AUTHOR:
Ms Tatiana Sango[1] iD
Mr Robert Prince[1] iD
Ms Sanet Steyn[1] iD
Mrs Precious Mudavanhu[1] iD

AFFILIATION:
[1]University of Cape Town, South Africa

High-stakes online assessments: A case study of National Benchmark Tests during COVID-19

Abstract

Owing to the COVID-19 pandemic, paper-based delivery of the National Benchmark Tests (NBTs) was not possible during the 2020 testing cycle. The NBTs, being a large-scale national assessment project, did not have alternative options, other than to offer the tests online. Moving these high-stakes tests online meant that certain considerations had to be considered to retain the credibility and security of the tests, without compromising the validity and reliability of the scores. Digitising the paper-based NBTs required an innovative, flexible and robust solution, which promotes fairness and ensures the quality of testing is maintained, while in many ways remains comparable to the paper-based implementation. To deliver the NBTs online, the following important considerations needed to be addressed: test security and integrity, test candidate identification processes, the prevention of dishonest behaviour, test scheduling and timing and technical support. The online testing solution chosen integrates the following aspects: it 1) enables all candidates to take the same test at the same time; 2) ensures the quality and similarity in experience of test delivery for all candidates as far as possible; 3) prevents candidates from accessing other applications and devices during the test; 4) enables proctoring before, during and after the tests to encourage appropriate behaviour similar to that expected during paper-based tests; 5) provides live support to assist candidates to deal with technical challenges and to guide them through the test sessions and 6) processes and presents data and scores in the same way as for the paper-based tests. In this article, we analyse the integration and complexity of the online NBTs solution, the opportunities and challenges associated with this form of delivery and reflect on test candidates' and the team's experiences. We discuss components of online assessment and wish to argue that this is also relevant to high-stakes course assessments. This case study should help to refine the scope of further research and development in the use of online high-stakes assessments.

Keywords: *National Benchmark Tests (NBTs); paper-based assessment; online assessment; Artificial Intelligence (AI); proctoring.*

1.   Introduction

1.1 The National Benchmark Tests Project

Designed with the aim of ensuring that the South African Higher Education (HE) sector understands and addresses the levels of academic readiness of entry-level students from diverse educational backgrounds, the National Benchmark Tests (NBTs) Project provides a national service to this sector using two instruments – the National Benchmark Test: Academic and Quantitative Literacy (NBTs AQL) assessment and the National Benchmark Test: Mathematics (NBTs MAT) assessment.

Since the official implementation of the NBTs in 2009, the tests had been administered to up to approximately 90 000 candidates annually in about 26 test sessions at test centres across South Africa. The assessments are administered in English and Afrikaans, the two mediums of instruction used at South African higher education institutions. The first assessment, the NBTs AQL, consists of two components – Academic Literacy and Quantitative Literacy – and is intended to be written by all NBT candidates. For the purposes of these tests, Academic Literacy refers to a candidate's "capacity to engage successfully with the demands of academic study in the medium of instruction within the context of higher education" (Cliff & Yeld, 2006) and Quantitative Literacy refers to the candidate's "ability to manage situations or solve problems of a quantitative nature in real contexts relevant to higher education" (Frith & Prince, 2006). The second instrument, the NBTs MAT assessment, is explicitly designed to probe higher education competencies (i.e., depth of understanding and knowledge) relevant to mathematically demanding disciplines within the context of the NSC[1] framework (CETAP, 2019: 14).

1.2 The impact of COVID-19 on the NBTs' assessment model

The coronavirus (COVID-19) pandemic has changed the educational landscape for most students, teachers and lecturers in 2020. The digital transformation that schools, colleges and universities have been slow to adopt, particularly in assessment modes and practices (Richardson & Clesham, 2021), have been accelerated dramatically during the pandemic, with the higher education sector seeking solutions to teaching, learning and assessment. Accelerated digital adoption across all sectors, including education, and higher education in particular, necessitated quick responses to the question of the mode of assessment. Many high-stakes examinations were postponed or cancelled entirely, and few formal in-course summative assessments were delivered face-to-face, in person or as sit-down assessments (The World Bank, 2020). In many cases assessments took the form of projects, take-home assessments, open-book assessments and other forms of remote assessments (Burnett & Fuentes, 2020). Being a high-stakes standardised assessment, the NBTs could not use these alternative assessment options easily, therefore, the only way was to offer the tests online.

