meet 50%-75% of the minimum requirements.
Low	<10	Fail to meet 50% of the minimum requirements.

Note. Adapted from DFID (2014).

Second, the strength of the evidence for each hypothesis was evaluated separately, based on the information aggregated from the sources corresponding to each hypothesis. This overall strength was evaluated for *quality*, *size*, and *context* of the sources, following the benchmarks in Table 2. A body of evidence was rated "High" when over 30% of its sources were rated "High," "Moderate" when no more than 30% of its sources were rated "High" but more than 30% of its sources were rated "Moderate," and "Low" when no more than 30% of its sources were rated either "High" or "Moderate." As for evaluating the size, a body of evidence was rated "Large" when containing more than ten sources, "Medium" with six to ten sources, and "Small" with no more than five sources. When evaluating the geographical context, a body of evidence was considered "Global" when it covered sources from more than five countries and "Specific" when it covered no more than five countries.

Table 2. *Assessing a body of evidence: Ratings and benchmarks in three dimensions.*

Quality rating	Size rating	Context rating
High (>30% sources rated "High")	Large (>10 sources)	Global (>5 countries)
Moderate (≤30% sources rated "High" AND >30% studies rated "Moderate")	Medium (6-10 sources)	Specific (≤5 countries)
Low (≤30% sources rated "High" AND ≤30% sources rated "Moderate")	Small (≤5 sources)	

Note. Adapted from DFID (2014) and Ashley et al. (2014).

Third, an overall rating was assigned to each body of evidence given its ratings in all the three aforementioned dimensions. This overall rating can be "Very strong," "Strong," "Medium," "Limited," and "No evidence." Table 3 shows how each overall rating was calculated and interpreted. When a body of evidence was marked with "High" in quality, "Large" in size, and "Global" in context, it was rated "Very strong." When it was identified with "High" or "Moderate" in quality, "Large" or "Medium" in size, and "Global" in context, it was rated "Strong." When a body of evidence was found to be "Moderate" in quality, "Medium" in size, and "Specific" in context, it was rated "Medium." When it was "Moderate" in quality, "Small" in size, and "Specific" in context, it was rated "Limited." The body of evidence was flagged with "No evidence" when it covered few sources.

Table 3. *Overall ratings and their benchmarks and interpretations.*

Overall Rating	Benchmarks	Interpretation
Very strong	High Quality, Large Size, Global Context.	We are very confident about supporting/refuting the hypothesis. The body of evidence is highly credible and diverse.
Strong	High-to-Moderate Quality, Large-to-Medium Size, Global Context.	We are confident about supporting/refuting the hypothesis. The body of evidence is credible and diverse.
Medium	Moderate Quality, Medium Size, Specific Context.	We may support/refute the hypothesis. The body of evidence shows significant shortcomings. Contextual differences may affect the outcomes.

Limited	Moderate Quality, Small Size, Specific Context.	We may support/refute the hypothesis. The body of evidence shows very significant shortcomings. Contextual differences may substantially affect the outcomes.
No evidence	No/few sources exist.	There is no plausible evidence to support/refute the hypothesis.

Note. Adapted from DFID (2014).

4. Results

Technology-based assessments are known by various names. In broad strokes, they are digitalized and implemented via computers, tablets, or smartphones. In the literature, technology-based assessments were described using umbrella terms like "online assessment," "remote assessment," and "digital assessment." Meanwhile, a wide range of specialized terms were found to be addressing certain types of technology-based assessments. They included but are not limited to "computer-based assessment," which refers to the assessment delivered via a computer; "computer/computerized adaptive testing," which can assign different items to each learner based on the learner's ability; "gamified assessment," which refers to assessments that integrate elements of digital gaming; "e-portfolio," which is a digitalized portfolio documenting a learner's learning achievements for the purpose of evaluation.

The final pool of sources for the final analysis consisted of 34 sources. Initially, 913 sources from 2018-2022 were identified when using the search terms and strategies specified above. I removed 812 sources that were not relevant to technology-based evaluative activities, one source that did not address learning outcomes, and two sources that did not concentrate on learners. Of the remaining 98 sources, I excluded 47 sources addressing higher education or teacher training (instead of pre-college education), 15 sources without sufficient details on the technology-based assessment, and two extra sources that did not specify the geographical context of the study. Figure 2 visualizes how the final pool of sources was obtained, following the updated PRISMA guidelines (Page et al., 2021).

Evidence from the selected sources was synthesized below, when assessing each of the six pre-specified hypotheses for analyzing technology-based assessments. In this synthesis, the names related to technology-based assessments were kept as they were in the selected sources.

4.1.At the learning level

4.1.1.Hypothesis 1: Learners' learning achievement is better measured by technology-based assessments

The selected sources mostly discussed measurement-related issues in the context of assessments of ultimate learning achievements. This type of assessments was often described using "assessment of learning," "summative

assessment," "high-stakes assessment," "final exam," and "term paper." Three assumptions were found to undergird this hypothesis about measurement.

Ratings: To test the hypothesis about the measurement of technology-based assessments, a total of 15 sources were analyzed. When evaluating this body of evidence, the quality was rated "Moderate," the size was "Large," and the context was rated "Global." The overall strength of this body of evidence was considered "Strong." Table 4 lists a few examples of the selected sources, with the aggregated ratings at the bottom.

Table 4. *Examples of relevant sources: Measurement of learning achievement.*

Country	No.	Sample	Technology-based Assessment	Conclusion (in Measuring)
Indonesia	1	577 Grade-11 students from 6 high schools in Kulonprogo Regency, Indonesia, and 11 experts involved in assessing the CAT mechanism; data collected in 2018.	*CAT-PhysCriTS*, a CAT program, to assess critical thinking skills in physics; via computers.	Positive: High-precision measurement.
Germany	7	98 learners from Grade 4 in Germany; data collection year not specified.	A fixed-form test on mathematical competencies; via tablets.	Positive: Test mode effect is negligible.
United Kingdom	6	159 assessments taken by at least 5,000 students as part of UK's GCSEs, A Levels or equivalents; data collected in 2013-2017.	Multiple CATs assessed for their reliability and predictive validity; via statistical modeling.	Negative: Limited reliability, although the studied CATS showed fine results.
Overall Rating		Quality	Size	Context
Strong		Moderate	Large	Global

Note. Complete information of the listed sources here can be found in Appendix 3, using the source number specified in the "No." column.

Assumption 1: Higher precision. It is generally agreed that well-designed adaptive assessments have higher precision than the traditional paper-and-pencil tests. In the latter, most of the items have a high measurement precision only when applied to assess learners of an average ability, not learners of extremely high or extremely low abilities (Lord, 1980). Many technology-based assessments, such as computer adaptive tests (CAT; Wainer, 2000) based on item response theory (IRT; Hambleton & Swaminathan, 1985), are designed to improve precision by adapting items to each learner's ability. That said, many existing computer-based assessment (CBA) simply

administers traditional paper-based test items via computers without adapting items. It has similar measurement characteristics to the ones in paper-based tests.

Findings: With some exceptions, the cross-country evidence overall supported the assumption that technology-based assessments have higher precision than paper-based assessments. Supporting evidence was found in China, Cyprus, Germany, India, Indonesia, Malaysia, and Turkey. Mixed evidence was present in the context of Saudi Arabia and the United Kingdom. For instance, in Indonesia, *CAT-PhysCriTS*, a CAT program, met precision requirements for the measurement of higher-order skills, such as critical thinking in physics (Abidin et al., 2019). In India, multiple adaptive assessments backed by artificial intelligence (AI) technologies recommended appropriate learning contents and strategies as part of the formative assessment process (Jaiswal & Arun, 2021). On the other hand, in France, the measurement precision of a CBA was limited because open-ended questions were difficult to be integrated for automatic scoring (Auphan et al., 2020). Based on a few psychometric studies, it was proclaimed that a CAT does not necessarily result in improved measurement precision, and an evaluation of 159 CATs in the United Kingdom yielded no meaningful differences between CAT and non-CAT assessments: The average reliability of non-CATs was found to be 0.015 lower than that of CATand this difference was negligible (Benton, 2021).

Assumption 2: Reduced cheating. Cheating hampers the integrity of test scores and makes it hard to understand the actual learning achievement. It exists in both traditional paper-based assessments and technology-assisted assessments. Many technology-based assessments are straight CBAs that administer the same test items to all learners. This fixed form of assessment makes it easy for learners to copy each other's responses, which renders CBAs highly vulnerable to cheating. In CAT, however, each learner is assigned different items, which makes it hard to copy other learners' responses. That said, content constraints, exposure restrictions, and a limited number of available items for selection all limit the "adaptability" of CAT in real-world settings (Chuah et al., 2006), associated with higher susceptibility to cheating.

