Beating a dead horse: investigating the continued use of ineffective assessment practices in modern education systems

Mohammad Awad AlAfnan[1], Siti Fatimah MohdZuki[2], Shefa Mohammad AlAfnan[3]
[1]Department of Liberal Arts, American University of the Middle East, Eqilah, Kuwait
[2]University Technology MARA, Shah Alam, Malaysia
[3]International Islamic University Malaysia, Kuala Lumpur, Malaysia

ABSTRACT

The study examines the persistence of outdated assessment practices in the education systems of Jordan, Malaysia, and Australia, utilizing the Dead Horse Theory as an interpretive lens. Despite ongoing educational reforms advocating for authentic, student-centered assessment models, traditional high-stakes and summative evaluations continue to dominate. The research adopts a qualitative multi-site case study approach, drawing data from semi-structured interviews, document analysis, and classroom observations. The 30 participants, including teachers, administrators, faculty, and policy analysts, were purposively selected across the three countries. Findings reveal three interconnected themes contributing to assessment stagnation: institutional inertia, cultural attachment to standardized metrics, and inadequate professional development. In Australia, systemic accountability pressures, such as National Assessment Program–Literacy and Numeracy (NAPLAN) and Australian Tertiary Admission Rank (ATAR) rankings, limit the innovation in assessment. In Jordan, the Jordanian High School General Examination (*Tawjihi*) exam significantly influences pedagogy and curriculum. In contrast, a dual reality prevails in Malaysia, with reform attempts often overshadowed by entrenched examination cultures. Teachers across all contexts expressed frustration with superficial policy changes and limited practical support, frequently engaging in performative compliance rather than genuine transformation. Applying the Dead Horse Theory highlights how education systems adopt symbolic or cosmetic reforms, 'buying a stronger whip' or 'changing the rider', instead of abandoning dysfunctional practices. The study concludes that meaningful reform requires structural and policy change, cultural shifts, and sustained teacher empowerment. The Dead Horse Theory thus serves as a powerful metaphor and diagnostic tool for identifying the illusion of progress and advocating for truly transformative educational assessment.

Corresponding Author:

Mohammad Awad AlAfnan
Department of Liberal Arts, American University of the Middle East
Block 6, Building 1, Egaila, Kuwait
Email: Mohammad.al-afnan@aum.edu.kw

1. INTRODUCTION

In the ever-evolving landscape of education, where innovation and evidence-based practices are touted as essential to progress, it is paradoxical that many educational institutions continue to rely on outdated and ineffective assessment methods [1]. Despite significant advancements in pedagogical research, many modern education systems continue to rely on traditional assessment tools, such as standardized tests,

multiple-choice exams, and rigid grading schemes. This phenomenon is aptly captured by the 'Dead Horse Theory', a satirical metaphor suggesting that when one realizes the horse is dead, the logical action would be to dismount rather than keep beating it [2]. However, in education, the proverbial dead horse is beaten, dressed in new harnesses, renamed, and rebranded, while the foundational problems remain unaddressed.

The 'Dead Horse Theory', often used humorously in organizational and institutional critique, illustrates the futility and absurdity of persisting in failed efforts [3]. Within educational contexts, it highlights how systemic inertia, resistance to change, and institutional tradition often outweigh reason and evidence. This theory underscores a critical and persistent issue in contemporary education: the unwavering reliance on assessment methods that no longer effectively serve students or educators. Many practices hinder learning, marginalize diverse learners, and promote a culture of high-stakes anxiety and superficial achievement.

Assessment, by definition, should be a tool to support learning, provide feedback, and inform instruction [4]. It should offer insights into student understanding, skill development, and educational progress. However, in many schools and universities worldwide, assessment has become an instrument of judgment rather than a tool for growth. The emphasis on scores, ranks, and test performance has overshadowed the need for meaningful feedback, critical thinking, and creativity [5]. In such systems, teaching is often reduced to 'teaching to the test', and students are trained to perform rather than to learn. This shift from learning-oriented assessment to performance-based evaluation reflects a fundamental misalignment between educational goals and assessment practices [6].

The continued use of ineffective assessments is not merely a pedagogical issue but a cultural and structural problem. Standardized testing, for example, was initially introduced to ensure equity and accountability. However, decades of use have revealed its limitations. These tests often fail to capture complex cognitive abilities, ignore socio-economic disparities, and disproportionately disadvantage students from marginalized communities [7]; nevertheless, they persist. Policymakers, educators, and institutions often justify their use with appeals to tradition, simplicity, and scalability, while ignoring the wealth of research highlighting their shortcomings.

Like multiple-choice tests and traditional exams, these assessment practices continue to dominate despite criticism for promoting surface-level learning [8]. These formats encourage students to memorize facts and select the correct option rather than engage in higher-order thinking or problem-solving. Although easier to administer and grade, they do little to reflect a student's accurate understanding or potential [9]. However, they remain the default method in many classrooms, especially in high-stakes environments such as national exams and university entrance tests.

Several factors contribute to the persistence of these outdated practices. First, there is institutional resistance to change. Educational systems are complex and slow-moving, often governed by bureaucracies that prioritize stability over innovation. Changing assessment practices requires systemic shifts in curriculum design, teacher training, policy formulation, and resource allocation—an undertaking many systems are ill-equipped or unwilling to make [10].

Second, there is a deep-rooted culture of tradition in education [11]. Many educators, having gone through traditional assessment systems, are inclined to replicate familiar methods. This cultural continuity, while comforting, can be detrimental when it perpetuates ineffective practices. Additionally, the fear of the unknown, uncertainty about new assessment methods, the potential for failure, or a loss of control discourages experimentation and adaptation. Third, accountability pressures play a significant role. Governments and educational authorities rely on standardized test scores to evaluate schools, allocate funding, and determine teacher effectiveness [12]. This creates a high-stakes environment where test performance becomes the primary indicator of success, prompting schools to adopt practices that 'work' in a narrowly defined sense, even if they are pedagogically unsound.

There is no professional development and support for alternative assessment practices [13]. While many educators recognize the limitations of traditional assessments, they may lack the training, tools, or institutional support to implement more effective alternatives, such as project-based learning, formative assessments, peer evaluations, and reflective portfolios. Teachers may feel unequipped to move away from the familiar without adequate professional learning opportunities.

Despite these challenges, a growing body of research and practical evidence supports the need for change. Educators, researchers, and reformers worldwide are advocating for assessment methods that are holistic, inclusive, and aligned with 21st-century learning objectives [14]–[16]. These include formative assessments that guide instruction, performance-based assessments that simulate real-world tasks, and self-assessments that foster metacognition and learner autonomy. Such methods provide a more accurate picture of student learning and empower students to take ownership of their education.

Furthermore, educational technologies are offering new possibilities for assessment [17]. Digital tools can facilitate personalized assessments, track learning progress in real time, and provide immediate feedback. Adaptive learning platforms, online simulations, and e-portfolios are being used innovatively to

assess a broader range of competencies. However, integrating these technologies effectively requires careful planning, pedagogical insight, and systemic support.

This article critically examines why ineffective assessment practices persist in modern education systems, despite widespread recognition of their limitations. Drawing on the Dead Horse Theory as a conceptual lens, it explores the cultural, institutional, and structural barriers that prevent meaningful change. It also highlights successful examples of innovative assessment practices and offers practical recommendations for moving beyond the "dead horse" of traditional evaluation. Ultimately, this inquiry seeks to contribute to a more reflective and progressive discourse on assessment in education, prioritizing learning over legacy and impact over inertia.

2.   LITERATURE REVIEW

Assessment has long been the backbone of educational systems, guiding instruction, informing policy, and evaluating student progress. Historically, standardized testing was introduced in the early 20th century to promote objectivity and comparability in measuring student achievement [18], [19]. Originally intended to close learning gaps and support equity, these assessments have since evolved into high-stakes tools that significantly influence curriculum, instruction, and educational policy [20], [21]. Over time, critics have argued that these tests have deviated from their intended purpose, often reducing learning to a narrow range of skills and fostering a culture of compliance and performance rather than genuine understanding [22], [23].