One of the key factors that influenced the decisions the team made in terms of the development and administration of the online NBTs was to ensure that the test experience (or testing conditions) for all candidates using the online testing environment would be comparable regardless of where they were writing from, and that it would also be as close as possible to the experience in the paper-based administration of the tests. These aims relate to important categories of validity evidence that form part of the NBTs' quality assurance process. Considerations such as the fairness of the assessment to different

---

1   The National Senior Certificate (NSC) examinations conclude twelve years of schooling in South Africa.

types of students, and the comparability of the thinking processes and skills needed to complete the assessment tasks, are crucial (Brookhart & Nitko, 2019; Nitko, 2001). Introducing a new mode of delivery complicates this, because there are not only clear differences between the modes of delivery, but several additional potential issues could compromise the test experience. Some of the challenges we faced were loadshedding , unstable internet connectivity, candidates who do not have access to devices that meet the minimum requirements that would allow the testing platform to function and enable proctoring, and administrators who had to have the necessary skills and permissions to make changes on the testing platform when troubleshooting issues. Candidates' level of computer literacy could further complicate test outcomes because it could introduce construct-irrelevant variance (Wise, 2019). A lower level of computer literacy may be a deterrent to taking the online NBTs – if a student is not confident in their own computer skills, the student is less likely to be comfortable with online assessments. Even students who are comfortable using computers, need to become used to the testing environment before using it (Khan & Khan, 2019).

We have chosen a qualitative approach for this case study and analysed experiences by comparing two modes of assessment, paper-based and online test administration, with reference to processes, milestones, advantages and challenges. With the introduction of the online mode of delivery during the pandemic, standard processes underwent rapid redesign to achieve the intended outcomes. Following the international practices, the local South African challenges were considered, and we reflect upon these in this article. The candidates' experiences were collected and documented in the post-test surveys and live chats data were screened for overall understanding of the major challenges. As the paper-based and online modes are envisaged to continue to be delivered, some elements of the quantitative data, such as test data and candidates' performance analysis, were also included in this article from the initial implementation round of online testing.

2. Literature overview

2.1 The use of standardised tests and high-stakes assessments

High-stakes examinations are mostly used for selection, progression and certification purposes, but they also play an important role in supporting policy making, accountability purposes and the making of strategic decisions that influence teaching and learning in classrooms (Black *et al.*, 2012). Standardised tests are used for placement and diagnostic decisions, and for providing feedback to parents and students. By virtue of these tests administration procedures and scoring using a nationally representative sample, the test scores are comparable which allows for important high-stakes decisions to be made in relation to the candidates' abilities (Kubiszyn & Borich, 2013). As a result, these tests are widely used for university admissions and placements of prospective students. For example, in the United States, the required standardised tests for undergraduate admissions usually include Scholastic Assessment Test (SAT), Test of English as a Foreign Language (TOEFL) and American College Testing (ACT), which play a significant role in the admissions decisions due to the fact that these tests are found to be the best predictor of a student's success (Leonard, 2020). Course specific admissions and aptitude tests are commonly used across Europe in addition to language proficiency examinations, such as the International English Language Testing System (IELTS) or TOEFL (McGrath *et al.*, 2014).

2.2 The use of NBTs

Institutions use the results from the NBTs for four main purposes: to "1) assess entry-level academic and quantitative literacy and mathematics proficiency of students; 2) assess the relationship between entry level proficiencies and school-level exit outcomes; 3) provide additional information in the admission and placement of students; and 4) inform the nature of foundation courses and curriculum responsiveness" (Griesel, 2006: 4). Consequently, there is a growing body of research that stems from these different uses. Some have investigated student performance in the three domains and variables that may influence and/or explain these trends (Prince & Frith, 2020) or compared NBT results to performance on exit-level examinations, such as the National Senior Certificate (Prince *et al.,* 2021). Other studies have looked at the use of these results for placement and selection for particular programmes and fields of studies (Mabizela & George, 2020) or how teaching and learning practices should respond to the performance trends that have been observed (Cliff, 2015; Nel, 2020; Steyn *et al.,* 2020).

2.3 The impact of COVID-19 on standardised tests and high-stakes assessments internationally

The review made by Magno (2020) for The Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP) highlights the most common adjustments that were made during the pandemic in the conduct of large-scale assessments. These adjustments are grouped into four broad categories: 1) scheduling arrangements, 2) modality of the assessments, 3) safety measures and 4) content modifications. The scheduling of assessment is often the first option countries are faced with, considering whether to postpone and reschedule, or to cancel or suspend the exams. The choice behind changing the modality of assessments and going online is largely based on the capability of the countries in terms of their internet infrastructure, as well as the students' access to devices and availability of the technology solution to support the assessment delivery. Magno (2020) reports on China, Pakistan and Thailand, who were able to conduct online and offline high-stakes national examinations. While in India and Indonesia, where high stakes examinations were completely cancelled, alternative measures for promotion were used, based on cumulative grades or other criteria. In the early months of the pandemic, some countries were able to conduct national high-stakes assessments under strict safety measures, in addition to decentralised seating arrangements and capacity considerations. Magno (2020) also reports on examples where the content modification strategies were used for the national exam, ensuring that specific sections of the curriculum are not included.

Comprehensive global reviews were collated by UNESCO (2020a) in the working document summarising national coping strategies on high-stakes examinations and assessments in the education sector. However, what is evident is that in the context of a global pandemic, the widespread reliance on paper-based administration for high-stakes examinations is problematic and we need to continue to seek alternative ways of test delivery (Richardson & Clesham, 2021).