Findings: Compared with traditional paper-based assessments, would technology-based assessments be associated with reduced cheating? Mixed evidence was found in this regard from the selected literature, which could be explained by the varying cheating prevalence across different school climates, socioeconomic contexts, and technologies in assessments. It was generally recognized that cheating widely existed in competitive school climates like the one examined in Indonesia (Dwiyono et al., 2021). Yet, cheating was found to be more frequently discussed with CBAs than with CATs, indicating that the latter were likely associated with reduced cheating. For instance, in Indonesia, the CAT-driven national exam was designed to assign different items to learners and administered via computers, and it was believed to reduce cheating (Dwiyono et al., 2021), whereas the CBA-like formative assessments were1 contaminated by cheating among many students (Ibrohim et al., 2021). In Kazakhstan, most of the surveyed teachers claimed that cheating was serious in CBAs administered in 2021-2022 (Tashbolatovna et al., 2022). In Saudi Arabia, academic integrity, associated with learners' cheating behavior, was a concern to many teachers and administrators in online assessments (O'Keefe et al., 2020) that were straight CBAs.

Assumption 3: Easier interpretation. Limited standardization, reliability, and generalizability of a traditional paper-based assessment make its scores hard to interpret (Fuchs et al., 1994; Lord, 1980). This interpretability challenge also exists in the computer-based tests without adaptive item selection, given the reason described before. That said, even CBAs are often seen as a better alternative than traditional paper-based assessments in terms of interpretability because the digital mode of the former often allows for automatic scoring, documenting, and monitoring (Benton, 2021), and this autonomy supports any additional work needed to interpret scores. As for CAT- or AI-driven assessments, they experience increased popularity because millions of users find their results highly interpretable (Jaiswal & Arun, 2021).

Findings: Evidence from the selected literature overall was positive about the interpretability of scores from technology-based assessments, apart from a few caveats. The *CAT-PhysCriTS* program in Indonesia (Abidin et al., 2019) produced interpretable scores for measuring critical thinking skills in physics. AI-backed adaptive tools in India (Jaiswal & Arun, 2021) had been used by millions of K-12 learners and teachers because of their interpretable scores and recommendations. However, a case study in France showed that the autonomous scoring in a CBA still needs additional work to synthesize and understand the results (Auphan et al., 2020). Also, a cross-country analysis with data about 20 countries in Europe, Latin America, and Asia, found that learners of different genders or socioeconomic statuses differed in their ability to access and use technologies; this gap should be attended to when interpreting results from technology-based assessments (Ercikan et al., 2018).

4.1.2. Hypothesis 2: Learners are better supported in the learning process by technology-based assessments

Many technology-based assessments are designed to support learning. They were often known as "assessments for learning," "formative assessments," "game-based assessments," "online self-assessments," "web-based peer assessments," and "e-portfolio" in the literature. Three assumptions were identified in the selected sources in this aspect.

Ratings: For testing this hypothesis about the support of learning from technology-based assessments, a total of 17 sources were analyzed. When evaluating this body of evidence, the quality was rated "Moderate," the size was "Large," and the context was "Global." The overall strength of this body of evidence was considered "Strong." Table 5 lists a few examples of the selected sources with the aggregated ratings.

Table 5. *Examples of relevant sources: Support for learning.*

Country	No.	Sample	Technology-based Assessment	Conclusion (in Supporting)
Cyprus	29	24 7-8-yearold learners from the same Grade-2 class, at a public primary school in rural Cyprus;	*Helping Nemo*, a formative assessment based on an augmented reality (AR) game, to support learners'	Positive: Improved engagement of learners.

		data collection year not specified.	storytelling and writing; via tablets and computers.	
Malaysia	17	70 teachers in Grades 4-6 at public primary schools in Sarawak, Malaysia; data collection year not specified.	*Padlet*, an online program, for writing assessment; medium not specified.	Positive: User-friendly; highly engaging.
Turkey	32	46 Grade-7 learners from 2 classes at a secondary school in Turky; data collected in 2016.	*Kahoot!*, a game-based system that can deliver quizzes; via computers.	Positive: Instant feedback; much more fun.

Overall Rating	Quality	Size	Context
Strong	Moderate	Large	Global

Note. Complete information of the listed sources here can be found in Appendix 3, using the source number specified in the "No." column.

Assumption 1: Higher engagement in learning. The integration of technologies in assessments could enhance the learning experience by engaging learners in active learning via evaluation activities. Digital game-based applications, for instance, including *Kahoot!*, *Plickers*, *Quizizz*, and *Quizlet*, are widely accepted to be effective learning tools in which the major activity is assessment. They embrace game elements such as avatars, virtual teams, content unlocking, and game levels (Werbach & Hunter, 2012), which motivates learners to learn. Badges or bonus points are offered throughout the assessment process, which can be seen as positive reinforcements and lead to more participation (Kim, 2015).

Findings: The evidence in the selected sources supported the assumption that technology-based assessments trigger higher engagement than traditional paper-based assessments, despite cautions on the meaningful match between technological elements and instructional needs or purposes. In Malaysia, many school teachers embraced *Quizizz* because it was highly engaging (Yunus & Hua, 2021). In Turkey, an experimental learner group that used the game-based *Kahoot!* In classroom learning was identified with a higher engagement level than a control group that used a non-game-based assessment (Turan & Meral, 2018). In Spain, gamified natural science exams were found to be helping primary school learners continue self-learning, even after school (Sánchez-Rivas et al., 2019). In Cyprus, *Helping Nemo*, a formative assessment based on an augmented reality (AR) game, increased engagement for learners of diverse backgrounds while helping them achieve their learning goals (Stylianidou et al., 2020). In Indonesia, smartphone-based competitive games, as part of an assessment, boosted self-learning activities (Herwin et al., 2021).

Assumption 2: More interaction with others. Interaction, in the form of verbal or written communication in the assessment process, is scarce in these traditional paper-based assessments (Fawns & O'Shea, 2018). In contrast,

technology-based assessments often feature dynamic interaction throughout the assessment process, in the form of teachers', peers', or machines' instant responses and learners' habituation and adaptation to the responses. This interaction in effective technology-based assessments was expected to help enrich the learning experience (Gaytan & McEwen, 2007). Via this interaction, learners can adjust their strategies, improve communication and collaboration skills, build relationships, and learn from different perspectives.

Findings: The evidence in the selected sources was generally positive about this assumption concerning more interaction associated with technology-based assessments, although with cautions on the quality of feedback. In Malaysia, it was found that the game-based assessment application *Quizizz* cultivates the social relatedness of learners to others (Yunus & Hua, 2021); *Padlet*, a digital application to assess writing skills, enabled learners to share ideas, collaborate with each other, and learn from such interaction (Jong & Tan, 2021). AI-backed digital assessment tools in India foster human-machine interactions, which support learning by offering continuous evaluation and recommendations based on each learner's unique learning trajectory (Jaiswal & Arun, 2021). It was cautioned that feedback in interactions should align with the instructional content and learners' actual needs (Barana et al., 2021). For instance, the feedback should be succinct enough to avoid overwhelming learners (Jaiswal & Arun, 2021); in writing assessment, delayed, face-to-face feedback is seen as more effective than instant feedback from peers' online discussion in Singapore (Chan, 2021).

Assumption 3: Better learning outcomes. Sufficient guidance, high engagement, good interaction with healthy relationship building, and effective feedback are expected to improve learning outcomes. Technology-based assessments can integrate all these elements, as described above in this section. Guidance throughout an assessment is meant to help learners take independent actions in learning (Feng et al., 2009). Also, increased motivation and engagement had been found associated with improved learning outcomes (Headden & McKay, 2015). Additionally, increased interaction brought by technology-based assessments had been found to improve academic achievement (Yulia et al., 2019) and lead to positive learning growth via interactive feedback (Hattie, 2008).

Findings: In the selected literature, mixed evidence was found regarding whether technology-based assessments lead to better learning outcomes. In a case study in Malaysia, many learners' overall performance improved after being administered game-based assessments via *Quizizz* (Yunus & Hua, 2021). However, it is possible these score differences resulted from a mode difference (i.e., digital mode vs. traditional paper-based mode) instead of an actual learning improvement caused by technology-based assessments. Thus, one could argue that this study's pre-post design was insufficient to validate its findings. Nevertheless, this difference, or mode effect, was found negligible given a close-to-zero average effect size of the score mean difference between paper-based and computer-based assessments from a meta-analysis on Turkey (Yilmaz, 2021) and an experiment in Germany (Blumenthal & Blumenthal, 2020). Yet, it was cautioned that this mode effect may vary for different learner groups,

considering the digital competency gap between people of different genders or socioeconomic statuses (Ercikan et al., 2018)

4.2.At the teaching level

4.2.1.Hypothesis 3: Teachers are better prepared for teaching by technology-based assessments

Technology-based assessments are expected to support teachers' diverse roles throughout the assessment process, including identifying individual learners' needs and refining instructional strategies and lesson plans. Two assumptions were identified in the selected sources about the possible benefits of technology-based assessments to teachers' teaching.

