

Applying Assessment Principles during Emergency Remote Teaching: Challenges and Considerations

Wid Hasen Allehaiby

Department of English, College of Science and Health Professions
King Saud bin Abdulaziz University for Health Sciences
Jeddah, Saudi Arabia
Corresponding Author: lehaibyw@ksau-hs.edu.sa

Sara Al-Bahlani

English Department, University of Technology and Applied Sciences
Nizwa, Oman

Received: 4/8/2021

Accepted: 10/4/2021

Published: 12/15/2021

Abstract

One of the main challenges higher educational institutions encounter amid the recent COVID-19 crisis is transferring assessment approaches from the traditional face-to-face form to the online Emergency Remote Teaching approach. A set of language assessment principles, practicality, reliability, validity, authenticity, and washback, which can be applied to any academic subject, are critical within the design of any task that aims to assess learning. This review paper discusses how assessment approaches need to be modified in a time of crisis. It determines the position assessment should adopt within emergency remote teaching methods and analyzes the fundamental characteristics of five principles of assessment and how they can be accomplished with emergency remote teaching approaches. Furthermore, this paper evaluates the vulnerability and viability of the five assessment principles, examines the application of online assessment on a holistic level, and puts forward a set of recommendations to ensure the assessment principles are achieved within emergency remote learning contexts. The paper concludes with the notion that the construct of time, which is inherent within the principle of practicality, is the most significant when developing online assessments as it is this characteristic that is the most at risk. In addition, we suggest that the assessments that are implemented during emergency remote learning involve open-ended, as opposed to close-ended, questions and highlight the importance of educators demonstrating flexibility and understanding toward their students.

Keywords: assessment principles, COVID-19 crisis, emergency remote teaching, formative assessment, summative assessment

Cite as: Allehaiby, W. H. , & Al-Bahlani, S. (2021). Applying Assessment Principles during Emergency Remote Teaching: Challenges and Considerations. *Arab World English Journal*, 12 (4) 3-18. DOI: <https://dx.doi.org/10.24093/awej/vol12no4.1>

Introduction

During the ongoing global coronavirus (COVID-19) pandemic, the education sector in most countries has had to suspend or transform educational programs. Many governments have made the tough but necessary choice to suspend face-to-face learning and substitute it with online learning for higher education institutions to safeguard students' and teachers' health (Cahapy, 2020; Huber & Helm, 2020; Huang, Tlili, Chang, Zhang, Nascimbeni, & Burgos, 2020). Consequently, and where feasible, educators have turned en masse to online teaching to cause the least disruption possible. For example, due to the pandemic, Australian universities have turned to online learning (Reedy, Pfitzner, Rook, & Ellis, 2021). Similar responses have been observed in Saudi Arabia (Alqabbani, Almuwais, Benajiba, & Almoayad, 2020), Jordan (Al-Salman & d Haider, 2021), Oman (Guangul, Suhail, Khalit, & Khidhir, 2020), China (Huang et al., 2020; Zhang, Wang, Yang, & Wang, 2020), and various other nations including Germany, the United Arab Emirates, and Malaysia (Crawford et al., 2020). Plans approved by the Ministry of Education in Saudi Arabia, for example, have led to the implementation of initiatives aimed at ensuring that online Learning Management Systems (LMS), such as Blackboard, can reach as many students as possible (MOE, 2020).

There is nothing new about online teaching in itself: Many educators were already making substantial use of online learning resources before the pandemic started. What has caused concern is the unprecedented rapidity and extent of the transformation from face-to-face to remote modes, especially for those who had never given, or received, an online lesson before the current crisis. In this situation, Hodges, Moore, Lockee, Trust, and Bond (2020) have coined the term *Emergency Remote Teaching* (ERT) to describe a temporary change in how learning experiences are delivered in response to a crisis. It often involves remote teaching and learning as opposed to face-to-face classroom instruction. The instructional delivery is typically returned to the initial mode after the crisis has come to an end.

ERT differs from standard 'online learning' in that it does not necessarily aim to recreate the structured educational environment of the classroom but, rather, to ensure a rapid and reliable means for students to gain the necessary temporary access to educators should an emergency occur (Hodges et al., 2020). The educational goal of ERT is to guarantee that course materials are made available to all students who need them and that the latter continue to receive all the pedagogical support necessary to carry on learning throughout the emergency. Hence, the onus under ERT is on educators, and the priority is on delivering information. As teachers, and the management of higher education institutions, have likely been obliged to switch to online teaching (whether synchronous or asynchronous) rather than chosen to make the change, it is probable that they have had insufficient time to prepare specific materials. Instead, they may well simply be reworking existing material designed for classroom use, including assessment measures.

Hodges et al. (2020) suggest that under ERT, the role of assessment is less prominent than during regular times. They stress that all actors in the learning process, including students, educators, parents, and carers, are likely to be living through a time of profound disruption and, therefore, may set aside their everyday priorities, including an insistence on the need to study. Consequently, ERT places less emphasis on checking whether students have achieved learning

outcomes and more on the speed with which organizers ensure that all the necessary resources are in place and operational (Hodges et al., 2020).