2.4 Digitising assessments and online assessment administration challenges

Transitioning to online assessment design and administration has several challenges (Aisbitt & Sangster, 2005; Frankl & Schratt-Bitter, 2012; Cramp *et al.*, 2019; Moss, 2020; Alsadoon,

2021). In particular, Aisbitt (2005) describes the design process and implementation problems encountered during the first year of use of online assessment, making a strong case for appropriate planning, strong project management, competent and adequate technical support, careful choice of software and compatibility checks to avoid implementation issues. Frankl (2012) discusses students' positive attitudes towards online assessment and yet, the biggest category of obstacles to successful implementation is related to technical issues, suggesting that technical support is provided for students throughout an examination. In the study of remotely invigilated exams by Cramp *et al.* (2019), the authors, emphasise the need to simulate the exam experiences prior to the actual exam day and provide students with real-time responsive technical support to ensure successful implementation. In the work by Moss (2020), these factors are also supported to mitigate cognitive load and reduce stress. Similarly, Alsadoon (2021) reports on the challenges experienced by the administrators of online assessments, specifically making a case for developing technical skills for instructors and students to overcome many of the challenges.

In the policy recommendations for online assessments as a choice, UNESCO (2020b) stresses the need for thoroughly examining access to infrastructure, connectivity, protection of personal data, security, integrity and online proctoring methods, as well as students' digital skills and gaps. Ngqondi, Maoneke and Mauwa (2021) proposed a secure online exams conceptual framework for the South African HE context, also outlining the need for policies development, design of processes and establishing the required ICT infrastructures, prior to the actual development or adoption of the online exam system.

2.5 The need for test security and integrity

Test security and integrity form an important part of the design scope for the assessment administration processes, regardless of the modality. Some risks associated with paper-based test delivery are removed in the online mode, but new risks are introduced and will need to be carefully mitigated. Prior to the COVID-19 pandemic, these risks (cyber security breaches, personal data security, data loss) and concerns about equity and fairness for all students (test scores validity, malpractice) often contributed to the slow adoption of technology in high-stakes assessments (Coombe, Lester & Moores, 2020). Specifically, Coombe (2020) suggests that malpractice causes unfairness, requiring the new structures to be created in the online mode of administration to replicate the stringent security measures that exist in the paper-based administration. Ngqondi *et al.* (2021) give a comprehensive review of current international practices of online exams security features and components, arguing for an integrated approach to include student identification, authentication and continuous monitoring of the exam sessions while offering students support and the necessary guidance.

3. Conceptual framework

The NBTs, being both high-stakes assessments and standardised instruments, need to adhere to strict requirements grounded in the principles of educational measurement (Brookhart & Nitko, 2019; Nitko, 2001). For these instruments to achieve the intended purpose and be effective, the high-stakes tests must be carefully constructed, administered and scored using statistical tools to determine the degree of confidence that can be placed in the results; their usefulness depends on whether they are used and interpreted correctly (Kubiszyn & Borich, 2013). Design decisions, whether part of the initial test development or in response to significant changes in the testing landscape – such as the shift from paper-based to online

administration – must be informed by the criteria we use to evaluate the instrument and its use. Brookhart and Nitko (2019) identify eight broader categories of evidence that are used in the process of validating a test instrument and should undergird the test development process.
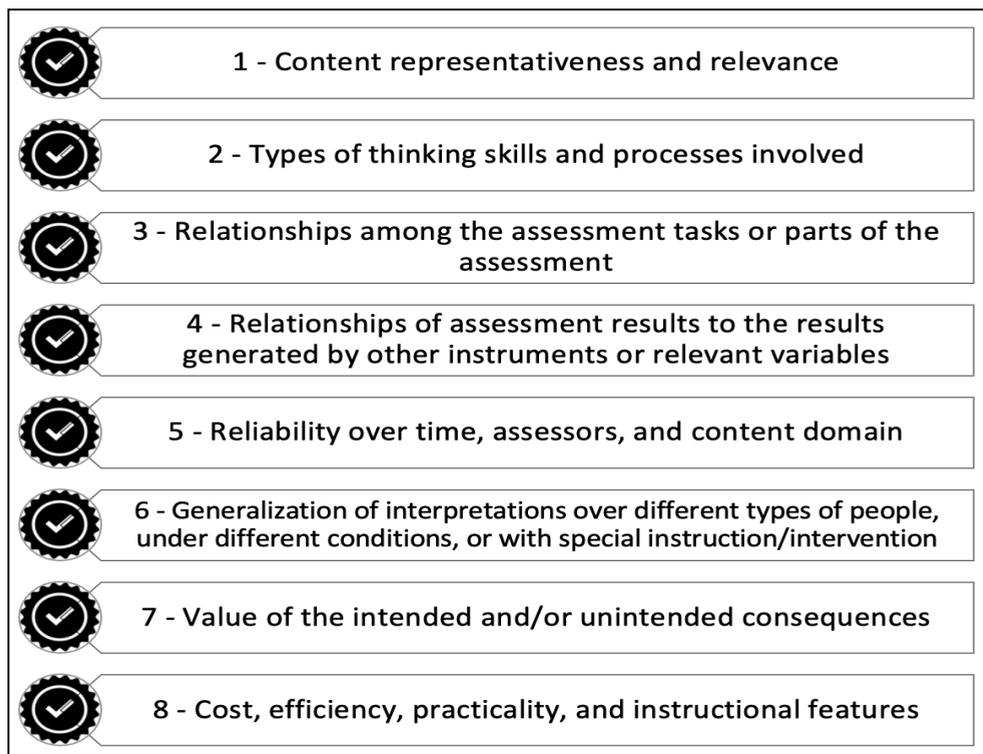
Figure 1: The eight broader categories of validity evidence (Brookhart & Nitko, 2019)