A key critique of standardized testing lies in its limited ability to assess a broad range of cognitive and affective abilities. Numerous studies have documented that standardized tests primarily measure lower-order thinking skills, such as recall and recognition, and are ill-suited for evaluating critical thinking, creativity, and collaborative competencies [24], [25]. Moreover, these assessments often contain cultural and linguistic biases that disadvantage students from minority, low-income, or multilingual backgrounds [26]. The resulting achievement gaps are frequently misinterpreted as deficiencies in student ability rather than systemic inequities inherent in assessment design [27], [28].

Equally concerning is the effect of traditional assessments on teaching practices. The "teaching to the test" phenomenon has become a widespread consequence of high-stakes testing regimes, particularly in countries such as the United States, the United Kingdom, and China [29], [30]. Under pressure to boost test scores, teachers often narrow their instruction to focus solely on tested content and formats, marginalizing non-tested subjects such as the arts, social studies, and physical education [31]. This environment limits students' exposure to a well-rounded education and contributes to teacher burnout and professional dissatisfaction [32].

The implications of these practices extend to issues of equity and access. High-stakes assessments can deepen disparities by disproportionately affecting students from under-resourced schools who may lack access to preparatory materials and support systems [33]. These assessments can also produce adverse psychological effects, including test anxiety, lowered self-esteem, and disengagement from school, particularly among students labelled as 'low-performing' [34]. Using test scores in school funding and teacher evaluations can encourage unethical practices such as data manipulation or excluding low-performing students from testing altogether [35].

In contrast to these traditional approaches, alternative assessment methods have gained increasing support in recent years. Formative assessments involving ongoing feedback to inform instruction and learning have significantly improved student achievement and motivation [36]. Rather than judging student performance with a final score, formative assessment emphasizes the process of learning and supports a growth mindset [37]. Performance-based assessments, such as projects, portfolios, and presentations, offer a more authentic evaluation of student abilities by requiring them to apply knowledge in real-world contexts [38]. These assessments encourage higher-order thinking, creativity, and interdisciplinary learning skills increasingly demanded in the 21st-century workforce. Moreover, culturally responsive assessment practices have become essential for promoting educational equity. These practices consider students' cultural backgrounds, values, and experiences in designing and interpreting assessment tasks [39]. Research suggests that culturally relevant assessments can increase student engagement, validate identity, and improve academic outcomes, particularly for historically marginalized populations [40]–[42].

Despite growing evidence supporting alternative assessment methods, several barriers hinder widespread adoption. One major obstacle is the policy landscape, which often mandates standardized testing as the primary mechanism of accountability [43]. Many teachers lack the autonomy, time, or training to implement more student-centered assessments [44]. Institutional inertia and fear of the unknown further contribute to the persistence of outdated practices, even when they no longer serve pedagogical purposes [45].

Several theoretical frameworks support the shift toward more effective and inclusive assessment models. Constructivist theory, for instance, emphasizes learners' active engagement in constructing knowledge and supports formative, process-oriented assessments that reflect individual learning paths [46]. Gardner's theory of multiple intelligences also challenges the narrow scope of standardized assessments, arguing for a broader evaluation of human capabilities [47]. Similarly, self-determination theory highlights the role of autonomy, competence, and relatedness in motivation, suggesting that assessments should empower students and support their psychological needs [48].

Empirical research supports the efficacy of these alternative approaches. For example, schools integrating portfolio assessments have reported increased student agency, deeper reflection, and improved learning outcomes [49]. Likewise, formative assessment strategies have demonstrated consistent gains across grade levels and subject areas [50]. When implemented effectively, culturally responsive assessments have been linked to enhanced classroom participation and academic achievement, particularly among Indigenous and multilingual students [51].

Researchers and educators advocate for systemic reforms to move beyond ineffective assessment models [52]. These include policy changes that support diverse forms of assessment, investment in teacher professional development, and greater collaboration among stakeholders to design assessments that reflect the needs of 21st-century learners [53]. As long as educational systems continue to "beat the dead horse" of traditional, ineffective assessment, meaningful learning will remain constrained by outdated measures of success.

3.   METHOD

This study employs a qualitative, multi-site case study approach to investigate the persistent use of outdated assessment practices in educational institutions. The investigation is framed through the lens of the Dead Horse Theory, which serves as a metaphor for how systems continue ineffective behaviors despite evidence of their failure. The study focuses on three national contexts: Australia, Jordan, and Malaysia, offering insights from different educational systems—Western, Middle Eastern, and Southeast Asian.

3.1. Research design

A qualitative case study design was carefully chosen. This was to provide a comprehensive and in-depth understanding of how institutional cultures and systemic norms contribute to the persistence of outdated assessment practices within educational systems. This design is particularly beneficial as it facilitates a nuanced exploration across three distinct educational contexts, thereby capturing both common patterns that may emerge and country-specific dynamics that illustrate the unique challenges and opportunities present in each locale.

To ensure a rich and diverse understanding of the multifaceted issues at play within the education sector, participants for the study were selected using a purposive sampling method. This strategic approach was specifically employed to guarantee that various voices and perspectives were represented, encompassing key stakeholders who play vital roles in education. Consequently, a diverse group of educators, school administrators, and policy actors, all directly involved in the design, implementation, or reform of educational assessments, was carefully invited to participate in the study, as shown in Table 1. This deliberate and thoughtful selection process is instrumental in enriching the data collected, resulting in more comprehensive and accurate overall study findings. By including such a diverse group of participants, the research can capture broader insights regarding the factors influencing assessment practices. This depth of understanding is critical, as it enables the study to reflect the complexities of the educational landscape and the diverse experiences and viewpoints of those directly engaged in the assessment process. The findings from this research can provide valuable implications for future policies and practices within the education system, ultimately contributing to the continuous improvement of assessment methods and their alignment with educational goals.

Table 1. Participant demographics by country

Nationality	Teachers	Administrators	Faculty	Policy analysts	Total
Australia	5	2	2	1	10
Jordan	6	2	1	1	10
Malaysia	4	1	2	3	10
Total	15	5	5	5	30

The participants involved in the study were carefully selected to encompass a diverse range of public and private institutions. This selection process involved individuals from various educational settings, including primary schools, secondary schools, and numerous entities within the higher education sector, such as community colleges, universities, and vocational training centers. By incorporating various educational institutions, the study aimed to capture a comprehensive representation of perspectives, opinions, and experiences, enriching the research findings and providing valuable insights into the academic landscape. This diversity among participants enhances the research's validity and ensures that a wide range of voices are heard and represented in the study's outcomes.

3.2. Data collection methods

In the research process, three distinct qualitative data collection strategies were strategically employed to enhance data triangulation and the findings' reliability. These strategies included in-depth interviews, focus group discussions, and participant observations, each chosen for its unique strengths in extracting detailed and contextual insights from participants, as shown in Table 2. By employing these diverse approaches, the study aimed to gather comprehensive insights and corroborate the findings, ensuring a more nuanced understanding of the subject matter under investigation. The incorporation of varied qualitative methods not only strengthens the validity of the research but also provides a richer context for analysis and interpretation. This multifaceted approach is particularly beneficial in qualitative research, as it enables the exploration of complex social phenomena and accommodates participants' diverse perspectives. Additionally, through triangulation, the researchers could assess the consistency of findings across different methods, further reinforcing the trustworthiness of the data collected and the conclusions drawn. Thus, the study contributes valuable knowledge to the field and exemplifies best practices in qualitative research design, demonstrating how the strategic selection of diverse data collection techniques can yield a more robust and comprehensive understanding of the research topic.