The events that unfolded in educational settings in the aftermath of the onset of COVID-19 represented one of the worst global interruptions to education in the 21st century. As such, there is very little prior research that has attempted to delineate the implementation of assessment principles when learning experiences are shifted abruptly from face-to-face to online contexts. Furthermore, to the best of our knowledge, no prior studies have put forward an emergency plan that can be followed to ensure a smooth transition to ensure assessment standards are upheld during times of crises. Therefore, this study aimed to elucidate on what assessment constructions need to be considered during emergency situations. It examines methods of navigating and ensuring assessment principles in a manner that ensures equitability for all students while also encouraging instructors to take into consideration the factors that ensure instruction continues and learning outcomes are evaluated without putting the factor of assessment at risk.

The current paper investigates the place of assessment within ERT. It is organized as follows: Firstly, the principles of traditional assessment are analyzed; following that, the operation of these principles under ERT is examined. Then, we explain online assessment in the wider context. Lastly, and in light of the discussion presented in the third section, we provide recommendations for using assessment measures within ERT.

Literature Review

Principles of Assessment and their Implementation within Emergency Remote Teaching

Given the technology now available to support remote learning and teaching, it might seem a simple matter to transfer modes of assessment from the classroom into the online environment. However, a closer investigation is necessary to ascertain whether assessment in the classroom is, in fact, both suitable and transferable to the context of online tuition, particularly ERT. As such, this study bridges the existing gaps in understanding by delineating a set of assessment principles and the factors that should be taken into consideration when educational institutions shift to online assessment in emergency situations. This section, therefore, presents Brown and Abeywickrama's (2010) assessment principles and examines how they aligned with assessment measures adopted under ERT.

Brown and Abeywickrama's (2010) focus was on the best way to structure educational assessments in language learning. They devised a series of assessment principles, namely practicality, reliability, validity, authenticity, and washback. Nonetheless, they also stated that these principles "apply to assessments of all kinds" (p.25). This is evidenced in multiple publications on assessment principles in education, including McAlpine (2002) and McMillan (2000).

The first of Brown and Abeywickrama's (2010) principles, practicality, requires an assessment to meet specific criteria, namely that it is cost-efficient, inexpensive, and should be marked within a reasonable amount of time through using a specific scoring process or rubric. Furthermore, an assessment does not need an unreasonable amount of time for either the student to complete or the teacher to mark. In addition, a test should not be difficult to administer (for example, any external equipment required, such as audio-visual devices, should be available and

in working order). Within the principle of practicality, then, it seems Brown and Abeywickrama (2010) particularly stress the value of time, suggesting that this criterion predominates over the other evaluation factors.

In transferring all five of Brown and Abeywickrama's (2010) principles from the classroom to an online context, it is the principle of practicality that suffers the most. Firstly, regarding the equipment necessary for students to complete the test under ERT and because of the shift to online learning, every student must have access to a computer or laptop, as well as good internet connectivity. This is not the case for all students, and whereas in regular times, students who do not have their computer might visit their university library and use a computer there, under ERT, this avenue is closed to them. Students who do not have their own means to access an online test are therefore at a gross disadvantage; hence, testing in an online setting, in this case, is impractical. Moreover, we cannot assume that all educational institutions in all countries have access to LMSs, such as Blackboard. Not only will institutions that do not have such resources have to find the funds to buy them and the personnel to install them, but doing so will take time which, under ERT, is lacking. Hence, their students will be at a disadvantage because their tests will be suspended or reduced until these resources are in place.

As noted above, the critical important element of the principle of practicality is time, and under ERT, there may simply be too little of it to create or recreate appropriate tests. Even in the best of cases, teachers will have to rework some elements of tests, such as classroom quizzes, to make them suitable for the new environment. In the worst of cases, they will have to change the nature of the assessments completely; for example, if students were going to be assessed by carrying out a laboratory experiment, under ERT, they might have to fill in a sheet of multiple-choice questions (MCQs).

The last element within the principle of practicality that has a bearing on assessment within ERT is the feasibility of transferring the scoring rubric or another evaluation process from the classroom to an online setting. For example, certain institutions have decided to replace end-of-term or midterm assessments with continuous assessment, simultaneously devising how to translate a scoring system developed for the first type (percentages, grades, pass/fail, etc.) to the second type. Discussion is ongoing as to the implications of making this type of change. For example, how does this change affect future employers or graduate admissions departments' understanding of the new types of assessment, and how can they compare between a class graduating under one system and classes which graduated under a different one as they consider accepting new applicants into their institutions.

The second principle enumerated by Brown and Abeywickrama (2010) to assess the construction of tests is reliability. For Brown and Abeywickrama (2010), a test is considered reliable if it meets two criteria: consistency and dependability. Thus, if the same test is taken on different occasions by the same or similar students, they will gain the same or similar grades. The factors contributing to the potential failure to meet the reliability criterion can be classified as student-related, rater-related, and test-related. Student-related reliability refers to the emotional or physical state of a student taking the assessment (e.g., feeling ill or suffering psychological problems such as anxiety). Rater reliability refers to the possibility that scoring is subject to human error in the form of the graders' inexperience, subjectivity, bias, or failure to

pay attention to the scoring criteria, as well as the possibility of disturbance in the examination hall, for example, from the noise outside. While test administration reliability affects the administration of specific assessments, rather than the test itself, test reliability refers to flaws in the construction of the assessment. For example, it could be very long, thus exhausting students, or it is required to be done within too short a time, thus pressurizing students. In worst cases, a test could include poorly written items that confuse students. All of these eventualities could negatively affect grades.