Ratings: For testing this hypothesis about technology-based assessments' support to teachers' teaching, a total of 17 sources were analyzed. When evaluating this body of evidence, the quality was rated "Moderate," the size was "Large," and the context was "Global." The overall strength of this body of evidence was "Strong." Table 6 lists a few examples of the selected sources with the aggregated ratings.

Table 6. *Examples of relevant sources: Support for teaching.*

Country	No.	Sample	Topic	Conclusion (in Teaching)
Kazakhstan	31	60 teachers teaching at primary and secondary schools in Kazakhstan; data collected in 2021-2022.	Teachers' perceptions of online assessments.	Negative: Efficacy of online assessments needs to be improved.
Philippines	24	51 teachers from a secondary school in San Jose del Monte, Bulacan, Philippines; data collection year not specified.	Teachers' practices and competences in online assessments.	Negative: Challenges in implementation; varied teacher competencies in online assessment.
Spain	23	477 teachers at kindergartens across Spain; data collected in 2017-18.	Teachers' use of ICT in assessments.	Negative: Scarce use of technologies in assessments.
Overall Rating		Quality	Size	Context
Strong		Moderate	Large	Global

Note. Complete information of the listed sources here can be found in Appendix 3, using the source number specified in the "No." column.

Assumption 1: More effective in monitoring learning progress. With technology-based assessments, learners' responses can be immediately available upon completion of the assessment activity, and the scoring can be automatic (Alruwais et al., 2018). These features are not available in traditional paper-based assessments. Over time, the data collected via technology-based assessments are assumed to help teachers more effectively monitor each learner's growth and diagnose their strengths and limitations (Robles & Braathen, 2002).

Findings: Mixed evidence was found about whether technology-based assessments more effectively support teachers' monitoring of learning outcomes. In Spain and Turkey, for instance, teachers of young children used digital assessment tools—*Transparent Classroom* in Barcelona and *SAP Fiori* in Istanbul—that require teachers to submit forms and enter data to create learners' records and monitor their academic progress. These tools, not allowing for automatic recording, were cumbersome to many teachers, who mentioned that paper-based assessments would have been sufficient. Also, these assessment tools were identified with technical limitations, such as not recording children's socio-emotional learning data and operating difficulties on the Android system (Simsek & Tugluk, 2021). In another source on Spain, technologies are scarcely used in assessments (Romero-Tena et al., 2020). On the contrary, in India, many advanced AI-backed assessment tools have been accepted by a large number of teachers for classroom use because they enable automatic recording and provide data analytics to teachers (Jaiswal & Arun, 2021). In an international school in Vietnam, educators employed learning analytics in combination with CAT to derive insights into each learner's growth (Aristizábal, 2018).

Assumption 2: More effective instruction. A greater understanding of learners' development ultimately is expected to support teachers' decision-making in instructional design and inform timely remedies or interventions when necessary. Technology-based assessment makes it possible to inform teachers of the learning outcomes and special learners' learning needs in time. This immediacy can help teachers quickly adapt to the constantly changing learning situation, and better align their teaching practice with learners' needs (Wiliam, 2006).

Findings: Mixed evidence was found in the selected sources on whether teachers' use of technology-based assessments is associated with better instructional design. For instance, in a writing class in Malaysia, *Padlet* was found to be helpful in providing information on learners' writing progress, which teachers could then use to provide immediate feedback to students in the instructional process (Jong & Tan, 2021). *E-portfolio* is a digital assessment tool that collects artifacts showing individual learners' learning progress. In a case study in Indonesia, *e-portfolio* was used in the context of a flipped classroom and significantly improved learners' English-speaking performance (Kusuma et al., 2021). However, teachers' digital competencies vary across contexts, and the presence of technology-based assessments alone does not guarantee positive outcomes in instruction. In recent surveys, over 70% of teachers reported that they could not administer online assessments effectively (Tas et al., 2021).

4.2.2. Hypothesis 4: Teachers are better supported in non-teaching activities by technology-based assessments

In addition to classroom teaching, teachers are committed to many non-teaching tasks, such as documenting learners' performance for terminal reports and interacting with learners' families about learners' progress. In the

selected sources, two assumptions were found concerning technology-based assessments' roles in supporting teachers' non-teaching activities.

Ratings: For testing this hypothesis about technology-based assessments' support for teachers' non-teaching activities, a total of 10 sources were analyzed. When evaluating this body of evidence, the quality was rated "Moderate," the size was "Medium," and the context was "Global." The overall strength of this body of evidence was "Strong." Table 7 lists a few examples of the selected sources with the aggregated ratings.

Table 7. *Examples of relevant sources: Support for non-teaching activities.*

Country	No.	Sample	Topic	Conclusion (in Non-Teaching)
India	16	4 subject matter experts in technologies and 4 senior managers from 4 leading Indian educational technology firms serving millions of K-12 learners in India; data collection year not specified.	Benefits of artificial intelligence (AI) technologies in education.	Positive: Customization of assessment materials; reducing repetitive tasks.
Italy	12	At least 241 teachers teaching in primary and secondary schools; data collected in 2020.	Teachers' perceptions of the online assessment.	Positive: Better teacher-learner interaction after school time.
Spain, Turkey	27	11 teachers from Istanbul, Turkey; 3 teachers from Barcelona, Spain; at Montessori kindergartens; data collection year not specified.	Teachers' use of digital assessment with children.	Positive: Easy to use; effective documentation of children's growth and communication with parents.
Overall Rating		Quality	Size	Context
Strong		Moderate	Medium	Global

Note. Complete information of the listed sources here can be found in Appendix 3, using the source number specified in the "No." column.

Assumption 1: Reduced workload. In traditional paper-based exams, a massive portion of teachers' time is invested into assessment question design, scoring, and distribution to learners. Many digital assessment tools allow for automatic recording and scoring, which frees up time that would normally be spent manually recording and grading in the paper-based mode. Also, these tools may come with rich assessment materials that facilitate the instructional

design and allow teachers to avoid repetitive work in distributing assessments to each learner. The traditional marking process can be burdensome and time-consuming, too slow to help learners benefit from the feedback (Al-A'ali, 2007), and error-prone (Cheang et al., 2003), particularly when teachers are juggling multiple tasks.

Findings: Mixed evidence was found in this domain concerning whether technology-based assessments help reduce teachers' workload in non-teaching activities. In Italy, 90% of the teachers in a survey mentioned that digital tools in assessment and other online educational activities expanded their work time (Ferretti et al., 2021). In Turkey, the *Education Informatics Network* (EBA) is a nationally recognized platform for digital learning that devotes part of its functions to assessment. Many teachers in a survey claimed that there were not enough test items in the *EBA* to support them (Tas et al., 2021). Teachers in Turkey also reported that technological limitations with *SAP Fiori* resulting from a slow system and required data entry wasted their time (Simsek & Tugluk, 2021). In India, contrarily, with CAT-driven assessment tools, dynamic question papers were prepared automatically for each student, which helped reduce a large part of teachers' time spent preparing exam sheets (Jaiswal & Arun, 2021). In Kazakhstan, a good portion (20%-35%) of surveyed teachers claimed that digital assessments saved them time (Tashbolatovna et al., 2022).

Assumption 2: Smoother communication with parents. Parents play an essential role in the education system (Care & Anderson, 2016). There is a wide range of activities related to parental involvement at home and in school that can affect learners' learning, including parents' participation in the curricular design for children at home or in school and their engagement in classroom activities and school-wide events with their children (Jowett & Baginsky, 1988). Unsurprisingly, teachers can spend a large amount of their work time communicating with parents to update them of their children's progress and solicit their feedback on different occasions about what could be done to improve the current work.

Findings: The evidence from the selected sources in this arena generally supported the assumption that technology-based assessments facilitate teachers' communication with parents. Three technology-based assessment systems found in three countries—*EBA* in Turkey, *KITALINO* in Germany, and *Transparent Classroom* in Spain—were found to be equipped with log-in components and interfaces that allow parents to access their children's information using personalized usernames and passwords. Teachers can use these tools to create reports about learners' progress and share them with parents (Özdil et al., 2021). Another source showed that teachers could use digital assessment tools to efficiently share selected information with parents (Simsek & Tugluk, 2021). In Indonesia, the benefits of the computer-based national examinations included faster and more detailed results to parents from computers (Dwiyono et al., 2021), saving teachers' efforts.

4.3.At the management level

Following Lock (2007), it is vital to strike a balance among quality, cost, and time (i.e., the Iron triangle). Without this equilibrium, a technology-based assessment may not be successfully implemented. For instance, overemphasizing quality could push up cost and time in assessment administration, making the technology-based assessment less

appetizing and consequently incurring institutional resistance against its coming into being. The evidence synthesized above primarily reflects the quality of technology-based assessments (i.e., how they supported learning and educating). Below came the evidence on the cost and time associated with technology-based assessments.