In-person data collection efforts were diligently carried out in two notably diverse countries, Jordan and Australia. This strategic choice allowed researchers to gather firsthand insights and personal experiences from participants in these unique locations. By engaging directly with individuals in their environments, the researchers could foster a deeper understanding of the cultural and contextual factors that influence the participants' responses. In Malaysia, a more innovative hybrid approach was implemented, skillfully combining online and in-person data collection methods. This dual strategy not only facilitated a broader reach but also provided enhanced flexibility in data collection. It ensured that participants could engage in the study through their preferred means of interaction, whether they chose to participate face-to-face or digitally. This inclusivity was particularly beneficial in accommodating varying comfort levels and accessibility needs, ultimately enriching the data gathered by capturing a wider range of perspectives and experiences.

Table 2. Data collection methods

Method	Purpose
Semi-structured interviews	Explore beliefs, attitudes, and narratives related to assessment.
Document analysis	Review national education policies, curriculum frameworks, and reports.
Classroom observations	Observe the implementation of assessment strategies in real-time contexts.

3.3. Data analysis

Thematic analysis, a method for analyzing qualitative data, was conducted following the comprehensive six-phase model proposed by Braun *et al*. [54]. This model provides a structured approach to ensure a thorough analysis and interpretation of the qualitative data collected throughout the study. The phases involved in the thematic analysis are as:

i)   Familiarization with the data: the first step involves immersing oneself in the collected data to gain an in-depth understanding and insight into its diverse content. Researchers often read, reread, and note any initial impressions during this phase, ensuring they are well-acquainted with the information before proceeding to the next stage. This step not only aids in comprehending the breadth of the data but also helps researchers identify potential areas of interest or concern that may warrant further exploration.

ii) Generating initial codes: in this phase, researchers identify and label significant data that seem relevant to the research questions. This coding process is crucial as it organizes the data into meaningful segments, allowing researchers to highlight essential elements and themes. Researchers typically use inductive and deductive approaches to create a comprehensive set of codes aligning with the research objectives. The initial codes can be the foundational building blocks for the subsequent theme generation process.

iii) Searching for themes: the next step is to identify broader patterns or themes within the generated codes after coding the data. Researchers assess how different codes may combine to form overarching themes and begin to see connections between various aspects of the data. During this phase, it may be beneficial for the research team to engage in discussions to collectively analyze the connections and implications of the codes, facilitating a richer understanding of the emerging themes.

iv) Reviewing themes: once potential themes are identified, they are reviewed against the dataset to ensure they accurately represent the data. This phase may involve refining the themes based on additional insights gained from further examination of the data. It requires critical evaluation to determine whether the themes effectively convey the essence of the data, ensuring clarity and coherence in their portrayal.

v) Defining and naming themes: in this phase, researchers articulate what each theme represents about the research question, providing a detailed definition for each theme that encapsulates its significance and relevance. This step also involves naming the themes to reflect their core message, making them accessible and understandable for audiences and stakeholders.

vi) Producing the final report: the final stage involves compiling a detailed report that conveys the findings of the thematic analysis, illustrating the key themes with supporting evidence from the data. This report is a communication tool synthesizing the analysis process and conclusions, providing comprehensive insights supported by data-driven evidence.

NVivo software was employed as a critical component in the research process to facilitate the coding and organization of qualitative data. This innovative tool significantly enhances the efficiency of the analysis process, enabling researchers to systematically approach significant volumes of data with ease and precision. By enabling intuitive management and quick retrieval of information, NVivo streamlines the workflow and supports the rigorous nature of thematic analysis. Researchers can categorize and synthesize diverse data types, fostering a more comprehensive understanding of complex themes and patterns that emerge from the research, ultimately leading to more robust findings and insights.

Thematic analysis was first applied within each country to identify unique themes. This initial analysis phase allowed researchers to develop a rich understanding of the individual cultural contexts and variations present within each dataset. Subsequently, a comparative study across countries was conducted to uncover shared themes and differing perspectives. This two-step approach highlights the individual cultural contexts. It emphasizes the broader patterns that emerge during cross-country comparisons, offering valuable insights into the research questions and contributing to a more nuanced understanding of the overall findings.

3.4. Emergent themes
3.4.1. Policy rigidity and administrative overload

Respondents from Australia highlighted the challenges of standardized testing mandates, notably the National Assessment Program–Literacy and Numeracy (NAPLAN). This system often imposes rigid structures on educational practices, leading to concerns about the administrative burden on teachers and schools. In contrast, educators from Jordan and Malaysia discussed the implications of centralized examination systems, which are often accompanied by top-down directives that limit flexibility in administering and developing assessments. This highlights a broader educational policy issue that can stifle innovation and responsiveness to local context.

3.4.2. Cultural perceptions of assessment

The cultural context in which assessments are viewed significantly influences educational priorities in various countries. In Australia, schools are under considerable community pressure to improve their rankings, which can lead to a narrow focus on high-stakes testing results. Meanwhile, in Jordan and Malaysia, the societal perception of education is intricately linked to the status conferred by grades. In these countries, academic achievement is often regarded as a measure of personal and family prestige, which can result in an intense focus on examinations and outcomes.

3.4.3. Professional development gaps

Across all three countries, educators expressed frustration regarding the limited scope of professional development. Many teachers reported feeling ill-equipped to design and implement alternative assessments that better serve students' diverse needs and learning styles. This gap in training highlights the need for more comprehensive professional development opportunities that focus on innovative assessment strategies and align with current educational trends and demands.

3.5. Use of Dead Horse Theory

The Dead Horse Theory provides a compelling framework for analyzing certain institutional behaviors and practices. It categorizes these behaviors into three specific approaches:

− Changing the rider: this approach involves altering the leadership within an institution while maintaining the underlying structure. This can often lead to a cycle of old problems, as the new leadership may not address the root causes of inefficiencies.
− Beating harder: this behavior indicates a tendency to persist with outdated methods, even when clear evidence demonstrates that these strategies are ineffective. Despite mounting evidence, institutions may resist change, believing that increased effort or persistence will yield better outcomes.
− Buying a stronger whip: the institution invests in high-stakes assessments or tools to improve performance without addressing the core issues that hinder success. This notion reflects a misconception that throwing resources at the problem can rectify fundamental flaws in processes or systems.

The Dead Horse Theory, ultimately, serves as a critical reminder of the need to recognize and actively reassess long-standing behaviors and practices within various institutions. This theory highlights how deeply ingrained habits can hinder progress and adaptation in a rapidly changing world. By fostering an environment that encourages reflection on these entrenched behaviors, institutions can pave the way for meaningful reform and transformation. The process involves questioning traditional methods, considering alternative approaches, and being open to change to ensure that organizational practices remain relevant and effective in achieving their goals.

3.6. Ethical considerations

Ethical approval was obtained thoroughly from each participating country's relevant university research boards. This process ensures that all research adheres to established ethical guidelines and standards aimed at protecting the rights and welfare of the study participants. Before their involvement in the study, participants were comprehensively informed about several key aspects:
− The purpose and scope of the study, which detailed the research objectives, methodology, and anticipated outcomes.
− Their rights as participants, including the assurance of voluntary participation, the right to withdraw from the study at any point without any implications or penalties.
− Confidentiality and data security measures were explicitly explained, assuring participants that their personal information would be treated with the utmost care, safeguarded against unauthorized access, and used solely for research purposes.

Additionally, it is essential to note that written informed consent was obtained from all participants involved in the study. This process demonstrates their understanding of the research and signifies their voluntary agreement to participate. By obtaining this consent, the researchers reinforce their ethical commitment to upholding transparency and respect for all study participants. The informed consent process is crucial for several reasons: it ensures compliance with legal and regulatory requirements governing research involving human subjects. It is vital in fostering trust and transparency between researchers and participants. Trust is essential in research settings, as it encourages open communication and cooperation, ultimately leading to more accurate and reliable data collection. Moreover, transparency regarding the study's objectives, procedures, potential risks, and benefits empowers participants, enabling them to make informed decisions about their involvement in the research.