Although all these categories were revisited during the process of implementing the online mode of delivery for the NBTs, some are particularly relevant to the case study presented in this article. For example, ensuring a comparable test experience across the different sessions and modes of delivery – an issue of fairness and a necessity in the administration of high-stakes standardised tests such as the NBTs – would involve the careful consideration of content representativeness (1) in the new test format, as well as the types of thinking skills and processes (2) that are involved in completing the test items in this new testing environment. Moreover, in addition to studying the reliability (5) of the instruments across administrations and the effects the change in mode might have on the logistics of administration and reporting (8), it is important to investigate the possible variations in testing conditions and their implications (6), and both the intended and unintended consequences that may be associated with using this new environment (7).

In assessment, and in particular high-stakes assessment, the integration of technology should not be seen as an easy and quick solution. Magno (2020) emphasises the importance of observing the issues of reliability and validity requirements while moving high stakes examinations online and lists five crucial factors by Luna-Bazaldua *et al.* (2020) that need to be considered when moving to an online format: 1) access to adequate device and internet connection; 2) possibility of software malfunctions; 3) possible legal challenges involved in

using remote proctoring; 4) maintaining psychometric standards of online exams when using test items previously administered in paper format; and 5) the need for high stakes exams to adhere to the principles of universal design to observe fairness towards students.

Taking these considerations and requirements as the conceptual framework for this article, we will discuss decisions that were made in the process of adapting these tests for online delivery and will compare their design and administration to their paper-based counterparts. We will do so to highlight what we have been able to accomplish to date, organise our observations and frame the need for further improvement.

4. Online assessment delivery model

### 4.1 Objectives of the NBTs' online assessment model

Taking exams online raises, inter alia, concerns about access, students' digital proficiency, academic dishonesty, integrity and the quality of online assessments. In addition to these concerns, students' data protection and privacy rights, as well as the reliability of data as evidence of students' learning, are important factors to consider when moving high-stakes assessments from paper-based to the online delivery mode. Digitising the paper-based assessment and adapting it for online delivery in the NBTs' context, required an innovative, flexible and robust solution to be comparable to the paper-based implementation. The choice of the online assessment solution for the NBTs was not only based on the ability to create and administer tests via digital devices, such as desktop computers and laptops, but also had to meet the objectives below.

The assessment solution had to:

- enable all candidates to take the same test at the same time;

- ensure the quality and similarity in experience of test delivery for all candidates as far as possible;

- prevent candidates from accessing other applications and devices during the test;

- enable proctoring before, during and after the tests to encourage appropriate behaviour similar to that expected during paper-based tests;

- provide live support to assist candidates to deal with technical challenges and to guide them through the test sessions; and

- process and present data and scores in the same way as for the paper-based tests.