4.3.1.Hypothesis 5: Technology-based assessments are more cost-effective in education

The economic efficiency of a technological project should be appraised before the project is applied at a large scale. Also, the associated cost should be calculated continuously throughout the entire life cycle of any technological product. Following Grunwald (2009), the costs of technology-based assessments can include (1) the cost to develop this product (i.e., "development costs"), (2) the cost to make it come to life (i.e., "manufacturing costs"), (3) the cost to use and maintain it (i.e., "operating costs"), and (4) the cost to dispose of it (i.e., "waste disposal costs"). Adapted from Jung and Rha's (2000) conceptualization of online education, the cost-effectiveness factors in a technological assessment may entail but not be limited to (1) the number of learners affected by the assessment, (2) the number of courses related to the assessment, (3) the type of medium used for the assessment, (4) the options for meeting different learners' needs, and (5) the assessment's usage rate.

Ratings: For testing this hypothesis about the managerial cost of technology-based assessments, a total of five sources were analyzed. When evaluating this body of evidence, the quality was rated "Moderate," the size was "Small," and the context was "Specific." The overall strength of this body of evidence was "Limited." Table 8 lists out a few examples of the selected sources with the aggregated ratings.

Table 8. *Examples of relevant sources: Cost of technology-based assessments.*

Country	No.	Sample	Technology-based Assessment	Conclusion (in Cost)
Germany	7	98 learners from Grade 4 in Germany; data collection year not specified.	A fixed-form test on mathematical competencies; via tablets.	Positive: Reducing expenditures when frequently using the computer-based exam.
Indonesia	10	5 teachers from a vocational high school in the northern part of Sangatta, East Kutai Regency, Indonesia; data collected around 2016.	The national exam; via computers.	Positive: Saving the cost from printing and distributing physical exam papers each year.
Germany, Spain, Turkey	22	Evidence in XX only.		
Overall Rating		Quality	Size	Context
Limited		Moderate	Small	Specific

Note. Complete information of the listed sources here can be found in Appendix 3, using the source number specified in the "No." column.

Findings: The evidence in the selected sources about the cost of technology-based assessments was somewhat limited, but it overall supported the assumption that the integration of technologies helps reduce the cost of assessment. It was generally agreed that technology-based assessments help reduce costs because they save educators from expenditures related to printing and distribution that are unavoidable in traditional paper-based exams (Blumenthal & Blumenthal, 2020; Dwiyono et al., 2021; Özdil et al., 2021). Additionally, technology-based assessments can be resource-efficient in their assessment designs (Abidin et al., 2019), reducing the administrative cost of designing new assessments. Despite that, the available evidence did not paint a complete picture of cost-effectiveness: it primarily concentrated on the "manufacturing costs" reduced by technology-based assessments, while generally neglecting possible costs associated with developing, operating, and disposing of technological assessment tools. These aspects of cost-effectiveness are understudied in the literature, and future research is imperative to provide more information.

4.3.2.Hypothesis 6: Technology-based assessments are more time-efficient in educational changes

Time efficiency is the third dimension of managing technology-based assessments, following the Iron Triangle. It is the constraint associated with the scheduled completion of a project. Here, time efficiency can specifically refer to the time spent implementing technology-based assessments in educational contexts. At the management level, it concerns whether there is enough committed time that matches the scope of a technology-based assessment. In other words, time efficiency is about whether technology-based assessments are prioritized with a reasonable agenda at the level of local or national decision-making. For comparing technology-based assessments with traditional paper-based ones, a natural question to ask is whether technology-based assessments are more time-efficient than paper-based assessments, on the condition that the quality and cost of the two are approximately the same.

Ratings: For testing this hypothesis about the time invested in technology-based assessments, a total of six sources were analyzed. When evaluating this body of evidence, the quality was rated "Moderate," the size was "Medium," and the context was "Global." The overall strength of this body of evidence was "Strong." Table 9 lists out a few examples of the selected sources with the aggregated ratings.

Table 9. *Examples of relevant sources: Invested time for technology-based assessments.*

Country	No.	Sample	Technology-based Assessment	Conclusion (in Time)
France	4	687 learners from Grades 2-9, at urban schools in the eastern part of France;	A newly designed reading assessment; via computers.	Mixed: Fast to implement with automatic recording

		data collection year not specified.		and scoring; but not for open-ended tasks.
Indonesia	1	577 Grade-11 students from 6 high schools in Kulonprogo Regency, Indonesia, and 11 experts involved in assessing the CAT mechanism; data collected in 2018.	*CAT-PhysCriTS*, a CAT program, to assess critical thinking skills in physics; via computers.	Positive: Fast measurement without compromising precision.
Vietnam	2	340 learners and 32 teachers in Grades 6-10 from an international school in Vietnam; data collected in 2017-2018.	Use of CAT data and learning analytics to improve learning in mathematics and reading.	Positive: Learning outcomes were reported in a timely manner.

Overall Rating	Quality	Size	Context
Strong	Moderate	Medium	Global

Note. Complete information of the listed sources here can be found in Appendix 3, using the source number specified in the "No." column.

Findings: Evidence from the selected sources did not directly address this hypothesis from a managerial perspective, but it nevertheless shed light on this dimension of technology-based assessments. It was generally agreed that technology-based assessments require less time to implement than traditional paper-based exams because of features like automatic scoring and instant score reporting (e.g., Abidin et al., 2019; Aristizábal, 2018; Auphan et al., 2020; Blumenthal & Blumenthal, 2020; Dwiyono et al., 2021; Özdil et al., 2021). At the management level, these save-time features suggest that technology-based assessments can be time-efficient for the institutional practice of documenting and reporting. More case studies examining the time efficiency of technology-based assessments at an institutional level are needed in this area of research.

4.4.Sensitivity analysis

The sensitivity analysis displayed in Table 10 shows the contribution of the selected sources to the synthesized evidence about the utilization of technology-based assessments. The sources were not evenly dispersed across the three layers (learning, educating, and management).. Slightly over half (62%) of the selected sources concentrated on one single layer to examine relevant issues in technology-based assessments at a deep level. A great portion of the sources targeted technology-based assessment issues in "teaching" (53%) of the educating process, as well as "measuring" (44%) and "supporting" (47%) for learners' learning. A smaller portion addressed the support of technology-based assessments in "non-teaching" (29%), "cost" (15%), and "time" (18%).

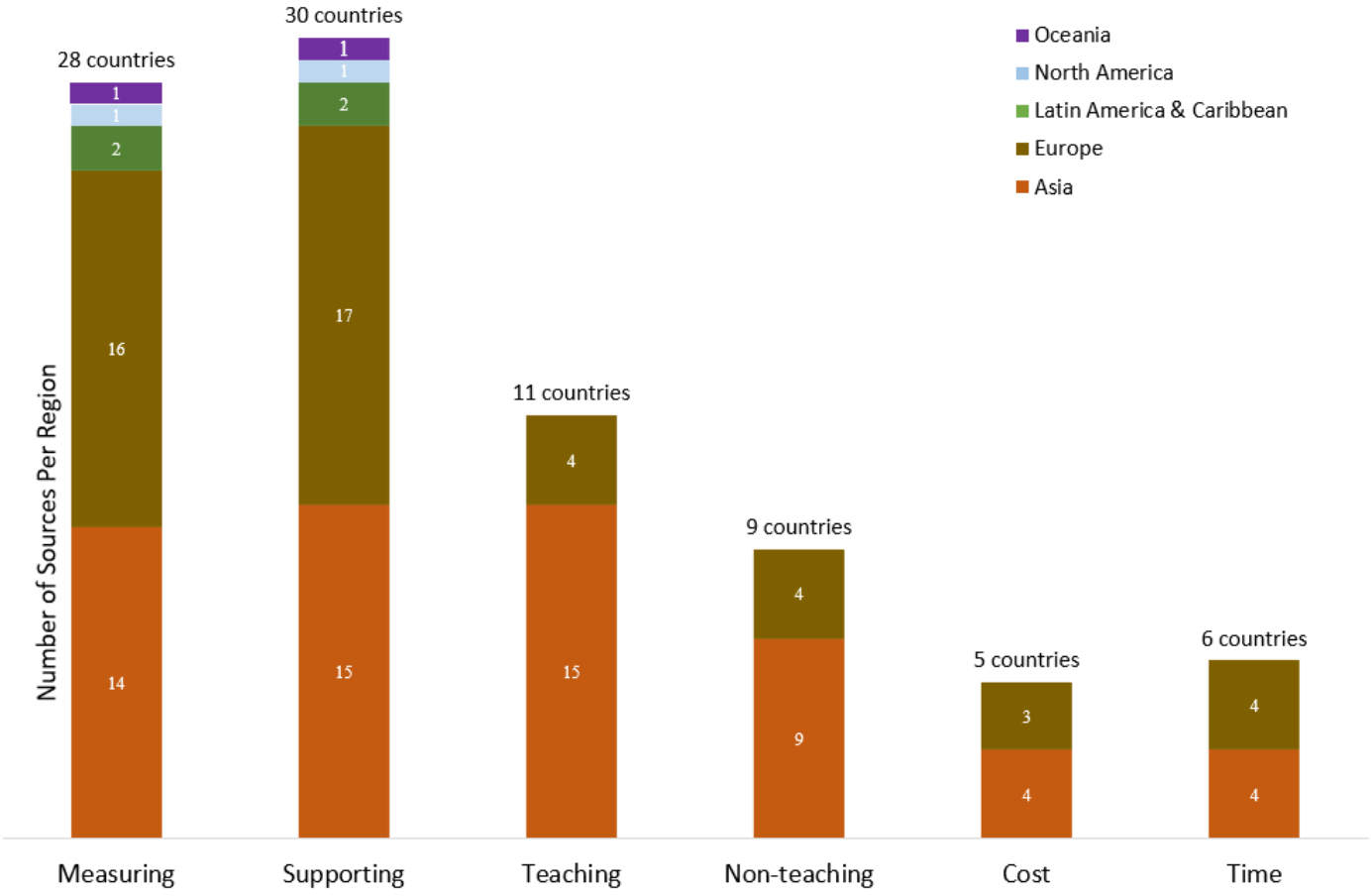
Figure 3 visualizes the geographical representation of the selected sources, specific to each hypothesis. Each unique color represents a unique geographic region, and the numbers inside the colored bars are the numbers of sources per region, for which one source covering multiple regions was counted multiple times. A total of five unique regions, covering 34 unique countries, were included in the selected sources. While most sources were related to teaching, as demonstrated above, the widest range of countries was covered by the evidence on learning (28 countries for “measuring” and 30 countries for “supporting”), followed by the evidence on the benefits of technology-based assessment for teachers’ “teaching” (11 countries). The distribution of countries was partly associated with the number of sources addressing each hypothesis, as described above. Within each body of evidence, the maximum number of sources per country was four, for Indonesia in Hypothesis 3; Malaysia was represented by three sources in Hypotheses 1 and 2, and Turkey was covered by three in Hypotheses 2 and 3.