3.7. Ensuring trustworthiness

Table 3 provides the strategies that were implemented to ensure research quality and credibility. Despite the political rhetoric that often advocates for educational reform, a comprehensive and robust methodological framework has substantially contributed to a detailed and nuanced exploration of the various ways outdated assessment practices continue to be upheld across three distinct countries: Australia, Jordan, and Malaysia. This framework facilitates a thorough examination of the educational systems in these nations, enabling a deeper understanding of the structural and cultural factors at play.

Table 3. Trustworthiness strategies

Criterion	Strategy
Credibility	Triangulation of data sources and methods
Transferability	Thick description of contexts and practices
Dependability	Audit trails and documentation
Confirmability	Researcher reflexivity and peer checks

Through this comparative analysis, we gain invaluable insights into the persistence of these antiquated practices, shedding light on the historical contexts and systemic challenges that contribute to their ongoing maintenance within these educational systems. It highlights the similarities and differences in how

these practices are implemented and the socio-political influences that shape educational policies and reforms in each country. This exploration can inform future discussions and strategic actions to foster meaningful educational change and improvement across diverse contexts.

4. RESULTS

This section provides an in-depth exploration of the findings derived from a comprehensive comparative analysis of educational assessment practices in three distinct countries: Australia, Jordan, and Malaysia. This analysis is fundamentally informed by the Dead Horse Theory, which serves as a conceptual framework for understanding the nuances and implications of these practices. Throughout the examination, three significant themes surfaced consistently across all analyzed contexts. The first theme, institutional inertia, refers to the resistance to change within educational institutions, impacting their ability to adapt to new assessment methods. The second theme highlights the cultural attachment to metrics, illustrating how various societies prioritize quantitative measures of educational success, often at the expense of more holistic approaches. Finally, the third theme, professional capacity gaps, points to the discrepancies in skills and knowledge among educational professionals, which can hinder the effective implementation of assessment practices. These themes are thoroughly supported by detailed analysis and supplemented with data tables that illustrate the findings, thereby providing a clear and comprehensive understanding of the comparative assessment practices in these diverse educational landscapes.

4.1. Institutional inertia and bureaucratic constraints

Educators across all three countries- Australia, Jordan, and Malaysia- reported experiencing persistent institutional constraints that significantly restrict their ability to innovate in the realm of assessment as shown in Table 4. Specifically, in Australia, participants highlighted the impact of national mandates, such as NAPLAN, which establishes standardized assessment measures that educators must adhere to. Meanwhile, educators in Jordan and Malaysia highlighted the significant influence of central examinations, such as the Jordanian High School General Examination (*Tawjihi*) in Jordan and the *Sijil Pelajaran Malaysia* (SPM) and *Pentaksiran Tingkatan 3* (PT3) in Malaysia. These examinations are crucial in shaping educational policies and practices, ultimately determining how assessments are designed and implemented within schools. The dominance of these standardized testing systems presents challenges for educators who seek to adopt more flexible and innovative assessment strategies that cater to the diverse needs of their students.

Table 4. Institutional features hindering assessment reform

Feature	Australia	Jordan	Malaysia
Centralized assessments	NAPLAN, ATAR	*Tawjihi* (national exam)	SPM, PT3
Curriculum flexibility	Moderate (state-dependent)	Very low	Moderate in the private sector
Teacher autonomy	Limited to the public sector	Extremely limited	Variable
Reform implementation	Partial, fragmented	Top-down, exam-driven	Slow and uneven

In Australia, many respondents have underscored a primary concern within the educational framework that has been attracting increasing attention. Despite progressive policies to enhance student learning experiences and broaden educational opportunities, a relentless emphasis on standardized performance metrics persists. This situation is primarily fueled by government reporting systems, which compel schools to prioritize these metrics above other potentially more enriching educational approaches that could better serve student development. A secondary school principal clearly expressed this concern, stating, "*NAPLAN data shapes school reputation. Even when we try formative methods, the pressure to lift scores pulls us back.*" This poignant quote emphasizes the pressure educators face and highlights the tangible and often detrimental impact of standardized testing on school identity. It reflects how the relentless focus on performance metrics can create an environment that hinders innovation in teaching and impacts the overall culture within educational institutions. Furthermore, it illustrates how overemphasizing these standardized measures can stifle pedagogical innovation and inhibit the adoption of a more holistic and student-centered approach to education, which values diverse learning methodologies and individual student needs above mere numerical performance indicators. Schools, caught in this cycle, may struggle to implement practices that foster more profound understanding and critical thinking among students, resulting in a potential disconnect between the prescribed curriculum and the actual learning experiences in the classroom.

In Jordan, the challenges faced within the educational system are particularly pronounced, echoing sentiments expressed by educators regarding a rigid and often inflexible framework. These educators have shared their experiences, shedding light on how the existing structure stifles innovation and adaptability in teaching. A central component of this system is the *Tawjihi* exam. This standardized test plays a crucial role in

shaping the curriculum and the methods employed in classrooms across the nation. The *Tawjihi* exam has far-reaching consequences, as it dictates the subjects that must be taught and the specific ways in which they are delivered to students. Consequently, this creates a uniform approach to education that fails to cater to different students' diverse needs and learning styles. This one-size-fits-all assessment model is detrimental on multiple levels. It curtails the creative potential of educators, who often feel constrained by the need to focus on test preparation rather than engaging with their students through innovative and stimulating teaching practices.

Furthermore, the pressure on students is immense; they find themselves in an educational environment that prioritizes rote memorization above all else. Such an approach inadequately prepares them for real-world challenges, as it undervalues critical thinking, analytical skills, and a deep, meaningful understanding of the subject matter. As a result, students are frequently left navigating a system that rewards surface-level knowledge rather than inspiring them to explore ideas, ask questions, and develop a passion for learning that extends beyond the classroom. In essence, the combination of a rigid educational framework and an overwhelming focus on standardized testing fosters an atmosphere that is counterproductive to promoting genuine intellectual curiosity and growth, with significant implications for the future of education in Jordan.

In Malaysia, a noticeable distinction exists between public and private educational institutions, particularly regarding the flexible teaching approaches employed within these settings. Educators working in public schools often express concerns about the considerable pressure they face due to the rigid standardized expectations imposed upon them. This challenge resonates with the sentiments expressed by their counterparts in countries such as Australia and Jordan, where teachers similarly navigate the constraints of uniformity in their curricula and assessments. On the other hand, private schools in Malaysia present a contrasting scenario. These institutions tend to offer educators greater autonomy and flexibility, allowing them to employ a broader range of instructional strategies. This flexibility empowers teachers to tailor their approaches, thereby better accommodating their students' diverse learning needs and individual strengths. Consequently, educators in private settings can often create more personalized learning experiences that enhance student engagement and understanding more effectively.

The concept of Dead Horse Theory offers a pertinent framework through which to interpret these systemic challenges. This theory articulates the behaviors frequently seen in bureaucratic systems, such as metaphorically 'appointing a committee to study the dead horse'—illustrating the futile attempts to analyze or reform ineffective policies without addressing the core underlying issues. Furthermore, this theory evokes practices such as 'changing the rider', wherein administrators may merely shuffle responsibilities and roles without fundamentally challenging or altering the entrenched logic of assessment that dictates educational outcomes. This reality highlights a significant observation: while discussions about educational reform are prevalent, actual, meaningful change often remains elusive, as the fundamental issues inherent in assessment and accountability frameworks continue to be overlooked and unaddressed, thereby preventing the creation of an educational environment conducive to genuine learning and growth.