4.2 Features of the model and the choice of digital provider(s)

Web-based assessment solutions are often preferred to stand-alone solutions to avoid the installation issues and complexities around data collection. However, even in a web-based environment, the installation of additional software to provide the needed security levels is often inevitable. The NBTs' online testing solution integrated the following components: test registration, test management and test delivery learning management system (LMS), LockDown Browser (LDB), live support (online and offline), pre-test, in-test and post-test proctoring, data analysis and reporting. Figure 2 below visually depicts this integration.
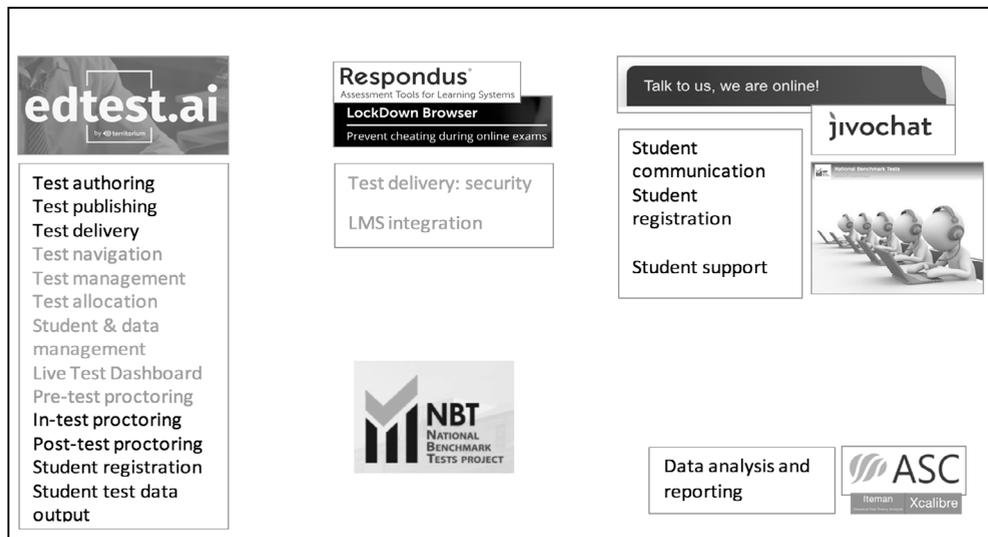
Figure 2: Components of the NBTs' online solution

The choice of a technology provider had to be made quickly in the pandemic-driven context, and we felt that referrals from educational partners in assessment will minimise the risks. A Mexican-based company, Territorium Life, offered a competitively priced digital solution and services, including support and training for staff, and was prepared to work with the team to customise the platform for the specific needs of the NBTs. Territorium's *edtest.ai* platform (https://www.territorium.com/edtest-ai) had been introduced as a secure, versatile and reliable assessment platform with artificial intelligence (AI). The platform was redesigned to handle candidates' registration, the facial recognition and identification process, test scheduling, test management and the delivery of online synchronous sessions in a way that closely matched the objectives of the NBTs.

The paper-based delivery model of the NBTs was redesigned using the available features of AI technology to help administer tests in a secure way. Monitoring testing sessions is made possible by proctoring before, during and after the tests to encourage candidates' appropriate behaviour, similar to that expected when paper-based tests are written under invigilation. The Respondus LDB software, a custom browser that locks down the testing environment within an LMS, is used to prevent any unauthorised browsing on the candidates' devices during the test session. The solution also integrates a live chat feature that provides communication between the support team and the candidates. The chat software (JivoChat and later Chatra) makes it possible for candidates to ask questions about technical issues, test-related matters and to make general enquiries about the test-taking process.

Candidates' response data is collected automatically, stored on Microsoft Azure cloud-based servers hosted locally in South Africa, retrieved, analysed and scored in line with the requirements of the paper-based NBTs. The Iteman and Xcalibre programs developed by Assessment Systems Corporation (ASC) are used to produce extensive item- and test-level output regarding the examinees' scores, to address important psychometric issues of the tests, and to ensure reviewing the NBTs in terms of item and test reliability so that the scores are defensible.

4.3 Test-taking and administration

For the first implementation cycle, four MAT tests (60 items each), four AQL tests (125 items each) and two sample tests were authored and published in an online environment. The NBTs, both AQL and MAT, consist of multiple-choice questions with four options of which one is correct, and each paper has a 3-hour time limit. In addition to this, the AQL paper-based test consists of seven timed sections of 25 to 30 minutes each.

4.4 NBTs candidates

In the traditional paper-based administration of the NBTs, candidates register on the NBTs website, book their test date and follow the prompts to complete the process. At the end of registration, candidates receive the venue details and their NBTs reference number against which to make payment. Candidates who take NBTs online complete a similar test registration process, but in addition also have to create a personal profile on the NBTs' online test delivery portal, the "venue", which is not hosted on the website. This process involves navigating to the testing portal, logging in with their NBTs reference number, accepting the standard declaration (results data use and session recordings), capturing their photo, completing ID verification and completing the pre-test simulation. In the paper-based and online-delivery modes, candidates log onto the NBT's website two to four weeks after the test to access and download their scores and personal reports.

In the paper-based administration, the candidates arrive at the venue with their official identification document (ID), which is checked before they are given an answer sheet with their name and ID number printed on the back. They are then given instructions by the chief invigilator on how to complete the answer sheet and make corrections using pencils and erasers. Candidates may not use anything else except pencils and erasers during the test, and their sessions are supervised by invigilators. On the day of the online NBTs session, candidates take the scheduled test at home or in a computer lab, using laptops or desktop computers with cameras and microphones that they have already set up as part of the registration process. Invigilation is replaced by a proctored environment where candidates access proctors or support team members through the platform's chat function to resolve technical and other test-related issues.

The more involved digital nature of the tasks in the online environment requires a different skill set of candidates compared to the paper-based assessment, and thus necessitates a different level of support and engagement by the team designing these experiences. For example, the pre-test simulations were designed so that candidates can have the full experience of taking the online test and successfully download the required software to prevent any unexpected problems on the test day.