Table 10. *Sensitivity analysis: Contribution of each selected source to the synthesized evidence.*

		Learning		Educating		Management	
No.	Source	Measuring	Supporting	Teaching	Non-teaching	Cost	Time
1	Abidin et al. (2019)	✗				✗	✗
2	Aristizábal (2018)			✗	✗	✗	✗
3	Aslam et al. (2020)			✗			
4	Auphan et al. (2020)	✗	✗				✗
5	Barana et al. (2021)		✗	✗			
6	Benton (2021)	✗					
7	Blumenthal & Blumenthal (2020)	✗	✗			✗	✗
8	Chan (2021)		✗	✗			
9	Çimen (2022)			✗			
10	Dwiyono et al. (2021)	✗			✗	✗	✗
11	Ercikan et al. (2018)	✗	✗				
12	Ferretti et al. (2021)			✗	✗		
13	Herwin et al. (2021)			✗			
14	Ibrohim et al. (2021)	✗		✗	✗		
15	Istiningsih (2022)			✗			
16	Jaiswal & Arun (2021)	✗	✗	✗	✗		
17	Jong & Tan (2021)	✗	✗	✗			
18	Kusuma et al. (2021)		✗	✗			
19	Liu (2022)	✗					
20	Lubiano (2018)		✗				
21	O'Keefe et al. (2020)	✗	✗				
22	Özdil et al. (2021)				✗	✗	✗
23	Romero-Tena et al. (2020)			✗			
24	Samosa (2022)			✗	✗		
25	Samsudin et al. (2020)	✗	✗				
26	Sánchez-Rivas et al. (2019)		✗				
27	Simsek & Tugluk (2021)			✗	✗		
28	Singh et al. (2022)			✗			
29	Stylianidou et al. (2020)	✗	✗				
30	Tas et al. (2021)			✗	✗		
31	Tashbolatovna et al. (2022)	✗		✗	✗		
32	Turan & Meral (2018)		✗				
33	Yilmaz (2021)	✗	✗				
34	Yunus & Hua (2021)		✗				

Note. The three shades used in the "Source" column indicate the quality of studies: ☐ – High quality, ☐ – Medium quality, ☐ – Low quality.

Figure 3. *Regions covered in each body of evidence.*



5. Discussion

While the digitalization trend is a defining feature of education in the current 21st century (Chen, 2021), there remains a digital gap in different countries' ability to access and use technology-based assessments. The switch from the traditional paper-based exam to the technology-based assessment has happened in many countries, including , Finland, Israel, Mexico, New Zealand, and the United States (Ercikan et al., 2018; Meadows, 2021), to name a few. Meanwhile, many other countries have been relatively isolated from this innovation, including Indonesia (Dwiyono et al., 2021; Kusuma et al., 2021) and Turkey (Çimen, 2022). Some countries (e.g., Germany), despite trying to catch up with this technological trend, have struggled to do so due to a limited supply of resources, such as computers, tablets, and digital terminals (Blumenthal & Blumenthal, 2020).

Via the present systematic review, both negative and positive evidence was found when testing hypotheses and related assumptions in the learning, educating, and management layers of technology-based assessments. This mixture was expected because digital tools, institutional support, teachers' competency, and social expectations in using technologies in assessments all vary across socioeconomic contexts. In summary, when compared with the traditional paper-based exams, *mixed* evidence was found about technology-based assessments' roles when testing assumptions about cheating reduction, learning boost, monitoring support, instructional improvement, and non-teaching workload reduction. It is worth noting that *strong supporting* evidence was found when testing assumptions regarding higher measurement precision, easier interpretation, higher learner engagement, and more interaction with others at the learning level, in addition to smoother communication with parents at the educating level. *Limited but positive* evidence at the management level suggested that technology-based assessments are more cost-effective and time-efficient when compared with paper-based ones.

A few important barriers to more effective usage of technology-based assessments were identified from this selected literature. First, *supportive policies and infrastructure for integrating technologies in assessments could be lacking* in a few countries. For instance, an investigation of secondary schools across 19 cities in Turkey found that almost all the teachers employed traditional paper-based assessments, not technology-based ones (Çimen, 2022). Few secondary schools in Indonesia allow students to bring digital devices to learn at school (Kusuma et al., 2021). In Saudi Arabia, many schools do not approve the application of technology-based assessments (O'Keefe et al., 2020). In Germany, technology-driven decision-making with assessment data has only partially gained acceptance in many schools (Blumenthal & Blumenthal, 2020). Second, *digital applications could be misaligned with learners' and educators' actual needs*. Educators should perform pre-assessments of the desired usage and outcomes of technology-based assessments before digitally transforming assessments (Özdil et al., 2021), and expectations of the institution and different stakeholders with respect to the transformation should be continuously communicated (O'Keefe et al., 2020). Third, *there could be little support from parents*. Families play an important role in the success of technology-based assessments; when parents do not allow their children to use digital gadgets to learn at home

(Ibrohim et al., 2021) or consider schools' technology-based assessment programs valuable (Simsek & Tugluk, 2021), they limit the effectiveness of technology-based assessments. Lastly, *many teachers could have limited competencies in technology-based assessments and low self-efficacy* in their ability to use them. They were found in countries such as Indonesia (Dwiyono et al., 2021) and Kazakhstan (Tashbolatovna et al., 2022).

The mixed signals from this systematic review demonstrate that the most important consideration for educators and institutions is the suitability of technology-based assessment to their specific sociocultural context. This suitability includes but is not limited to psychometric fit with required standards, content relevance to the chronological school grade or knowledge/skill domain, and practice alignment with institutional rules or social expectations. There is no one-size-fits-all resolution, given the complexities that make each context unique. Hence, it will be critical for educators, administrators, and policymakers to adapt their choices about technology-based assessments to reality. They are the individuals most familiar with the needs and pain points of their local education system through their daily work; they should, accordingly, be empowered to make important decisions about whether to join the digitalization trend in assessment. For instance, when a digital assessment system in Turkey was facing glitches and wasting time, a teacher mentioned it would be sufficient to assess learners via the paper-based exam (Simsek & Tugluk, 2021). When geographic regions lack the basic infrastructure and access necessary to support the implementation of standardized digital assessments with high psychometric rigor, it would be judicious to seek support from community-based, citizen-led approaches as alternatives to technology-based assessment. For instance, many rural communities in Asian and African countries seek assessment support from ASER centers via the cross-continental network People's Action for Learning (PAL Network, n.d.).