4.2. Cultural attachment to traditional metrics

In all three countries, cultural values surrounding grades and test scores perpetuate outdated assessment practices. This societal emphasis is a powerful barrier to reform, as shown in Table 5. An Australian teacher provided insight into the effects of standardized testing, specifically mentioning that, "*parents compare NAPLAN results across schools. That creates fear and competition, so teachers play it safe.*" This statement highlights the detrimental impact that parental pressure and comparisons can have on educators, leading them to refrain from innovative teaching methods in favor of ensuring high test scores. In Jordan, participants emphasized the significant weight of academic performance in their society, revealing that grades are intrinsically linked to family honor and future career prospects. One administrator said, "*a child with low Tawjihi scores is seen as a failure. It is not just academic; it is personal and social.*" This perspective highlights the intense societal pressure that students and their families face, where academic assessments can significantly impact one's social standing and sense of personal worth.

In Malaysia, the educational landscape presents a varied picture. Urban parents are increasingly advocating for a more holistic approach to education, recognizing the importance of developing the whole child rather than solely focusing on examination performance. However, despite this progressive view gaining traction, the traditional exam culture continues to dominate within public schools, reflecting a lingering preference for standardized assessment as the primary measure of academic success. Dead Horse Theory interpretation: cultural factors mirror 'buying a stronger whip' or 'doubling the feed'—exerting more effort into an ineffective system to maintain appearances.

Table 5. Cultural drivers of exam-centric practices

Factor	Australia	Jordan	Malaysia
Parental expectations	High (ranking and comparisons)	Extremely high (status, identity)	High (university and scholarship)
Community trust in exams	Strong (especially NAPLAN)	Deep-rooted	Strong in the public system
Student identity and pressure	Increasingly exam-driven	Dominated by final exam results	Heavily grade-focused

4.3. Lack of training and resources

A comprehensive analysis of the educational systems in the three countries revealed a notable and concerning disparity in the preparation of teachers, particularly in response to the demands of contemporary assessment methodologies, as shown in Table 6. Participants in the study, encompassing a diverse range of educators and stakeholders, articulated their varied experiences regarding the training they received. Many indicated that the training was predominantly theoretical, often touching on basic concepts related to assessment techniques without delving into the practical applications necessary for real-world classroom situations. In many instances, the training sessions were described as too brief to leave a lasting impact on the educators' skills and confidence or, in some unfortunate cases, completely nonexistent. Such a glaring lack of adequate preparation not only raises concerns regarding the individual effectiveness of educators but also casts doubt on their ability to proficiently adapt to and implement modern assessment techniques within educational settings. This situation underscores a pressing need for systemic changes in teacher training programs to ensure that educators are adequately prepared to meet the evolving demands of teaching and assessment in today's educational landscape.

Table 6. Assessment, training, and resource access

Criteria	Australia	Jordan	Malaysia
Frequency of PD workshops	Moderate	Low	Inconsistent
Focus of training	Mixed (some practical tools)	Mostly theoretical	Varied by region/school type
Resource availability	Moderate in metro areas	Poor	Limited to the public sector
Teacher confidence in assessment reform	Low to moderate	Very low	Moderate in select schools

Australian educators have raised significant concerns about the current approach to assessment reform within the education system, noting that it frequently falls upon individual schools to implement changes without a cohesive national strategy. One secondary school teacher highlighted assessment methods' confusion, stating, "*we are told to assess for learning, but no one shows us how. We Google rubrics or share templates informally.*" This indicates a disconnect between policy and practice, where teachers are expected to adapt to new assessment strategies without proper training or resources. In comparison, educators from Jordan and Malaysia reported even fewer support systems, particularly within rural and public schooling environments. In Malaysia, many teachers voiced their eagerness for reform and improvement in assessment techniques; however, they are hindered by a significant absence of practical resources and limited time to dedicate to such initiatives. The challenges these educators face underscore a critical need for improved infrastructure and support in educational assessment, particularly in areas that have traditionally been underserved or under-resourced. Dead Horse Theory interpretation: this reflects strategies like 'changing the saddle' or 'repainting the horse' cosmetic changes that fail to address the root of the problem.

4.4. Comparative summary

The cross-national comparison reveals intriguing insights, as it indicates that, despite the contextual differences in each of the three countries, they exhibit remarkably similar behaviors that align with the Dead Horse Theory. This theory suggests a stubborn adherence to outdated practices, an observation that holds across diverse cultural backdrops. In these nations, institutions persistently enforce and rely on assessment systems that are no longer effective or relevant. This continuation occurs due to several interrelated factors: first, institutional inertia plays a significant role, as organizations often resist change and prefer to maintain the status quo. Second, prevailing cultural expectations within each society encourage adherence to traditional methods, discouraging innovation. Finally, a notable lack of professional development opportunities further hampers the adoption of more progressive and effective assessment strategies. Consequently, combining these elements creates a landscape where outdated practices thrive, hindering advancement and adaptation.

In Australia, Jordan, and Malaysia, outdated assessment practices remain a significant challenge, entrenched by systemic, cultural, and institutional inertia, as shown in Table 7. This persistence reveals a reluctance or inability to move beyond established methods, hindering the potential for educational and

professional growth in these regions. The Dead Horse Theory is a poignant metaphor for the symbolic efforts to 'fix' these outdated systems. Despite the good intentions behind these initiatives, they often lack the depth and scope required for meaningful transformation. As a result, while stakeholders may engage in various activities aimed at reform, they frequently fail to achieve genuine progress, perpetuating a cycle of insufficient change and disappointment. It raises an important question about effectively addressing the inertia and catalyzing real reform that impacts educators and learners alike.

Table 7. Summary of challenges by country (1=low, 3=high)

Challenge	Australia	Jordan	Malaysia
Bureaucratic rigidity	2	3	2
Cultural attachment to exams	2	3	2
Lack of practical teacher training	2	3	2
Pressure from standardized testing	3	3	3
Readiness for alternative assessment	2	1	2

5. DISCUSSION

The findings of this comprehensive study reveal the deeply entrenched nature of assessment practices that persist across diverse educational systems within Australia, Jordan, and Malaysia. This is particularly striking given the prevailing policy narratives that emphasize the importance of innovation and student-centered learning approaches. Through the Dead Horse Theory lens, the study thoroughly unpacks the reasons behind the education systems' clinging to dysfunctional tools and outdated practices. Rather than engaging in systemic reform, these systems pursue symbolic or superficial changes that do not address underlying issues.

All three countries have a significant institutional and societal preference for high-stakes, standardized testing, which continues to dominate educational practices. In Australia, for instance, the NAPLAN and the Australian Tertiary Admission Rank (ATAR) are critical benchmarks for measuring school and student performance. Meanwhile, in Jordan and Malaysia, national examinations such as the *Tawjihi*, the SPM, and the PT3 are central to the educational landscape. These testing instruments have become institutional 'dead horses', representing structures that, while no longer meeting the needs of contemporary education, are retained mainly for their symbolic significance rather than their functional contributions.

This concerning phenomenon aligns with findings from earlier studies, specifically Clare *et al.* [7], which argue that standardization often replaces genuine accountability with superficial quality metrics. Additionally, critiques of Australia's test-driven accountability culture, as noted by Lingard *et al.* [55], highlight how such practices are mirrored in Jordan and Malaysia, where limited decentralization coupled with entrenched examination ideologies contribute to the issues at hand [56], [57]. The metaphorical 'beating of the dead horse' is evident in the increased frequency of testing, the emergence of new performance metrics, and superficial tweaks to curricula, all of which fail to address the fundamental inefficiencies of existing assessment models.

Cultural context is essential in shaping attitudes toward assessment across these nations. In both Jordan and Malaysia, academic grades are frequently viewed as a proxy for social status, a measure of family honor, and a determinant of economic opportunities. This deep cultural attachment to grades fosters a substantial resistance to alternative assessment forms, such as portfolio assessment, peer reviews, or diagnostic feedback. In Australia, despite a facade of progressive rhetoric around education, parental and media focus on rankings and school comparisons reinforces a reliance on numerical outcomes [58], further entrenching the issues related to standardized assessment.