4.5 NBTs team

In the traditional paper-based administration of the NBTs, the team assumes different roles and is expected to support various stages of test delivery: test development and research, test printing, operations and logistics, as well as data management and administration. With a major shift in the project implementation processes, many team roles and daily tasks needed to be redesigned to support the new way of working. The individuals' transition required by the shift to online assessment necessitated steep learning to adapt to the rapid redesign of processes, the use of new technology and the shift in normal day-to-day roles and responsibilities. In addition to this, the engagement with the technology solution provider to

create the online testing environment hinges on the ability to translate the conventional tasks of the paper-based delivery into new ways of digital implementation. For example, preparing the tests for online delivery involved critical engagement with the design of the user interface (UI), as well as the test navigation and instructional design, prior to the actual digitisation and authoring of the test items. The latter also involved a certain level of technical editing expertise, especially in the case of the mathematics items. These skills are predominantly found in the job descriptions of instructional designers and digital editors. Another skill set involved creating detailed instruction manuals and other communication instruments to explain the online test procedures, and to give test navigation support during the live sessions.

To accelerate the acquisition of new skills and to promote understanding of the new ways of test implementation, many staff training sessions were designed. In addition to proctor training, everyone went through the test simulation process, downloading the software, going through the photo and ID verification, and taking the sample test as a candidate. The ability to provide support requires a deep understanding of candidates' experiences and possible difficulties.

5. Discussion and analysis of experiences

5.1 Access

When looking at educational solutions, one of the most important concepts is access, which is seen as students' ability to gain the necessary learning and assessment opportunities. There are two aspects to access in terms of higher education that is pertinent to teaching and learning and assessment opportunities. The first is physical access and the second is epistemological access (Morrow, 2009). In terms of the paper-based and online modes of assessments there are challenges of physical access that need to be overcome. However, the challenges are different for the two modes of assessment.

Complex online teaching and assessment solutions generally require stable internet connectivity to deliver the desired learning and assessment experiences (Luna-Bazaldua, 2020). The online high-stakes examinations are no exception, with it being even more important not to lose connectivity during the session, which could result in losing data in the middle of the test. During the live testing sessions, losing connectivity creates issues with being able to pick up where one has left off, potentially also leading to a loss of focus and increased stress levels, which could impact the test performance of the candidate. Some candidates also reported issues with not being able to see images, graphs and mathematical symbols correctly when logged into the test. For example, candidates would see and report "Math processing error" messages being displayed, as shown in Figure 3 below, or would report that "the maths is not showing how I am used to it", referring to traces of LaTeX code that did not fully render. This was mostly due to slower internet connections, and in some cases, candidates had to be advised to reschedule their test sessions and seek alternative connectivity solutions.
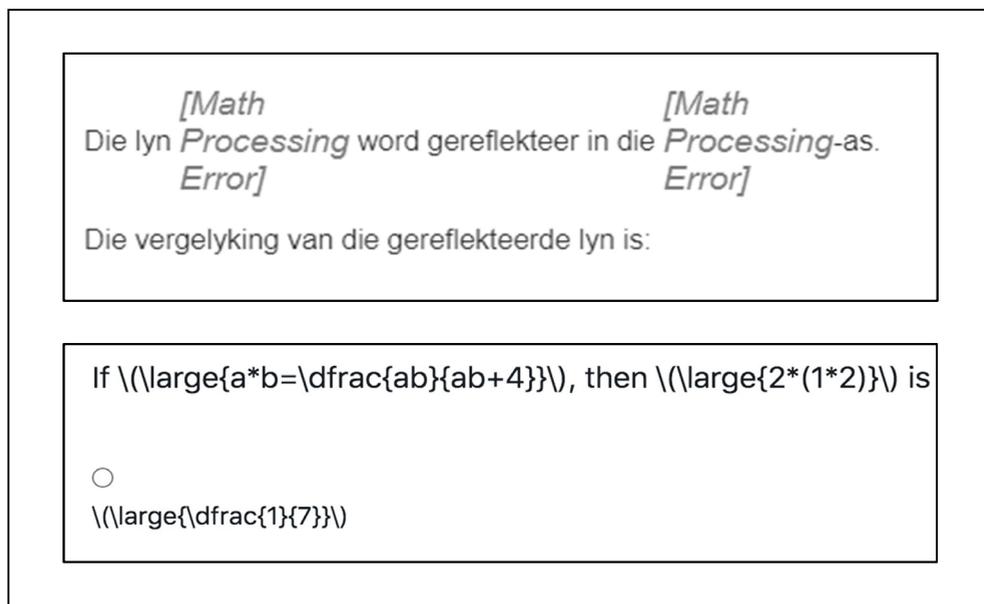
Figure 3: Examples of maths rendering errors

The NBTs' online solution requires 512 kb/s (recommended 1 MB/s) bandwidth or internet line speed. The additional test security features, such as downloading the LockDown Browser and maintaining the proctored sessions' video streams, incur additional data costs. Candidates were advised to have approximately 500MB of data to complete each test. In South Africa, data costs remain relatively high, with the cost of data varying "from R149 per one GB to R99 with effect from 1 April 2020" (Chinembiri, 2020), and although internet penetration in South Africa "increased by 1,7 million (+4,5%) between 2020 and 2021", it "stood at 64,0% in January 2021" (Kemp, 2021).