The sensitivity analysis revealed that there is limited literature on technology-based assessments' roles in teachers' non-teaching activities and management-level investment in cost and time. Further research is needed in these areas. Teachers' non-teaching activities can encompass building connections with learners, developing relationships with parents, and collaborating with colleagues. These activities are as important as teachers' teaching activities and can matter more to a learner's long-term growth and lifelong interest in learning. Regarding the cost issue, the World Bank has published a cost analysis of technological projects in education (Potashnik & Adkins, 1996) and a primer on the possible costs of a large-scale national assessment (Clarke & Luna-Bazaldua, 2021). Both fall short because they did not vet the usage rate or waste disposal costs of assessments, and the latter did not address the costs associated with technology. As for the possible time constraints, a Brookings report reflected that at least two problems exist with assessment programs across countries: data collection cycles are long, and a lag exists between the collection and use of assessment data, due to the time-intensive nature of data processing and reporting in existing assessment programs (Vista et al., 2018). It remains unknown whether these time-related issues could be resolved by integrating advanced technologies in assessments at the high level.

This systematic review was limited by the varied quality of the selected sources. A few of these sources did not specify the subject area or grade covered by the technology-based assessment and did not disclose the year when

the data were collected. Also, many of these articles in English contained grammatical errors because the authors were not from English-speaking countries. Six of these sources contained a conflict-of-interest statement declaring that there was no interest conflict, and ten contained a statement about whether the study was funded. Eight sources were supported by external funds from research institutions or government, and no source was supported by an assessment company. Nevertheless, all the selected sources were considered valuable and retained in the final analysis because they met the eligibility criteria, offered practical insights relevant to our hypotheses, and highlighted context-specific realities across numerous countries.

ACKNOWLEDGEMENTS

The author wishes to express sincere appreciation and gratitude to Dr. Priyadarshani Joshi (Senior Research Analyst, UNESCO) and Dr. Patrick Montjourides (Senior Program Specialist, UNESCO), who offered considerable feedback throughout the writing process for this background report. Also, the author wishes to thank Dr. Jinming Zhang (Professor, UIUC) for his support and the following doctoral colleagues at UIUC for their excellent assistance (alphabetized by last name): Amos Jeng, Bradley Kaptur, Shiyu Wang, Jing Zhang, and Yingbin Zhang. Any remaining errors or omissions are the author's own.

The comments and conceptualizations presented here represent those of the author and should not be attributed to any individual or organization mentioned above.

REFERENCES

Abidin, A. Z., Istiyono, E., Fadilah, N., & Dwandaru, W. S. B. (2019). A computerized adaptive test for measuring the physics critical thinking skills. *International Journal of Evaluation and Research in Education*, *8*(3), 376–383.

Al-A'ali, M. (2007). Implementation of an Improved Adaptive Testing Theory. *Educational Technology & Society*, *10*(4), 80–94.

Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of using e-assessment. *International Journal of Information and Education Technology*, *8*(1), 34–37. https://doi.org/10.18178/ijiet.2018.8.1.1008

Aristizábal, J. (2018). Using learning analytics to improve students' reading skills: A case study in an American international school with English as an additional language (EAL) students. *GIST Education and Learning Research Journal*, *17*, 193–214.

Ashley, L., Mcloughlin, C., Aslam, M., EngeL, J., Wales, J., Rawal, S., Batley, R., Kingdon, G., Nicolai, S., & Rose, P. (2014). *The role and impact of private schools in developing countries: A rigorous review of the evidence* (Education Rigorous Literature Review). Department for International Development.

Auphan, P., Ecalle, J., & Magnan, A. (2020). The high potential of computer-based reading assessment. *Canadian Journal of Learning and Technology*, *46*(1), 1–23.

Barana, A., Marchisio, M., & Sacchet, M. (2021). Interactive feedback for learning mathematics in a digital learning environment. *Education Sciences*, *11*, 279–299.

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Computer-Based Testing and the Internet*, *1*(1), 201–217. https://doi.org/10.1002/9780470712993.ch11

Benton, T. (2021). Item response theory, computer adaptive testing and the risk of self-deception. *Research Matters*, *32*, 82–100.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Blumenthal, S., & Blumenthal, Y. (2020). Tablet or paper and pen? Examining mode effects on German elementary school students' computation skills with curriculum-based measurements. *International Journal of Educational Methodology*, *6*(4), 669–680.

Burns, M. (2021). *Technology in education: Background paper prepared for the 2023 Global Education Monitoring Report* (Report No. ED/GEMR/MRT/2023/T1/1). UNESCO Global Education Monitoring Report team. https://learningportal.iiep.unesco.org/en/library/background-paper-prepared-for-the-2023-global-education-monitoring-report-technology-and

Card, N. A. (2012). *Applied meta-analysis for social science research*. The Guilford Press.

Care, E., & Anderson, K. (2016). *How education systems approach breadth of skills* (Skills for a Changing World). The Brookings Institution.

Chan, K. T. (2021). Embedding formative assessment in blended learning environment: The case of secondary Chinese language teaching in Singapore. *Education Sciences*, *11*, 360–371.

Cheang, B., Kurnia, A., Lim, A., & Oon, W.-C. (2003). On automated grading of programming assignments in an academic institution. *Computers & Education*, *41*(2), 121–131. https://doi.org/10.1016/S0360-1315(03)00030-7

Chen, D. (2021). Toward an understanding of 21st-century skills: From a systematic review. *International Journal for Educational and Vocational Guidance*. https://doi.org/10.1007/s10775-021-09511-1

Chuah, S. C., Drasgow, F., & Luecht, R. (2006). How big is big enough? Sample size requirements for CAST item parameter estimation. *Applied Measurement in Education*, *19*(3), 241–255. https://doi.org/10.1207/s15324818ame1903_5

Çimen, S. S. (2022). Exploring EFL assessment in Turkey: Curriculum and teacher practices. *International Online Journal of Education and Teaching*, *9*(1), 531–550.

Clarke, M. (2012). *What matters most for student assessment systems: A framework paper* (Working Paper No.1). World Bank Group. https://openknowledge.worldbank.org/bitstream/handle/10986/17471/682350WP00PUBL0WP10READ0web04019012.pdf?sequence=1&isAllowed=y

Clarke, M., & Luna-Bazaldua, D. (2021). *Primer on large-scale assessments of educational achievement*. World Bank. https://doi.org/10.1596/978-1-4648-1659-8

DFID. (2014). *Assessing the strength of evidence* (How to Note). Department for International Development. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf

Dwiyono, Y., Mulawarman, W. G., Pramono, P. O., Salim, N. A., & Ikhsan, M. (2021). Implementation of national examination based on computer based test at Vocational School 1 North Sangatta. *Cypriot Journal of Educational Sciences*, *16*(1), 86–95.

Ercikan, K., Asil, M., & Grover, R. (2018). Digital divide: A critical context for digitally based assessments. *Education Policy Analysis Archives*, *26*(51), 1–18.

Fawns, T., & O'Shea, C. (2018). Distributed learning and isolated testing: Tensions in traditional assessment practices. *Proceedings of the 11th International Conference on Networked Learning*, 132–139.

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*(3), 243–266. https://doi.org/10.1007/s11257-009-9063-7

Ferretti, F., Santi, G. R. P., Del Zozzo, A., Garzetti, M., & Bolondi, G. (2021). Assessment practices and beliefs: Teachers' perspectives on assessment during long distance learning. *Education Sciences*, *11*, 264–280.

Fisher, M., Maylahn, P., Pota, V., & Rappallini, J. (2022). *Effective assessment and progress monitoring in an online environment: A study in six countries*. T4 Education. https://24886661.fs1.hubspotusercontent-eu1.net/hubfs/24886661/T4_Assessment%20and%20Progress%20Monitoring%20Online-F1.pdf

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1994). Strengthening the connection between assessment and instructional planning with expert systems. *Exceptional Children*, *61*(2), 138–147.

Gaytan, J., & McEwen, B. C. (2007). Effective online instructional and assessment strategies. *American Journal of Distance Education*, *21*(3), 117–132. https://doi.org/10.1080/08923640701341653

Grunwald, A. (2009). Technology assessment: Concepts and methods. In *Philosophy of Technology and Engineering Sciences* (pp. 1103–1146). Elsevier. https://doi.org/10.1016/B978-0-444-51667-1.50044-6

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer-Nijhoff Pub. ; Distributors for North America, Kluwer Boston.

Hattie, J. A. C. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (1st edition). Routledge.

Headden, S., & McKay, S. (2015). *Motivation matters: How new research can help teachers boost student engagement*. Carnegie Foundation for the Advancement of Teaching. https://www.carnegiefoundation.org/wp-content/uploads/2015/07/Motivation_Matters_July_2015.pdf

Herwin, H., Hastomo, A., Saptono, B., Ardiansyah, A. R., & Wibowo, S. E. (2021). How elementary school teachers organized online learning during the COVID-19 pandemic? *World Journal on Educational Technology: Current Issues*, *13*(3), 437–449.

Ibrohim, Sudrajat, A. K., & Saefi, M. (2021). Assessing Indonesian teachers' perspective on the implementation of distance learning due to COVID-19 based on online survey. *Journal of Turkish Science Education*, *18*, 46–59.