The Dead Horse Theory offers valuable insights into how educational systems rationalize the continued use of outdated assessment practices to meet cultural expectations. An illustrative example is the notion of 'buying a stronger whip', which illustrates policies aimed at raising scores through stricter enforcement or expanded testing measures, rather than exploring innovative alternative assessments that may offer more meaningful insights into student learning. Such strategies are often politically safer and more quantifiable, appealing to policymakers even if they misalign with pedagogical best practices.

A notable and consistent finding across all three examined countries is the persistent disconnect between assessment policies and teachers' preparedness to implement them effectively in the classroom. Many educators face inadequate training, a glaring lack of continuous support, and significant resource shortages, all of which leave them ill-equipped to adopt alternative assessment methods. As a result, this gives rise to performative compliance, where teachers adopt the language of 'formative' or 'authentic' assessment without making meaningful changes to their teaching practices [53], [59]. In the Dead Horse

Theory context, this scenario aligns with actions such as 'repainting the horse' or 'changing the saddle'—superficial gestures that fail to confront and challenge the entrenched educational norms.

For instance, teachers in Jordan have reported attending workshops that present theoretical concepts without offering practical applications for their classrooms. Similarly, educators in Australia have expressed frustration at being instructed to innovate without being provided the time, tools, or clarity to do so effectively. While the core issues identified across these educational systems are shared, their manifestations vary significantly from one context to another. Australia, for example, exhibits more structural potential for educational reform, particularly through innovations at the state level and initiatives driven by the private sector. However, this potential is often neutralized by the strong influence of performance rankings and the public discourse surrounding test results, which detracts from genuine educational flexibility.

In contrast, Jordan's deeply centralized and exam-centric system is entrenched both institutionally and culturally. Reform efforts in this context are frequently more symbolic than substantive, often propelled by international benchmarks rather than a focus on grassroots needs [60]. Malaysia, on the other hand, appears to occupy a middle ground, with signs of change emerging particularly in urban areas and private schools; however, inconsistencies persist in implementing these reforms within the public sector and in securing community buy-in for them. This comparative analysis reveals an important insight: no educational system is immune to the 'dead horse' dilemma, where efforts to reform fail not due to a lack of awareness or understanding, but rather because of the complex interdependencies between policy, cultural attitudes, and system capacity.

5.1. Implications for policy and practice

Based on the outcome of this study, there are list of implications for policy and practice:

i) Rethinking accountability: systems must shift from test-based accountability to holistic models incorporating multiple student learning measures. Countries like Finland and Singapore have demonstrated how broader conceptions of learning can coexist with high performance [61].

ii) Cultural reframing: policymakers must engage communities in dialogues that reframe success beyond exam scores. Without a cultural shift, even well-designed reforms will fail to take root.

iii) Sustained professional development: short-term workshops are insufficient. Teachers need ongoing, practice-based professional learning communities (PLCs) that support reflective assessment design and evaluation.

iv) Systemic coherence: reform must be vertically aligned, from policy to practice, and horizontally supported across curriculum, pedagogy, and assessment; piecemeal interventions (changing the rider) risk reinforcing dysfunction rather than eliminating it.

v) Political courage and narrative change: transitioning away from legacy assessment systems requires a strong commitment to change. Leaders must be prepared to face public backlash, navigate uncertainty, and communicate a clear vision of meaningful learning.

5.2. Dead Horse Theory as a diagnostic framework

One of the most significant contributions of this study lies in its innovative application of the Dead Horse Theory, which serves as a valuable heuristic for diagnosing the stagnation of reform efforts within various institutions. This theory, presented with humor, effectively highlights the absurdity and inefficiency of investing resources in broken systems that yield no significant improvement or progress. Moreover, it sheds light on the various coping mechanisms institutions adopt in the face of such stagnation. These coping strategies range widely, encompassing denial, where stakeholders refuse to acknowledge the issues at hand, distraction, where focus is shifted to less critical problems instead of addressing the root causes, and symbolic overcompensation, where superficial changes are made to create an illusion of progress.

By creatively labeling these behaviors, such as "buying a stronger whip", "changing riders", or even "rebranding the horse", the Dead Horse Theory not only provides a satirical commentary on ineffective practices but also equips educators and policymakers with a nuanced vocabulary. This vocabulary is instrumental in recognizing these performative changes that do not lead to substantial improvement. Ultimately, the theory advocates a shift towards transformative actions, urging stakeholders to move beyond surface-level reforms and engage in more profound and meaningful change.

6. CONCLUSION

The study has examined how outdated and ineffective assessment practices persist in the educational systems of Australia, Jordan, and Malaysia, despite decades of reform rhetoric and innovation initiatives. Using the Dead Horse Theory as an analytical lens, the research identified core patterns of institutional behavior that sustain these legacy systems. These include bureaucratic rigidity, cultural fixation on examination results, and inadequate professional development infrastructure, all of which contribute to the

maintenance of assessment methods that no longer serve learners effectively. The metaphor of 'beating a dead horse' vividly illustrates how education systems often react to systemic failure with surface-level solutions: renaming policies, increasing test frequency, or rotating leadership, all while leaving the underlying structure untouched. These behaviors manifested differently in each country, shaped by national context, cultural values, and policy histories. However, the common denominator was a reluctance or inability to fundamentally question the logic of high-stakes, summative assessment as the primary indicator of student success.

In Australia, the education system experiences considerable pressure on schools due to standardization and league tables. Despite this, there remains some flexibility and policy space available for innovation within schools, allowing for creative teaching and learning approaches. In Jordan, the centralized control of the education sector and the deeply rooted social expectations surrounding examination scores are significant barriers to change. These societal norms make it difficult for schools and policymakers to implement reforms that deviate from traditional assessment methods. Although reform efforts have been initiated in Malaysia to introduce alternative approaches to education in specific contexts, the entire system still faces challenges such as uneven implementation across regions and resource disparities. Teachers in all three countries find themselves caught in a difficult position, balancing the pressure to adopt innovative pedagogical practices with the ongoing influence of traditional, culturally reinforced metrics emphasizing exam performance. This tension highlights the complex dynamics educators face in striving to promote meaningful educational change amidst existing constraints.

The findings underscore the critical need for a systemic paradigm shift in educational assessment that moves beyond symbolic gestures and embraces truly transformative approaches. This requires policy reform, cultural change, professional empowerment, and reimagining how we define and measure learning. Teachers must have conceptual frameworks and practical tools to implement authentic, learner-centered assessments. Policymakers must be willing to disrupt legacy systems and withstand the political and societal challenges. Ultimately, the Dead Horse Theory is more than a humorous metaphor; it is a cautionary framework. It warns against the comfort of routine and the illusion of progress. It challenges educators and decision-makers to ask the difficult questions. To create meaningful change, education systems must not only stop 'beating the dead horse' but also dare to dismount entirely, letting go of what no longer serves, and co-create new assessment models that are inclusive, responsive, and aligned with the complex realities of 21st-century learning.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mohammad Awad AlAfnan	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	
Siti Fatimah MohdZuki		✓		✓	✓	✓		✓	✓	✓	✓			
Shefa Mohammad AlAfnan	✓		✓				✓	✓		✓	✓			

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT
The authors state no conflict of interest.

INFORMED CONSENT
We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL
The research followed all national rules and institutional policies and was approved by the authors' institutional review board.

DATA AVAILABILITY
The data that support the findings of this study are available from the corresponding author, [MAA], upon reasonable request.