In the ERT context, too, test reliability can be undermined by several factors. As for student-related reliability, learners who are unfamiliar with online learning are likely to feel anxious when they are suddenly asked to take a test or assessment online, especially given the circumstances that have made this transfer necessary (i.e., ERT). At the same time, however, it is possible to strengthen assessment administration reliability under ERT, as students are less likely to suffer the anxiety which accompanies the experience of being in an examination hall. Instead, they may feel more relaxed and engaged taking an exam in their own homes, comfortably dressed, able to make a drink and visit the bathroom when they wish. Nonetheless, assessment administration reliability can also be undermined, given that in ERT, and within the broader context of online learning generally, it is far more difficult to prevent cheating. For example, there is no guarantee that the name on the assessment belongs to the person who sat it or that they did not consult their notes, books, and search engines during the test period. Hence, it is challenging to gain a clear view of whether a student has achieved the set learning outcomes. To partly circumvent this problem, teachers can use a setting under which students only see one question at a time and cannot go back to previous questions to change their answers. Alternatively, they might require students to answer within a set period, for example, 15 seconds, after which they are timed out and have to move on to the next question.

Although these measures may prevent cheating, they can have a counterproductive effect by increasing anxiety and stress levels or forcing students to choose an answer too quickly, thus making student performance unreliable for different reasons. This is particularly true for learners who are unused to online tuition and testing and who expect the less rigid norms of a classroom assessment. Finally, test administration reliability under ERT may frequently encounter technical challenges, among both teachers and students, that can impede the smooth running of an examination and even require IT professionals to provide assistance. This could cause disruption, add stress to students, and impact their grades.

Brown and Abeywickrama's (2010) third principle, which is arguably both the most important and the most complex of the five assessment principles, is that of validity. For a test to meet the validity criterion, it must precisely measure the skill or information it sets out to measure. Within language education, Brown and Abeywickrama's (2010) particular field, a writing test which sets out to investigate whether students can put together an argumentative essay in their second language must assess them on that ability, rather than, for example, testing whether they can translate vocabulary items between languages. The second area of validity refers to the fact that the test must reveal how well the person taking the test has achieved the stated learning objectives. When transferring tests from the classroom to the ERT context, it is essential that teachers must continue to respect the principle of validity. Learning objectives will not change when a curriculum leaves the face-to-face classroom to an online one; hence, the tests

accompanying them into the new online format must continue to measure effectively whether students achieve those objectives. The principal concern under ERT is how teachers can assess via online tests whether and to what extent students have achieved the learning outcomes required for them to graduate. One major issue with assessment alignments with the intended learning objectives is how well these objectives are communicated and taught online. Whether all students receive the same exposure or amount of learning they need to achieve a specific learning outcome is integral to assess how valid an online assessment is.

One can argue that disadvantaged students who have problems accessing their classes either because of financial constraints or lack access to a good network or device receive less training on the intended learning outcomes. Therefore, the validity of online assessments that measure specific learning outcomes is compromised by the inequity of students' exposure to these learning outcomes. As a result, many institutions have made recording online classes a must during ERT. Thus, if means are devised to change the assessment format without changing its fundamental focus on a particular outcome or set of outcomes, the validity of a test held under ERT should not be compromised.

Authenticity, Brown and Abeywickrama's (2010) fourth principle of assessment, is a measure of the extent to which test items reflect real-life instances which students use to interpret and demonstrate their knowledge. This is perhaps the most difficult of all the principles to maintain, particularly within the ERT context and in most online testing contexts. Providing authentic test items requires consideration of several factors, which include grasping which skill or knowledge the teacher is seeking to test for; ensuring that the information embedded in the test item is realistic and relatable; and, finally, judging how to present information within the test.

To implement the principle of authenticity within ERT, the information given in online test items must continue to meet all these standards. In some cases, this is relatively simple: For example, when asking students to read a text which depicts a realistic scenario and answer MCQs to demonstrate their understanding, authenticity is as easy to achieve in a test administered online as one distributed in the classroom. However, if students are taught to carry out some physical operation, such as using a laboratory tool, in the classroom, but then required to demonstrate their grasp of the technique under ERT, difficulties arise. They may not have the device or, if they do, the assessor may not be able to get a good view of how they manipulate it. In these circumstances, the assessment will cease to be authentic.

Finally, Brown and Abeywickrama (2010) consider 'washback', the effect exerted by any given test on how curricula are designed and taught, and what learning behaviors are expected from. According them, "a test that provides beneficial washback positively influences what [and] how teachers teach and positively influences what and how learners learn" (p.38). It also provides feedback that induces learning improvement. It should also allow for teacher feedback so students can enhance their learning (Brown & Abeywickrama, 2010). However, washback is considered more formative than summative because the latter is limited to a letter grade 'feedback' as opposed to detailed feedback that a student can build on. For washback to be beneficial, interaction between the teacher and students is needed. Hence, to achieve this principle under ERT, student and teacher interaction is necessary during and after an assessment. As such, it is crucial to identify ways to ensure that such an interaction exists. One plausible

solution is to encourage individual and whole-class discussions after each online assessment. One-to-one discussions between a teacher and a student can generate individual feedback that learner could use to understand what they need to improve. The whole class feedback discussion enhances learner understanding and acknowledgment of needs improving. While it can be time-consuming to deliver feedback of this nature, it reflects what a teacher would do in a traditional face-to-face mode. Finding the suitable platform for such interactions and feedback negotiation might be another challenge that can be eliminated by providing the right LMS and proper training for teachers and students on various technologies that allow smooth and easy one-to-one and whole-class interaction.