Jaiswal, A., & Arun, C. J. (2021). Potential of artificial intelligence for transformation of the education system in India. *International Journal of Education and Development Using Information and Communication Technology*, *17*(1), 142–158.

Jong, B., & Tan, K. H. (2021). Using Padlet as a technological tool for assessment of students' writing skills in online classroom settings. *International Journal of Education and Practice*, *9*(2), 411–423.

Jowett, S., & Baginsky, M. (1988). Parents and education: A survey of their involvement and a discussion of some issues. *Educational Research*, *30*(1), 36–45. https://doi.org/10.1080/0013188880300105

Jung, I., & Rha, I. (2000). Effectiveness and cost-effectiveness of online education: A review of the literature. *Educational Technology*, *40*(4), 57–60.

Kim, B. (2015). Designing gamification in the right way. *Library Technology Reports*, *51*(2), Article 2.

Kusuma, I. P. I., Mahayanti, N. W. S., Adnyani, L. D. S., & Budiarta, L. G. R. (2021). Incorporating e-portfolio with flipped classrooms: An in-depth analysis of students' speaking performance and learning engagement. *The JALT CALL Journal*, *17*(2), 93–111.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.

Lock, D. (2007). *Project management* (9th ed.). Gower Publishing Company Ltd.

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates, Inc.

Meadows, M. (2021, June 15). *Dr Michelle Meadows speech at City of London Schools Conference 2021*. GOV.UK. https://www.gov.uk/government/speeches/dr-michelle-meadows-speech-at-city-of-london-schools-conference-2021

Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research*, *5*(2), 147–149. https://doi.org/10.1177/104973239500500201

O'Keefe, L., Dellinger, J. T., Scragg, B., Amelina, N., & Mathes, J. (2020). The state of online learning in the Kingdom of Saudi Arabia: A COVID-19 impact study for K-12. In *Online Learning Consortium*. Online Learning Consortium, Inc. https://eric.ed.gov/?id=ED608877

Özdil, G., Simsek, F. M., & Tugluk, M. N. (2021). Comparison of digital assessment and documentation systems used in the early childhood education in Turkey, Germany and Spain. *Southeast Asia Early Childhood*, *10*(2), 1–15.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

PAL Network. (n.d.). *Assessment: Tests and tools*. Retrieved August 22, 2022, from https://palnetwork.org/tools/

Potashnik, M., & Adkins, D. (1996). *Cost analysis of information technology projects in education: Experiences from developing countries*. World Bank Group. https://documents1.worldbank.org/curated/en/444741468247804554/pdf/multi-page.pdf

Robles, M., & Braathen, S. (2002). Online assessment techniques. *Delta Pi Epsilon Journal*, *44*(1), 39–49.

Romero-Tena, R., Lopez-Lozano, L., & Gutierrez, M. P. (2020). Types of use of technologies by Spanish early childhood teachers. *European Journal of Educational Research*, *9*(2), 511–522.

Sánchez-Rivas, E., Ruiz-Palmero, J., & Sánchez-Rodríguez, J. (2019). Gamification of assessments in the natural sciences subject in primary education. *Educational Sciences: Theory and Practice*, *19*(1), 95–111.

Simsek, F. M., & Tugluk, M. N. (2021). Making learning visible in the 21st century; examining of the use of digital assessment tools in Montessori education. *Pegem Journal of Education and Instruction*, *11*(2), 72–86.

Stylianidou, N., Sofianidis, A., Manoli, E., & Meletiou-Mavrotheris, M. (2020). "Helping Nemo!"—Using augmented reality and alternate reality games in the context of universal design for learning. *Education Sciences*, *10*, 95–118.

Tas, Y., Eminoglu, S., Atila, G., Yildiz, Y., & Bozkurt, U. (2021). Teachers' self-efficacy beliefs and opinions about distance education during the COVID-19 pandemic. *Turkish Online Journal of Distance Education*, *22*(4), 229–253.

Tashbolatovna, M. K., Duisenbaevna, U. Z., Kaldyhanovna, K. R., Koyshibaevna, A. G., Ulmeken, Z., & Gulzat, S. (2022). Development of the evaluative activities of teachers in the conditions of updated education. *Cypriot Journal of Educational Sciences*, *17*(4), 1304–1319.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357. https://doi.org/10.1093/intqhc/mzm042

Turan, Z., & Meral, E. (2018). Game-based versus to non-game-based: The impact of student response systems on students' achievements, engagements and test anxieties. *Informatics in Education*, *17*(1), 105–116.

Vista, A., Kim, H., & Care, E. (2018). *Use of data from 21st century skills assessments: Issues and key principles* (Optimizing Assessment for All). The Brookings Institution. https://www.brookings.edu/wp-content/uploads/2018/10/EffectiveUse-Vista-Kim-Care-10-2018-FINALforwebsite.pdf

Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Lawrence Erlbaum Associates, Inc.

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, *11*(3–4), 283–289. https://doi.org/10.1080/10627197.2006.9652993

Yamamoto, K., Khorramdel, L., & Shin, H. J. (2018). Introducing multistage adaptive testing into international large-scale assessments designs using the example of PIAAC. *Psychological Test and Assessment Modeling*, *60*(3), 347–368.

Yilmaz, F. (2021). Comparing paper-pencil and computer-based tests: A meta-analysis study in the sample of Turkey. *Eurasian Journal of Educational Research*, *93*, 279–300.

Yulia, A., Husin, N. A., & Anuar, F. I. (2019). Channeling assessments in English language learning via interactive online platforms. *Studies in English Language and Education*, *6*(2), Article 2. https://doi.org/10.24815/siele.v6i2.14103

Yunus, C. C., & Hua, T. K. (2021). Exploring a gamified learning tool in the ESL classroom: The case of Quizizz. *Journal of Education and E-Learning Research*, *8*(1), 103–108.

APPENDICES

Appendix 1: Search Queries

Themes	Search terms (including related terms)
Assessment	assessment OR test
	AND
Technology	technology OR digital OR computer-based OR adaptive OR game-based OR online
	AND
Learner	learner OR examinee OR student OR pupil
	AND
Achievement	achievement OR competency OR ability
	AND
Quality	quality OR accuracy OR reliability OR validity
	NOT
Higher Edu	higher education OR college students OR undergraduate students OR college faculty
	NOT
United States	United States OR US

Appendix 2: Protocol for Assessing the Quality of a Source

Criteria	Associated questions	Score (0/1)
Conceptual framing	Does the source acknowledge existing research?	
	Does the source construct a conceptual framework?	
	Does the source pose a research question or outline a hypothesis?	
Transparency	Does the source present or describe the raw data it analyzes?	
	Does the source specify the geography/context?	
	Does the source declare sources of support/funding?	
Appropriateness	Does the source identify a research design?	
	Does the source identify a research method?	
	Does the source demonstrate why the chosen design and method are well suited to the research question?	
Cultural sensitivity	Does the source explicitly consider any context-specific cultural factors that may bias the analysis/findings?	

Validity	Does the source demonstrate measurement validity?	
	Does the source consider/show the internal validity?	
	Does the source consider/show the external validity?	
	Does the source consider/show the ecological validity?	
Reliability	Are the measures used in the source stable?	
	Are the measures used in the source internally reliable?	
	Are the findings likely to be sensitive/ changeable depending on the analytical technique used?	
Cogency	Does the author 'signpost' the reader throughout?	
	Does the author consider the source's limitations and/or alternative interpretations of the analysis?	
	Are the conclusions clearly based on the source's results?	
	TOTAL	**__/20**

Note. Adapted from DFID (2014). 1 – Meet the criterion; 0 – Fail to meet the criterion.