REFERENCES
[1] A. D. Larsen, "Is critical thinking and innovation possible within the current compliance environment? Perceptions of Arizona Principals," Ph.D. dissertation, Northern Arizona University, Flagstaff, Arizona, United States, 2023.
[2] D. C. Youvan, "The Dead Horse of American culture: a reflection on decline, denial, and the limits of reform," 2025, doi: 10.13140/RG.2.2.22742.77122.
[3] A. Shalaby, "The digital sustainable growth model is disrupting the traditional macro-microeconomics in the 3D economy," *SSRN*, pp. 1–25, 2024, doi: 10.2139/ssrn.5160079.
[4] R. Morris, T. Perry, and L. Wardle, "Formative assessment and feedback for learning in higher education: a systematic review," *Review of Education*, vol. 9, no. 3, p. e3292, Oct. 2021, doi: 10.1002/rev3.3292.
[5] R. J. Shavelson, O. Zlatkin-Troitschanskaia, K. Beck, S. Schmidt, and J. P. Marino, "Assessment of university students' critical thinking: next generation performance assessment," *International Journal of Testing*, vol. 19, no. 4, pp. 337–362, Oct. 2019, doi: 10.1080/15305058.2018.1543309.
[6] H. H. Tabrizi and J. House, "Moving beyond conventional tests: performance assessment as an alternative for evaluating students' academic translations," *Contrastive Pragmatics*, vol. 6, no. 3, pp. 608–639, 2025, doi: 10.1163/26660393-bja10139.
[7] I. Clare *et al.*, "In their own words: removing barriers for 'more able' disadvantaged students in secondary schools - Scoping review," Project Report. York St John University, 2025. [Online]. Available: https://ray.yorksj.ac.uk/id/eprint/11951/
[8] Y. Oc and H. Hassen, "Comparing the effectiveness of multiple-answer and single-answer multiple-choice questions in assessing student learning," *Marketing Education Review*, vol. 35, no. 1, pp. 44–57, Jan. 2025, doi: 10.1080/10528008.2024.2417106.
[9] D. D. Stevens, *Introduction to rubrics: an assessment tool to save grading time, convey effective feedback, and promote student learning*, 2nd ed. New York: Routledge, 2023, doi: 10.4324/9781003445432.
[10] S. K. Benson, "The evolution of Jordanian inclusive education policy and practice," *FIRE: Forum for International Research in Education*, vol. 6, no. 1, pp. 102–121, Jan. 2020, doi: 10.32865/fire202061177.
[11] L. Shan, "A gene for maintaining sustainable cultural specificity of contemporary architecture—national deep-rooted nature," *Sustainability*, vol. 14, no. 3, p. 1330, Jan. 2022, doi: 10.3390/su14031330.
[12] R. Rays, "The relationship between Georgia public school educational funding sources and academic achievement," Ph.D. dissertation, Walden University, Minneapolis, Minnesota, United States, 2018.
[13] A. M. Oliemat, M. M. Al-Gaseem, and M. H. Alkhazaleh, "Professional development courses and alternative assessment strategies for in-service science education teachers," *International Journal of Education and Practice*, vol. 13, no. 1, pp. 178–189, Nov. 2024, doi: 10.18488/61.v13i1.3990.
[14] G. K. Joseph, "Bridging cultures fostering inclusive assessment of 21st century skills," in *21st Century Teaching and Learning in Classrooms*, G. K. Joseph, K. Y. Benedict, S. P. K., U. P., Eds., Chikmagalur: Iterative International Publishers (IIP), Selfypage Developers Pvt Ltd., 2024, pp. 1–9, doi: 10.58532/nbennurctch1.
[15] S. Srivastava, "The evolution of education: navigating 21st-century challenges," *International Journal for Multidisciplinary Research*, vol. 5, no. 5, pp. 1–9, 2023, doi: 10.36948/ijfmr.2023.v05i05.6314.
[16] M. M. M. B. Díaz *et al.*, "The power of curriculum to transform education: How education systems incorporate 21st century skills to prepare students for today's challenges," Jul. 2022, doi: 10.18235/0004360.
[17] M. A. AlAfnan, "Taxonomy of educational objectives: teaching, learning, and assessing in the information and artificial intelligence era," *Journal of Curriculum and Teaching*, vol. 13, no. 4, pp. 173–191, Aug. 2024, doi: 10.5430/jct.v13n4p173.
[18] C. Yang, L. Luo, M. A. Vadillo, R. Yu, and D. R. Shanks, "Testing (quizzing) boosts classroom learning: a systematic and meta-analytic review," *Psychological Bulletin*, vol. 147, no. 4, pp. 399–435, Apr. 2021, doi: 10.1037/bul0000309.
[19] G. F. Madaus and M. Clarke, "The adverse impact of high stakes testing on minority students: evidence from 100 years of test data," in *Raising standards or raising barriers? Inequality and high stakes testing in public education*, G. Orfield and M. Kornhaber, Eds., New York: The Century Foundation, 2001, pp. 1–53.
[20] A. Zakharov and M. Carnoy, "Does teaching to the test improve student learning?" *International Journal of Educational Development*, vol. 84, p. 102422, Jul. 2021, doi: 10.1016/j.ijedudev.2021.102422.
[21] A. Kohn and L. Bridges, *The case against standardized testing: raising the scores, ruining the schools*. Portsmouth: Heinemann, 2000.
[22] D. Ravitch, "The myth of charter schools," *New York Review of Books*, no. 11, pp. 1–11, 2010.
[23] Y. Zhao, *Catching up or leading the way: American education in the age of globalization*. Alexandria, VA: ASCD, 2009.
[24] J. Whalen and C. Mouza, "ChatGPT: Challenges, opportunities, and implications for teacher education," *Contemporary Issues in Technology and Teacher Education*, vol. 23, no. 1, pp. 1–23, 2023.

[25] H. L. Andrade and S. M. Brookhart, "Classroom assessment as the co-regulation of learning," *Assessment in Education: Principles, Policy and Practice*, vol. 27, no. 4, pp. 350–372, 2020, doi: 10.1080/0969594X.2019.1571992.

[26] J. Noonan, "An affinity for learning: teacher identity and powerful professional development," *Journal of Teacher Education*, vol. 70, no. 5, pp. 526–537, 2019, doi: 10.1177/0022487118788838.

[27] C. Gist, I. Jackson, B. Nightengale-Lee, and K. Allen, "Culturally responsive pedagogy in teacher education," in *Oxford Research Encyclopedia of Education*, Oxford UK: Oxford University Press, 2019, pp. 1–26, doi: 10.1093/acrefore/9780190264093.013.266.

[28] J. A. Banks, *Transforming multicultural education policy and practice: expanding educational opportunity. Multicultural Education Series.* New York: Teachers College Press, 2021.

[29] K. H. Au, *Literacy achievement and diversity: keys to success for students, teachers, and schools*. New York: Teachers College Press, 2011.

[30] N. Miled, "Educational leaders' perceptions of multicultural education in teachers' professional development: a case study from a Canadian school district," *Multicultural Education Review*, vol. 11, no. 2, pp. 79–95, 2019, doi: 10.1080/2005615X.2019.1615249.

[31] W. Au, "When multicultural education is not enough," *Multicultural Perspectives*, vol. 19, no. 3, pp. 147–150, 2017, doi: 10.1080/15210960.2017.1331741.

[32] J. Hong, D. C. Francis, and P. A. Schutz, "Reconceptualizing teacher identity development," *Educational Psychologist*, vol. 59, no. 3, pp. 159–176, 2024, doi: 10.1080/00461520.2023.2292713.

[33] C. N. Carter, "How a high-stakes standardized biology assessment affects teachers' pedagogical planning and implementation for minority students of low socioeconomic status," Ph.D. dissertation, Cabrini University, Radnor, Pennsylvania, United States, 2020.

[34] D. W. Putwain and N. P. von der Embse, "Cognitive–behavioral intervention for test anxiety in adolescent students: do benefits extend to school-related wellbeing and clinical anxiety," *Anxiety, Stress and Coping*, vol. 34, no. 1, pp. 22–36, 2021, doi: 10.1080/10615806.2020.1800656.