Clearly, there is an interplay among all these factors when assessment is recreated under ERT: For example, teachers under the pressure of time (the principal construct within practicality) to ensure that students are taught the entire curriculum after the delay imposed by the onset of the ERT may find that tests which are hastily composed to save time and get results in by a predetermined date do not meet the principles of validity and washback. Equally, not all these five assessment principles have to be recreated entirely in ERT; failing to comply with certain constructs completely will not weaken an assessment as much as the failure to completely comply with others, as further discussed below.

In the university context in Saudi Arabia, there has been concern that implementing online assessment so suddenly could sacrifice one or more of the above-mentioned principles and endanger students' performance by failing to assess their learning accurately. Furthermore, due to an inability to provide students with a fair opportunity and the required technologies and amid efforts to prevent the risk of contracting the coronavirus, the Ministry of Education took the decision that all students would be permitted to progress to the next level and that the grades from the first semester would also be held for the second semester. Furthermore, the Ministry also decreed that their grades would not be taken into consideration in their final GPAs (SPA, 2020). Although this plan may be far from ideal, it was deemed appropriate given the time pressure, need to come up with a decision whether to conduct high-stakes exams, the requirement to take infrastructure limitations into account, and the risk of further spread of the virus. Consequently, it may be that under ERT, priority must be given to ensuring that courses continue to be delivered, with testing and assessment relegated to a position of secondary importance, as was the case during the first wave of the pandemic in April 2020 in Saudi Arabia (SPA, 2020). The section below investigates this argument in more detail by discussing the standard approaches to assessment of online tuition. After that, recommendations are made for the implementation or modification of these standard approaches to fit the context of ERT.

Assessment in Online Learning Settings

The essential aim of any assessment of learning, either face-to-face or remote, is to check whether, and to what extent, a student has achieved predetermined learning outcomes. Educational institutions in the 21st century are seeking ways to understand which outcomes are appropriate to a constantly changing globalized and digital world and devise ways to deliver their educational programs accordingly (Jerald, 2009; Kereluik, Mishra, Fahnoe, & Terry, 2013). Consequently, these institutions must also understand and develop new means or modify old ones to evaluate these new outcomes and pedagogical approaches (Boitshwarelo et al., 2017). Several scholars have examined how online testing is used by educators and the relative

advantages and disadvantages of their methods, while others have proposed new or revised models for online assessment.

Brown and Abeywickrama (2010) find that designers creating tests intended for online use select only certain types of test items, such as those that require a closed-ended response from students, which has the added advantage of making the creation process more manageable. Boitshwarelo et al. (2017) also find that online tests tend to ask MCQs alongside other computer-friendly question types, such as true/false, pairing, and preset short answers. Although we do not recommend that test creators should stick entirely within these confines, as other types of assessment are possible, even online, the sort of tests described by Brown and Abeywickrama (2010) and Boitshwarelo et al. (2017) do have the advantage that teachers can quickly and easily create them, and they can be easily scored and administered to students.

As for the five assessment principles discussed in the previous section, in our view, practicality is not only the most important, but also the one which is most vulnerable to compromise during the transfer from traditional assessment to online assessment. As noted above, the principle of practicality encompasses the constructs of time and cost-effectiveness; as concerns the latter, the risk is that equal opportunity will not be afforded to all students undertaking online tuition or testing because some do not have access to the necessary equipment, such as computer and a good internet connection; hence, they may incur considerable expense in obtaining the necessary resources when university and public libraries are not available. This is even more the case in the context of ERT. If teachers are aware that their students do lack these technologies, they will be in the difficult position of being unable to administer tests or of having to find alternative means to include this group of disadvantaged students. Given the central importance of practicality, if this principle cannot be adhered to, it is of little use adhering to any of the others as if not all students can participate, then testing is, in effect, rendered meaningless.

Gipps (2005) and Means, Toyama, Murphy, Bakia, and Jones (2010) argue that online assessment can, in fact, offer advantages over face-to-face assessment: In terms of efficiency, both the assessment and the marking and feedback processes can be carried out more quickly and reliably, especially where large groups of learners are involved. Moreover, as Pifia (2013) points out, the need for manual input is eliminated when an LMS is used, reducing the time taken to generate tests and decreasing the likelihood of marker bias. However, previous studies on cheating have found that cheating can occur during online examinations because exam administrators cannot check whether students are looking up answers or conferring with each other via social networking apps, for example (Arnold, 2016; Harmon, Lambrinos, & Buffolino, 2010; Varble, 2014). Multiple studies have found that cheating is as likely to occur online as it is in face-to-face contexts (Grijalva, Nowell, & Kerkvliet, 2006; Burrus, McGoldrick, & Schuhmann, 2007; Watson & Sottile, 2010). In light of this, academic dishonesty has been highlighted and described as a significant concern during ERT (Guangul et al., 2020). Some researchers regard the following points as more serious disadvantages that may entail online testing is only suitable for assessing lower levels of understanding (McAllister & Guidice, 2012); the format offers little flexibility for educators who are seeking to test students of different needs, backgrounds, and levels (Stupans, 2006); and test creators may find themselves overly dependent on the lists of standard MCQs drawn up by educational publishers (Masters, Hulsmeyer, Pike,

Leichty, Miller, & Verst, 2001). In normal circumstances and with the current affordances and advances of technology, we would probably disagree with the first and second arguments because technology allows for a myriad of both testing methods and mediums that are suitable for different types of learners (for example, special needs learners can take online tests that have certain features such as audio clips that read aloud the choices and questions) while allowing for the creation of examinations and test items of all levels of difficulty. For example, computer-adaptive testing in high-stakes testing changes the difficulty of test items presented to the test taker depending on how well they perform on the test (see Chalhoub–Deville & Deville, 1999). However, because of the push to immediately shift to ERT, teachers might create tests in a rush in which both the level of difficulty and accessibility for different types of learners are compromised. In addition, teachers might succumb to incorporate previously published questions to relieve any stress and ensure they meet assessment deadlines.