Appendix 3: Selected Sources

No.	Source	Full reference	Context	Quality
1	Abidin et al. (2019)	Abidin, A. Z., Istiyono, E., Fadilah, N., & Dwandaru, W. S. B. (2019). A computerized adaptive test for measuring the physics critical thinking skills. *International Journal of Evaluation and Research in Education*, *8*(3), 376–383.	Indonesia	Moderate
2	Aristizábal (2018)	Aristizábal, J. (2018). Using learning analytics to improve students' reading skills: A case study in an American international school with English as an additional language (EAL) students. *GIST Education and Learning Research Journal*, *17*, 193–214.	Vietnam	Moderate
3	Aslam et al. (2020)	Aslam, R., Khan, N., & Ahmed, U. (2020). Technology integration and teachers' professional knowledge with reference to International Society for Technology in Education (ISTE)-Standard: A causal study. *Journal of Education and Educational Development*, *7*(2), 307–327.	Pakistan	Moderate
4	Auphan et al. (2020)	Auphan, P., Ecalle, J., & Magnan, A. (2020). The high potential of computer-based reading assessment. *Canadian Journal of Learning and Technology*, *46*(1), 1–23.	France	Moderate
5	Barana et al. (2021)	Barana, A., Marchisio, M., & Sacchet, M. (2021). Interactive feedback for learning mathematics in a digital learning environment. *Education Sciences*, *11*, 279–299.	Italy	High
6	Benton (2021)	Benton, T. (2021). Item response theory, computer adaptive testing and the risk of self-deception. *Research Matters*, *32*, 82–100.	United Kingdom	Moderate
7	Blumenthal & Blumenthal (2020)	Blumenthal, S., & Blumenthal, Y. (2020). Tablet or paper and pen? Examining mode effects on German elementary school students' computation skills with	Germany	Moderate

		curriculum-based measurements. *International Journal of Educational Methodology*, *6*(4), 669–680.		
8	Chan (2021)	Chan, K. T. (2021). Embedding formative assessment in blended learning environment: The case of secondary Chinese language teaching in Singapore. *Education Sciences*, *11*, 360–371.	Singapore	Low
9	Çimen (2022)	Çimen, S. S. (2022). Exploring EFL assessment in Turkey: Curriculum and teacher practices. *International Online Journal of Education and Teaching*, *9*(1), 531–550.	Turkey	Low
10	Dwiyono et al. (2021)	Dwiyono, Y., Mulawarman, W. G., Pramono, P. O., Salim, N. A., & Ikhsan, M. (2021). Implementation of national examination based on computer based test at Vocational School 1 North Sangatta. *Cypriot Journal of Educational Sciences*, *16*(1), 86–95.	Indonesia	Moderate
11	Ercikan et al. (2018)	Ercikan, K., Asil, M., & Grover, R. (2018). Digital divide: A critical context for digitally based assessments. *Education Policy Analysis Archives*, *26*(51), 1–18.	Argentina, Australia, Canada, Chile, Croatia, Czech Republic, Denmark, Germany, China, South Korea, Lithuania, Netherlands, Norway, Poland, Russia, Slovakia, Slovenia, Switzerland, Thailand, Turkey	Moderate
12	Ferretti et al. (2021)	Ferretti, F., Santi, G. R. P., Del Zozzo, A., Garzetti, M., & Bolondi, G. (2021). Assessment practices and beliefs: Teachers' perspectives on assessment during long distance learning. *Education Sciences*, *11*, 264–280.	Italy	Moderate
13	Herwin et al. (2021)	Herwin, H., Hastomo, A., Saptono, B., Ardiansyah, A. R., & Wibowo, S. E. (2021). How elementary school teachers organized online learning during the COVID-19 pandemic? *World Journal on Educational Technology: Current Issues*, *13*(3), 437–449.	Indonesia	Moderate
14	Ibrohim et al. (2021)	Ibrohim, Sudrajat, A. K., & Saefi, M. (2021). Assessing Indonesian teachers' perspective on the implementation of distance learning due to COVID-19 based on online survey. *Journal of Turkish Science Education*, *18*, 46–59.	Indonesia	Moderate
15	Istiningsih (2022)	Istiningsih, I. (2022). Impact of ICT integration on the development of vocational high school teacher TPACK in the Digital Age 4.0. *World Journal on Educational Technology: Current Issues*, *14*(1), 103–116.	India	Moderate
16	Jaiswal & Arun (2021)	Jaiswal, A., & Arun, C. J. (2021). Potential of artificial intelligence for transformation of the education system in India. *International Journal of Education and Development Using Information and Communication Technology*, *17*(1), 142–158.	Malaysia	Moderate
17	Jong & Tan (2021)	Jong, B., & Tan, K. H. (2021). Using Padlet as a technological tool for assessment of students' writing skills in online classroom settings. *International Journal of Education and Practice*, *9*(2), 411–423.	Malaysia	Moderate
18	Kusuma et al. (2021)	Kusuma, I. P. I., Mahayanti, N. W. S., Adnyani, L. D. S., & Budiarta, L. G. R. (2021). Incorporating e-portfolio with flipped classrooms: An in-depth analysis of students' speaking performance and learning engagement. *The JALT CALL Journal*, *17*(2), 93–111.	China	Moderate
19	Liu (2022)	Liu, K. (2022). Success plan for the remote assessments: Lessons learnt in COVID-19 outbreak. *International Journal*	Philippines	Low

		of Multidisciplinary and Current Educational Research, *4*(1), 37–42.		
20	Lubiano (2018)	Lubiano, M. L. D. (2018). Interactive e-learning portal for enrichment of conceptual understanding of Grade 8 learners in Physics. *Tilamsik: The Southern Luzon Journal of Arts and Sciences*, *10*, 37–50.	New Zealand	High
21	O'Keefe et al. (2020)	O'Keefe, L., Dellinger, J. T., Scragg, B., Amelina, N., & Mathes, J. (2020). The state of online learning in the Kingdom of Saudi Arabia: A COVID-19 impact study for K-12. In *Online Learning Consortium*. Online Learning Consortium, Inc.	Turkey, Germany, Spain	High
22	Özdil et al. (2021)	Özdil, G., Simsek, F. M., & Tugluk, M. N. (2021). Comparison of digital assessment and documentation systems used in the early childhood education in Turkey, Germany and Spain. *Southeast Asia Early Childhood*, *10*(2), 1–15.	Indonesia	High
23	Romero-Tena et al. (2020)	Romero-Tena, R., Lopez-Lozano, L., & Gutierrez, M. P. (2020). Types of use of technologies by Spanish early childhood teachers. *European Journal of Educational Research*, *9*(2), 511–522.	Philippines	Moderate
24	Samosa (2022)	Samosa, R. C. (2022). Reframing remote learning assessment practices of teachers': Input for school based testing reforms. *International Journal of Academic Pedagogical Research*, *6*(1), 4–20.	Malaysia	Moderate
25	Samsudin et al. (2020)	Samsudin, M. A., Chut, T. S., Ismail, M. E., & Ahmad, N. J. (2020). A calibrated item bank for computerized adaptive testing in measuring science TIMSS performance. *EURASIA Journal of Mathematics, Science and Technology Education*, *16*(7), 1–15.	Spain	High
26	Sánchez-Rivas et al. (2019)	Sánchez-Rivas, E., Ruiz-Palmero, J., & Sánchez-Rodríguez, J. (2019). Gamification of assessments in the natural sciences subject in primary education. *Educational Sciences: Theory and Practice*, *19*(1), 95–111.	Spain, Turkey	Moderate
27	Simsek & Tugluk (2021)	Simsek, F. M., & Tugluk, M. N. (2021). Making learning visible in the 21st century; examining of the use of digital assessment tools in Montessori education. *Pegem Journal of Education and Instruction*, *11*(2), 72–86.	Malaysia	Moderate
28	Singh et al. (2022)	Singh, C. K. S., Muhammad, M. M., Mostafa, N. A., Yunus, M. M., Noordin, N., & Darm, R. (2022). Exploring ESL teachers' alternative assessment strategies and practices in the classroom. *Journal of Language and Linguistic Studies*, *18*(1), 411–426.	Cyprus	Moderate
29	Stylianidou et al. (2020)	Stylianidou, N., Sofianidis, A., Manoli, E., & Meletiou-Mavrotheris, M. (2020). "Helping Nemo!"—Using augmented reality and alternate reality games in the context of universal design for learning. *Education Sciences*, *10*, 95–118.	Indonesia	Moderate
30	Tas et al. (2021)	Tas, Y., Eminoglu, S., Atila, G., Yildiz, Y., & Bozkurt, U. (2021). Teachers' self-efficacy beliefs and opinions about distance education during the COVID-19 pandemic. *Turkish Online Journal of Distance Education*, *22*(4), 229–253.	Turkey	Moderate
31	Tashbolatovna et al. (2022)	Tashbolatovna, M. K., Duisenbaevna, U. Z., Kaldyhanovna, K. R., Koyshibaevna, A. G., Ulmeken, Z., & Gulzat, S. (2022). Development of the evaluative activities of teachers in the conditions of updated education. *Cypriot Journal of Educational Sciences*, *17*(4), 1304–1319.	Kazakhstan	Moderate

32	Turan & Meral (2018)	Turan, Z., & Meral, E. (2018). Game-based versus to non-game-based: The impact of student response systems on students' achievements, engagements and test anxieties. *Informatics in Education*, *17*(1), 105–116.	Turkey	Moderate
33	Yilmaz (2021)	Yilmaz, F. (2021). Comparing paper-pencil and computer-based tests: A meta-analysis study in the sample of Turkey. *Eurasian Journal of Educational Research*, *93*, 279–300.	New Zealand	High
34	Yunus & Hua (2021)	Yunus, C. C., & Hua, T. K. (2021). Exploring a gamified learning tool in the ESL classroom: The case of Quizizz. Journal of Education and E-Learning Research, 8(1), 103–108.	Malaysia	Moderate

unesco

Global Education
Monitoring Report