[35] D. Koretz, *The testing charade pretending to make schools better*. Chicago: University of Chicago Press, 2017.

[36] R. N. Wafubwa, "Role of formative assessment in improving students' motivation, engagement, and achievement: a systematic review of literature," *International Journal of Assessment and Evaluation*, vol. 28, no. 1, pp. 17–31, 2020, doi: 10.18848/2327-7920/CGP/V28I01/17-31.

[37] I. Rissanen, S. Laine, I. Puusepp, E. Kuusisto, and K. Tirri, "Implementing and evaluating growth mindset pedagogy – a study of Finnish elementary school teachers," *Frontiers in Education*, vol. 6, p. 753698, Sep. 2021, doi: 10.3389/feduc.2021.753698.

[38] L. Indriati, N. Mai, and H. T. Yeen-Ju, "Enhancing authentic assessment in large-class design education through authentic project-based learning," *International Journal of Learning, Teaching and Educational Research*, vol. 23, no. 9, pp. 432–452, 2024, doi: 10.26803/ijlter.23.9.22.

[39] R. P. Loweth, S. R. Daly, J. Liu, and K. H. Sienko, "Assessing needs in a cross-cultural design project: student perspectives and challenges," *International Journal of Engineering Education*, vol. 36, no. 2, pp. 712–731, 2020.

[40] M. E. Walker, M. Olivera-Aguilar, B. Lehman, C. Laitusis, D. Guzman-Orth, and M. Gholson, "Culturally responsive assessment: provisional principles," *ETS Research Report Series*, vol. 2023, no. 1, pp. 1–24, 2023, doi: 10.1002/ets2.12374.

[41] E. P. O'Dwyer, J. R. Sparks, and L. N. Oláh, "Enacting a process for developing culturally relevant classroom assessments," *Applied Measurement in Education*, vol. 36, no. 3, pp. 286–303, 2023, doi: 10.1080/08957347.2023.2214652.

[42] L. Zhou and H. C. Hassan, "Role of culturally relevant instruction in enhancing social-emotional learning for marginalised students," *Conhecimento & Diversidade*, vol. 16, no. 44, pp. 450–469, 2024, doi: 10.18316/rcd.v16i44.12267.

[43] A. Verger, C. Fontdevila, and L. Parcerisa, "Reforming governance through policy instruments: how and to what extent standards, tests and accountability in education spread worldwide," *Discourse*, vol. 40, no. 2, pp. 248–270, 2019, doi: 10.1080/01596306.2019.1569882.

[44] S. Pastore and H. L. Andrade, "Teacher assessment literacy: a three-dimensional model," *Teaching and Teacher Education*, vol. 84, pp. 128–138, 2019, doi: 10.1016/j.tate.2019.05.003.

[45] H. Aksom, "Institutional inertia and practice variation," *Journal of Organizational Change Management*, vol. 35, no. 3, pp. 463–487, 2022, doi: 10.1108/JOCM-07-2021-0205.

[46] X. Xu, W. Shen, A. Y. M. A. Islam, and Y. Zhou, "A whole learning process-oriented formative assessment framework to cultivate complex skills," *Humanities and Social Sciences Communications*, vol. 10, no. 1, pp. 1–15, 2023, doi: 10.1057/s41599-023-02200-0.

[47] H. Gardner, *Frames of mind: the theory of multiple intelligences*. New York: Basic Books, 1983.

[48] F. Guay, "Applying self-determination theory to education: regulations types, psychological needs, and autonomy supporting behaviors," *Canadian Journal of School Psychology*, vol. 37, no. 1, pp. 75–92, 2022, doi: 10.1177/08295735211055355.

[49] D. P. Wolf, "Portfolio assessment: sampling student work," *Educational Leadership*, vol. 46, no. 7, pp. 35–39, Apr. 1989.

[50] D. Fisher, N. Frey, and J. Hattie, *Teaching literacy in the visible learning classroom, grades K-5*. Thousand Oaks, CA: Corwin Press, 2017.

[51] G. Ladson-Billings, "Three decades of culturally relevant, responsive, & sustaining pedagogy: what lies ahead?" *Educational Forum*, vol. 85, no. 4, pp. 351–354, 2021, doi: 10.1080/00131725.2021.1957632.

[52] R. E. Slavin, "How evidence-based reform will transform research and practice in education," *Educational Psychologist*, vol. 55, no. 1, pp. 21–31, 2020, doi: 10.1080/00461520.2019.1611432.

[53] L. A. Shepard, W. R. Penuel, and K. L. Davidson, "Design principles for new systems of assessment," *Phi Delta Kappan*, vol. 98, no. 6, pp. 47–52, 2017, doi: 10.1177/0031721717696478.

[54] V. Braun, V. Clarke, and N. Hayfield, "'A starting point for your journey, not a map': Nikki Hayfield in conversation with Virginia Braun and Victoria Clarke about thematic analysis," *Qualitative Research in Psychology*, vol. 19, no. 2, pp. 424–445, 2022, doi: 10.1080/14780887.2019.1670765.

[55] B. Lingard, S. Sellar, and S. Lewis, "Accountabilities in schools and school systems," *Oxford Research Encyclopedia of Education*, Oxford UK: Oxford University Press, 2017, pp. 1–28, doi: 10.1093/acrefore/9780190264093.013.74.

[56] M. A. F. Al-Khawaldeh, "The effect of simulation activities on improving speaking skill among fourth basic grade female students in Jordan," *Ulum Tarbawiyyah (Educational Sciences)*, vol. 32, no. 2, pp. 57–84, 2024, doi: 10.21608/ssj.2024.353860.

[57] Y. Wan *et al.*, "Improving automatic source code summarization via deep reinforcement learning," in *ASE 2018 - Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, 2018, pp. 397–407, doi: 10.1145/3238147.3238206.

[58] J. Rose, S. Low-Choy, P. Singh, and D. Vasco, "NAPLAN discourses: a systematic review after the first decade," *Discourse: Studies in the Cultural Politics of Education*, vol. 41, no. 6, pp. 871–886, 2020, doi: 10.1080/01596306.2018.1557111.

[59] B. Huisman, N. Saab, J. van Driel, and P. van den Broek, "Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance," *Assessment and Evaluation in Higher Education*, vol. 43, no. 6, pp. 955–968, 2018, doi: 10.1080/02602938.2018.1424318.

[60] M. A. Mizher, A. A. Mazhar, and M. A. A. Mizher, "A review of mobile cloud computing in education during the COVID-19 pandemic in Jordan," in *ACM International Conference Proceeding Series*, 2021, pp. 187–193, doi: 10.1145/3489088.3489101.

[61] S. J. Ball, *The education debate*, 4th ed. Bristol, UK: Policy Press, 2021.

BIOGRAPHIES OF AUTHORS

Mohammad Awad AlAfnan is an associate professor of Applied Linguistics/Business Communication. He participated in a number of international conferences and published journal articles and books on AI in education, teaching and learning, critical discourse analysis, and mass/social media. He taught undergraduate and postgraduate courses, moderated master's theses and dissertations, and supervised Ph.D. dissertations. He can be contacted at email: Mohammad.al-afnan@aum.edu.kw.

Siti Fatimah MohdZuki studied Electrical Engineering at RMIT University in Melbourne, Australia, and holds a Master's degree in Computer Science from Universiti Teknologi MARA (UiTM) in Shah Alam. She has published extensively on topics related to Artificial Intelligence, as well as teaching and learning methodologies. She can be contacted at email: SitiFatimahMohdZUki@yahoo.com.

Shefa Mohammad AlAfnan is affiliated with the International Islamic University, Kuala Lumpur, Malaysia, and has diverse research interests, including learning and communication. Her publications include work on educational taxonomies and learning frameworks. She can be contacted at email: ShefaMohammadAlAfnan@yahoo.com.