Some of the test sessions happened during periods of national power shortages, when the Electricity Supply Commission (Eskom) moves the country's energy supply to a reduced level and introduces rolling blackouts based on a rotating schedule ("loadshedding"). This problem was beyond our control and, again, the best solution we could offer affected candidates was to reschedule their NBT session.

Digital education assumes students' ability to use digital technologies and tools to navigate the learning and assessment set-up process. This ability hinges on access to or the ownership of digital devices with the required technical specifications. Shared devices present challenges with administrator rights, device-specific security features (firewalls, access to video and voice, etc.), which prevented some candidates from downloading additional software and installing it correctly. The type of devices required to enable the delivery of the secure assessment is also limited: smartphones and tablets are currently not supported, which becomes a limiting factor for those candidates who do not personally own or have access to a laptop or a desktop computer. Other challenges reported by some candidates, which affected test presentation, relate to the technical specifications of equipment, software versions and licensing, and screen sizes and resolution.

For the online assessment, candidates' access to devices with the required technical specifications, high data costs, unstable internet connectivity and/or electricity or power supply, as well as software licensing challenges need to be overcome. These concerns are related to the potential so-called hidden costs involved in getting access to appropriate devices or additional costs that may arise when technical issues are experienced and a candidate's test session is prolonged to try to resolve those issues. We must be aware of the impact these costs may have on a candidate's ability to complete an online version of the NBTs under standardised conditions.

5.2 Accommodating students' capacity

Accommodating students' capacity, in the context of high-stakes assessment, means to design assessment experiences with students' abilities, skills and expertise in mind (Nusche, 2013: 139). Whether the test is paper-based or online, high-stakes assessment experiences lead to unintended stress and increased cognitive load (Cramp *et al*., 2019). Test candidates are preoccupied with the test session arrangements and might be stressed about the content of the test and the consequences of not doing well in the test, seeing the outcome of their performance as a barrier to higher education. It is important to note that while the school-leaving National Senior Certificate (NSC) is a statutory requirement for entry into higher education, the NBTs are optional.

The paper-based examination format is familiar to the majority of high school learners in South Africa (Ngqondi *et al.,* 2021), whereas online learning and assessment is a novelty in many cases, even though this mode of learning and assessment became a necessity during the pandemic. Given this situation, it could be expected that the introduction of a different, unfamiliar mode of assessment, increases stress and affects candidates' cognitive load (Prisacari & Danielson, 2017). To minimise or reduce this impact requires student-centric high-stakes assessment design. The experience needs to be accommodating and supportive at all stages of the candidates' engagement.

"Usability means quality and puts the users and their real needs in the center" (Zaharias, 2004). The usability or quality of the online assessment solution can be evaluated by how well the solution fits the purpose, how easy it is to navigate and use it, and whether the engagement with it is effective. When designing online experiences, usability factors should be carefully considered to prevent the introduction of construct-irrelevant variance when taking the test online, minimising the need for computer experience during the test session. The main goal was to simplify the UI, increase the quality of the on-screen presentation and to make navigation through the test predictable. To avoid excessive scrolling, we decided to use an interface that shows only one question per page and repeats the vignette or source material where necessary, as shown in Figure 4 below. The questions and the answer options are not randomised in the online tests, preserving the exact structure of the paper-based tests.

Figure 4: Examples of the UI and item layout

It is important to note that because additional security features are needed to be activated and stay active during the entire online session, candidates still need a certain level of technical proficiency or computer literacy outside of the test-taking environment. For example, candidates need to enable their internet browsers to take photos and to bypass certain built-in security features in order to get access to the testing environment, as shown in the example in Figure 5 below. These factors can negatively influence candidates' performance in the test. For this reason, the simulation process became compulsory after the first testing session – its importance cannot be overemphasised. Having dealt with all the necessary settings on their devices a few days before the test, and having experienced the testing environment, is important in reducing candidates' cognitive load (Cramp *et al.*, 2019; Moss, 2020). It also normalises their stress levels, ultimately positively impacting test performance.
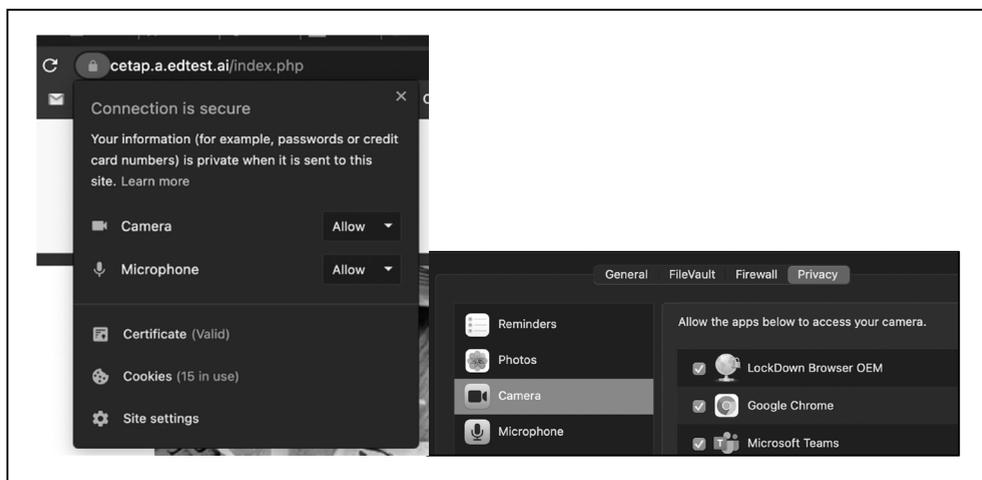
Figure 5: Example of camera activation for in-browser use