The two basic types of test assessment are formative and summative. Of these, Meyen, Aust, Bui, and Isaacson (2002) and Boitshwarelo et al. (2017) suggest the first is more appropriate in the online context, in which they argue assessment of achievement should be an integral and continuous part of the teaching process. Moreover, it offers the most opportunity for feedback. However, Sewell, Firth, and Colvin (2010) find that both forms of assessment are appropriate and feasible for online learning. The former type aims more to provide an opportunity for feedback than to grade students' progress and to enable teachers to see how far students have progressed towards achieving the necessary outcomes (Sewell et al., 2010). Creators of formative tests can choose from a variety of formats according to their aim. For example, if an educator is seeking to evaluate affective learning, then classroom surveys and discussions will allow them to gauge the values, beliefs, reflectivity, and other affective attributes of the students, as well as gain feedback to improve the testing process itself. Formal summative assessments, on the other hand, are of value at the end of programs to check the progress made towards learning objectives, for example via graded examinations, essays, or quizzes. These assessments are usually regarded with greater trepidation by students because they offer a single chance to achieve a high grade. Hence, if an exam grade represents 20% upwards of an entire course grade, it is extremely vital to protect the integrity of administrators and students by carrying out the examination under the supervision of a proctor (Sewell et al., 2010). However, Black and Wiliam (2009) prefer not to make such distinctions between the two forms of assessment as summative assessments can be used as formative when the purposes of the latter are appropriately employed.

It seems that dealing with formative assessment under normal situations and ERT might be easier for both teachers and students since interaction exists amongst them while the assessment is being conducted. However, summative assessment during ERT should be approached with more caution. Students should not be disadvantaged due to the sudden move to online assessment. Offering training interventions on the format and instructions associated with summative assessments can be effective. In fact, some schools do provide their students with mock exams in the normal situation, if time allows, so adopting this type of approach to support the ERT is deemed more appropriate. Another plausible and more advisable method would be assessing students using multiple timed summative assessments rather than relying on one midterm exam and a further final exam with a 50% grade weight for each. Timed lengthy exams

might need advanced proctoring equipment on the side of both educational institutions and students, and that would be rather daunting money-wise.

Taking advantage of the various affordances of advanced technology in conducting different performance-based assessments is viable. Live presentations or speaking tests could be performed via video conferencing, for example. This will be further discussed in the following section.

Discussion

Any discussion of assessment within ERT must be prefaced by a discussion of how assessment can be implemented within the wider context of online learning. However, despite a recent surge of interest in remote learning, particularly within higher education, little research has been carried out into how learning outcomes and student performance can be assessed in this context (Kim, Smith, & Maeng, 2015). Moreover, there is a lack of clarity around which methods are best applied to assessing the skills and knowledge required more generally for digital literacy and employment in a digitally-driven workplace (Boitshwarelo, Reedy, & Billany, 2017). Perhaps due to the paucity of research, national governments, as well as the educational profession, have recently taken up the search for appropriate means of preparing students for life in the digital age, and evaluating their progress toward readiness to join a workforce characterised by uncertainty and change (Gupta & Ndahi, 2002; Kinash et al., 2015).

In this pandemic and given the particular lack of research into how online education can continue in emergency circumstances, educators have turned to technology to continue delivering their courses with the minimum delay and disruption. However, having done so, we find ourselves challenged by the need to devise new strategies of assessment in order to ensure that all our students, including those who do not have access to technologies, continue to achieve the necessary learning outcomes, so they all have an equal opportunity to progress to the next stage of their education or into the workforce (Cahapay, 2020).

Different studies adopted different approaches to the study of assessment in online learning. Some focused on academic integrity and whether it is compromised in the online learning, proctoring options (See Reedy et al., 2021), student achievement in the online mode as opposed to the traditional mode and students perceptions of online assessment (See Igaz, Afacan, 2020). While such studies provide insights on the gains of shifting online, its costs, place and time benefits as compared to traditional examinations, they fail to address the quality of the assessments used in the online mode and or ERT in terms of the widely-accepted assessment principles. Validity, reliability, washback, authenticity and practicality should be given a priority in assessment research as they represent the pillars of assessment design and evaluation whether in the traditional setting or the online and ERT mode.

Considerations to Consider for Implementing Online Assessments in ERT

The argument in the previous section enables us to consider the feasibility of carrying out summative and formative assessments in an ERT context. As noted in the first section of this paper, the construct of time is particularly vulnerable in this context, given that an emergency, by its very nature, requires a sudden and relatively unplanned shift from one mode of action to another - in this case, classroom to online learning. In these circumstances, educators are

attempting to translate planned assessments from one medium to another. Certainly, it is more feasible to modify existing assessments to fit the new environment than to begin the process of creating entirely new ones, as doing so will minimize delay in delivering teaching programs and alleviate stress for both students and educators. This last point is of particular importance given that stress levels will probably be higher than normal anyway due to the emergency, and that the overarching objective is to safeguard wellbeing and ensure the earliest continuation of education. Hence, going ahead with planned assessments such as multiple-choice exams, final projects, dissertations, papers, or presentations is recommended, alongside the use of online discussion boards and email consultations to provide a method of communication which, if a little ‘rough and ready’, will nonetheless allow formative assessment to continue.

Alternatively, educators could dispense entirely with the heavily weighed final summative assessment normally used at the end of a course. Given, as discussed above, that some students cannot access the necessary technology, it cannot be assumed that all students will be able to sit certain types of exams while ERT continues. Furthermore, there is no way in such conditions to implement the recommendation made by Sewell et al. (2010), that all such high-stakes examinations should be proctored. Instead, the one-off summative examination could be broken down into several components, such as quizzes, with the final grade being the sum of the grades for each component and the marking rubric adjusted accordingly.

Whatever assessment plan is formulated to deal with the conditions of ERT, it is essential that students must be kept informed, so they are not taken aback by a change in format, and that they have the time to prepare themselves, gather the necessary equipment, and make any arrangements, for example ensuring they will not be disturbed. Significantly, educators should not consider ERT as an opportunity to become overly creative around how they carry out an assessment: Students need as much continuity as possible, so adding new assessments, or creating several new types of assessment, is not advised.

When it comes to ERT and academic dishonesty, we first need to ask: What role and what form does academic integrity play? To answer this question, it is necessary for us to investigate how digital environments affect academic integrity. Reedy et al. (2021) state that it is crucial to maintain academic integrity through a

[r]econceptualization of digital exams and assessment that are fit for purpose in a digital age. This requires thoughtful and innovative assessment design and deployment, aligned to a teaching and learning approach to academic integrity [...] and a move away from outdated approaches to academic integrity that are not effective when translated to online contexts. (pp. 20-21)

It is important to note that adequately addressing academic integrity is not limited to employing technology, such as E-proctoring software. According to Stockwell (2020), technological tools do not represent a satisfactory solution to cheating during high-stakes examinations in invigilated online environments. Instead, developing exams that consist of test items developed by faculty could significantly reduce academic dishonesty by creating test items that are unique and authentic. However, Reedy et al. (2021) highlight how it can be very difficult to deter students from cheating during online tests if they have already developed an intention to cheat. Stockwell

(2020) emphasizes that, despite the technologies that have been developed to prevent students from cheating, they consistently find methods of overcoming the underlying mechanisms.

Parnther (2020) suggests that, specifically in the context of ERT, we consider the real meaning of academic integrity in relation to what ERT is doing to students in order to rethink what academic integrity is. For instance, there is a need to consider how the pandemic and the sudden switch to online learning have impacted students' concerns and anxieties. Parnther (2020) asserts that learning extends beyond preparing for examinations because it involves considering students' requirements outside the academic environment, for example, providing them with the skills they need to find a job. It is crucial these factors are taken into consideration to adequately address the factors that may lead to academic dishonesty. As Parnther suggests, it is essential to re-evaluate courses and learning outcomes, as well as effectively communicate with students during ERT and explain what is expected of them. The approach described here may be compassionate toward students, but it still does not address how academic integrity can be achieved during online assessment. Undoubtedly, external factors beyond the educational environment can impact how students perform and may motivate them to violate academic integrity standards; however, academic integrity represents a critical aspect of education and cannot be sacrificed, even during a difficult period such as ERT. As a result, it might be useless to move forward with the learning/teaching process since teachers would not be aware of their students' performance, hence jeopardizing their chances of achieving their goals. To address this issue, we need to educate students on the critical nature of academic integrity. Reedy et al. (2021) found that students who perceived academic integrity to be of critical importance were significantly less likely to cheat, regardless of the ease with which they could potentially bend the rules.

A final advice or consideration for designing assessments during ERT is that educators must remember to show understanding and flexibility, whatever type of assessment they decide to implement during ERT. In these exceptional circumstances, educators must demonstrate that they trust their students and accept that the justification offered by a student to explain, for example, a missed deadline or a low grade, is probably the truth, given the impossibility of proving otherwise. It is important to bear in mind that not all students have the technological means at home to keep up with their learning or participate in online group activities and may well ask to be given a second chance. Undoubtedly, this adds to the teacher's workload because she then has not only to create a second assessment, but also administer and mark it. To preclude this possibility, we recommend that teachers, where possible, set assignments based on open-ended, rather than closed-ended, questions, while ensuring that similar elements and difficulty levels are included that were present in the main assessment and aligning that with the learning objectives to ensure equality and equal opportunity for all students. An example could be setting a question which requires students to write an argumentative essay, which can be completed in their own time. However, this solution is not appropriate for certain subjects, such as assessing knowledge of math or grammar.

Conclusion

The aim of this study was to examine how assessment principles can be put in place and determined in emergency situations in which learning experiences are shifted from face-to-face to online modalities. It also reviewed some of the factors educators should take into

consideration when developing assessment methods. The above discussion suggests that the construct of time was one of the most at-risk factors when implementing assessments during crises due to the unpredictability of the events that unfold. As such, educators need to adopt a flexible approach to assessment and be open-minded in terms of the options that are pursued; for instance, multiple-choice questions or writing a final paper.

As we delve into the different approaches of assessment and methods of achieving assessment principles (Brown & Abeywickrama, 2010) under ERT, it is worth considering what decisions different educational institutions have taken in terms of assessment during ERT. In addition, questions of how and/or to what extent those decisions affected the various assessment principles and what teachers and students have learned from, and thought about, assessment during ERT are worth investigating. Future research should also consider the various strategies that transform paper-and-pencil assessments into online or computer-based assessment under ERT. A final learning point from the ERT experience is that it has proven to many educators, including those who were not keen on online teaching or had not considered such an approach, that teaching via this mean could be beneficial. In fact, there is a broad consensus that education may never return to the form in which it took prior to the pandemic (Zhu & Liu, 2020), and that technology will be increasingly employed in combination with face-to-face approaches via an integrated approach (Bellini, Pengel, Potena, Segantini, & ESOT, 2021; Goh & Sanders, 2020). It is necessary to recognize that the shift to online learning during ERT might have been temporary, but technology is still imperative to education, even though it cannot be relied upon exclusively. In addition, COVID-19 has had a significant impact on education, requiring that we consider possible development opportunities concerning the provision of open, free, and accessible resources to all students. Consequently, a special focus should be placed on developing digital solutions that facilitate the development of relevant content, effective teaching techniques, and learning environments that are supportive, in addition to ensuring that teachers are appropriately qualified and prepared.

About the Authors:

Dr. Wid Allehaiby is an Assistant Professor in the Department of English at King Saud bin Abdulaziz University for Health Sciences, Saudi Arabia. Her research merges the fields of applied linguistics and educational technology and focuses on the implementation of language instruction, pedagogical practices, and facilitation of learners' engagement using non-conventional learning tools, such as social networking sites. ORCID: <https://orcid.org/0000-0002-4320-7429>

Dr. Sara Al-Bahlani is an Assistant Professor and Head of the English Department at the University of Technology and Applied Sciences-Nizwa in Oman. Her research focused on language assessment and teacher assessment literacy, digital assessment literacy and integrating technology in the language classroom. She worked as an English, sociolinguistics and academic skills lecturer and course coordinator. ORCID: <https://orcid.org/0000-0002-9136-4576>

References

Arnold, I. J. M. (2016). Cheating at online formative tests: does it pay off? *The Internet and Higher Education*, 29, 98–106.

- Alkhazim, M. A. (2020). Distance education is more than Blackboard. *Al-Jazirah*, 5 May. Available at: <http://www.al-jazirah.com/2020/20200405/ln22.htm>
- Alqabbani, S., Almuwais, A., Benajiba, N., & Almoayad, F. (2020). Readiness towards emergency shifting to remote learning during COVID-19 pandemic among university instructors. *E- Learning and Digital Media*, 18(5), 460–479. <https://doi.org/10.1177/2042753020981651>.
- Al-Salman, S., & Haider, A. S. (2021). Jordanian university students' views on emergency online learning during COVID-19. *Online Learning*, 25(1), 286–302. <https://doi.org/10.24059/olj.v25i1.2470>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Bellini, M. I., Pengel, L., Potena, L., Segantini, L., & ESOT COVID19 Working Group. (2021). COVID19 and education: Restructuring after the pandemic. *Transplant International*, 34(2), 220-223. <https://doi.org/10.1111/tri.13788>
- Boitshwarelo, B., Reedy, A. K., & Billany, T. (2017). Envisioning the use of online tests in assessing twenty-first century learning: a literature review. *Research and Practice in Technology Enhanced Learning*, 12, 1-16.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). White Plains, NY: Pearson Longman
- Burrus, R. T., McGoldrick, K., & Schuhmann, P. W. (2007). Self-reports of student cheating: Does a definition of cheating matter?. *The Journal of Economic Education*, 38(1), 3-16.
- Cahapay, M. B. (2020). Rethinking education in the new normal post-COVID-19 era: A curriculum studies perspective. *Aquademia*, 4(2), 1-5. <https://doi.org/10.29333/aquademia/8315>
- Chalhoub-Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273-299.
- Crawford, J. et al. (2020). COVID-19: 20 countries' higher education intraperiod digital pedagogy responses. *Journal of Applied Learning & Teaching*, 3(1), 1–20.
- Gipps, C. V. (2005). What is the role for ICT based assessment in universities? *Studies in Higher Education*, 30(2), 171–180. <http://dx.doi.org/10.1080/03075070500043176>
- Goh, P. S., & Sandars, J. (2020). A vision of the use of technology in medical education after the COVID-19 pandemic. *MedEdPublish*, 9. <https://doi.org/10.15694/mep.2020.000049.1>
- Grijalva, T., Nowell, C., & Kerkvliet, J. (2006). Academic honesty and online courses. *College Student Journal*, 40(1), 180-185. Available at <http://www.projectinnovation.com/college-student-journal.html>
- Guangul, F. M., Suhail, A. H., Khalit, M. I., & Khidhir, B. A. (2020). Challenges of remote assessment in higher education in the context of COVID-19: A case study of Middle East College. *Educational Assessment, Evaluation and Accountability*, 32, 1-17. doi: 10.1007/s11092-020-09340-w.
- Gupta, A., & Ndahi, H. (2002). Meeting the digital literacy needs of growing workforce. *Reading Matrix: An International Online Journal*, 2(1), 1-10.
- Harmon, O., Lambrinos, J., & Buffolino, J. (2010). Assessment design and cheating risk in online instruction. *Online Journal of Distance Learning Administration*, 13(3). Available at http://www.westga.edu/~distance/ojdla/Fall133/harmon_lambrinos_buffolino133.html

- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*, 27. Available at <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- Huang, R. et al. (2020). Disrupted classes, undisrupted learning during COVID-19 outbreak in China: Application of open educational practices and resources. *Smart Learning Environments*, 7(1), 1-15. <https://doi.org/10.1186/s40561-020-00125-8>
- Huber, S. G., & Helm, C. (2020). COVID-19 and schooling: evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability*, 32(2), 237-270. <https://doi.org/10.1007/s11092-020-09322-y>
- Ilgaz, H., & Adanır, G. A. (2020). Providing online exams for online learners: Does it really matter for them?. *Education and Information Technologies*, 25(2), 1255-1269. <https://doi.org/10.1007/s10639-019-10020-6>
- Jerald, C. D. (2009). Defining a 21st century education. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.8011&rep=rep1&type=pdf>
- Kereluik, K., Mishra, P., Fahnoe, C., & Terry, L. (2013). What knowledge is of most worth: Teacher knowledge for 21st century learning. *Journal of digital learning in teacher education*, 29(4), 127-140.
- Kim, N., Smith, M. J., & Maeng, K. (2008). Assessment in online distance education: a comparison of three online program at a university. *Online Journal of Distance Learning Administration*, 11(1). Available at <https://www.westga.edu/~distance/ojdl/spring11/kim111.html>
- Kinash, S., Knight, D., & McLean, M. (2015). Does digital scholarship through online lectures affect student learning? *Journal of Educational Technology and Society*, 18(2), 129-139.
- Masters, J. et al. (2001). Assessment of multiple-choice questions in selected test banks accompanying test books used in nursing education. *Journal of Nursing Education*, 40(1), 25-32.
- McAllister, D., & Guidice, R. M. (2012). This is only a test: A machine-graded improvement to the multiple-choice and true-false examination. *Teaching in Higher Education*, 17(2), 193-207. <https://doi.org/10.1080/13562517.2011.611868>
- McAlpine, M. (2002). *Principles of assessment*. Luton: CAA Centre, University of Luton.
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research, and Evaluation*, 7(1). Available at <https://scholarworks.umass.edu/pare/vol7/iss1/8>
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.
- Meyen, E. L., Aust, R. J., Bui, Y. N., & Isaacson, R. (2002). Assessing and monitoring student progress in an e-learning personnel preparation environment. *Teacher education and special education*, 25(2), 187-198. Retrieved from <https://www.learntechlib.org/primary/p/24323/>
- MOE. (2020). (rep.). *The Saudi MOE Leading Efforts to Combat Coronavirus Pandemic COVID-19*. Retrieved from <https://iite.unesco.org/wp->

- content/uploads/2020/10/The-Saudi-MOE- Leading-Efforts-to-Combat-Coronavirus-Pandemic-COVID-19.pdf
- Parther, C. (2020, March 24). *Emergency Driven Remote Learning*. ICAI. Available at <https://academicintegrity.org/blog/146-emergency-driven-remote-learning?highlight=WyJwYXJudGhlciJd>.
- Pifia, A.A. (2013). Learning management systems: A look at the big picture. In Y. Kats (Ed.), *Learning management systems and instructional design: Best practices in online education*(pp. 1-19). Hershey: Idea Group Inc (IGI).
- Reedy, A., Pfitzner, D., Rook, L., & Ellis, L. (2021). Responding to the COVID-19 emergency: Student and academic staff perceptions of academic integrity in the transition to online exams at three Australian universities. *International Journal for Educational Integrity*, 17(1), 1-32.
- Rogers, C. F. (2006). Faculty perceptions about e-cheating during online testing. *Journal of Computing Sciences in Colleges*, 22(2), 206-212. Available at <https://dl.acm.org/citation.cfm?id=1181936>
- Sankey, M. (2020). COVID-19 Exam Software Survey – 2020. *Acode*, 1–9. Retrieved from https://www.acode.edu.au/pluginfile.php/8244/mod_resource/content/2/examsWhitepaper.pdf
- Sewell, J. P., Frith, K. H., & Colvin, M. M. (2010). Online assessment strategies: A primer. *MERLOT journal of online learning and teaching*, 6(1), 297-305.
- SPA. (2020, April 16). The Minister of Education directs schools and universities to progress their students to the next grade. Available at <https://www.spa.gov.sa/2075525>.
- Stockwell, S. (2020, June 24). *Tech experts want online exam monitoring canned*. Triple J Hack. <https://www.abc.net.au/triplej/programs/hack/online-exam-monitoring/12388630>.
- Stupans, I. (2006). Multiple choice questions: can they examine application of knowledge? *Pharmacy Education*, 6(1), 59–63.
- UN. (2020). (issue brief). *Policy Brief: Education during COVID-19 and beyond*. United Nations. Retrieved from https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf
- Varble, D. (2014). Reducing cheating opportunities in online test. *Atlantic Marketing Journal*, 3(3), 131-149.
- Watson, G., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? *Online Journal of Distance Learning Administration*, 13(1). Retrieved from <http://www.westga.edu/~distance/ojdl/spring131/watson131.html>
- Zhang, W., Wang, Y., Yang, L., & Wang, C. (2020). Suspending classes without stopping learning: China's education emergency management policy in the COVID-19 outbreak. *Journal of Risk and Financial Management*, 13(55), 1-6.
- Zhu, X., & Liu, J. (2020). Education in and after Covid-19: Immediate responses and long-term visions. *Postdigital Science and Education*, 2(3), 695–699.