Live support is another essential aspect of the online assessment (Frankl, 2012; Cramp *et al.*, 2019). Different levels of support were identified during the online test delivery: technical support related to devices (integration with the platform), support with processes and test navigation, and special needs assistance. We anticipated that having comprehensive instruction guides, YouTube videos and FAQs, the team will be less involved in live support activities. However, candidates' ability to follow the process by reading the instruction guides, navigating multiple channels of communication and dealing with technical challenges was overestimated. Most candidates who were able to log into the system, were not shy to ask for help via the chat feature and had a high level of expectation to be assisted in real time. The number of conversations in which each agent participated ranged between 1 700 and 2 800 chats during the first two sessions and some chats involved extensive engagement, as can be seen in the example in Figure 6 below, where candidates' wait time was recorded at 13 min 33 seconds, followed by 55 messages in almost an hour.

Figure 6: Example of live support engagement with a candidate

In addition to this, another important factor is candidates' ability to communicate their challenges, by using clear language or creating visual references to support their query. The team needs to be aware of the potential challenges in communication and patient enough to elicit comprehensive responses to understand the issue a candidate is experiencing. This engagement during large live sessions could become time-consuming and challenging due to human capacity, leading to candidates' increased frustration and inability to write the test.

5.3 Secure test delivery and proctoring

In the paper-based examinations, invigilators are appointed to administer the tests and to supervise the candidates at the physical test venue. The invigilators are responsible for ensuring a rigorous test-taking environment, in which candidates are following required procedures and do not engage in any misconduct. In the case of online proctoring, the

invigilation process is essentially the same, where some of the features are automated and some are still controlled in real time by a human. The NBTs' online proctoring uses the model where online proctors are available before the exam, while candidates are taking their photos, and during the exam. The proctors are assisted by the AI technology in response to suspected violations, videos and audios are recorded to support identification of such violations.

To alleviate concerns about candidates' privacy and data protection, the NBTs team chose not to outsource the proctoring task, handling all the elements of the process within the team. NBTs proctoring team was relatively small and during the active test sessions, proctors were mostly focused on assisting candidates when the AI system automatically raised flags and alerted them of a potential problem. Many of these alerts had something to do with movements away from the camera, low lighting in the room or the detection of foreign objects. A proctor receiving a message from a candidate was able to check the problem by reviewing photos or video clips uploaded to the dashboard. Not all alerts meant that there was a violation; sometimes a picture frame on the wall or a door frame behind the candidate was identified and flagged as a foreign object; or when a candidate looked down to work out the answers to Maths questions, the candidate was flagged by the system as "not detected".

Even though proctoring has attracted negative publicity over the years, becoming the focus of discussions during the COVID-19 pandemic (SPARC, 2021), the NBTs did not receive many complaints from candidates or their parents in this regard. Instead, some concerns were raised about the additional software downloads (Respondus LDB), where candidates were unsure of the impact of the modifications it might have made on their devices.

From the test delivery and logistical issues, we can highlight the potential cost effectiveness of online sessions in a proctored environment, eliminating the need to have physical testing centres and removing travelling costs for invigilators and candidates. Having said that, the human capacity to support the online testing sessions remains a concern. Most of the challenges with the online NBTs were related to human capacity in a large-scale implementation model – not only to support candidates during the test sessions, but also to engage with the sessions' post-test data when decisions need to be made about the validity of each individual's test session based on AI-reported violations.

5.4 Candidates' data: online versus paper-based

There were four sessions under review for this case study. More than 6 000 candidates completed the pre-test online simulations, 7 000 candidates completed the NBT MAT test online and 8 000 candidates completed the NBT AQL test online in the first six months of implementation. All the candidates who faced challenges during each session have been encouraged to rewrite in the next available session. During the first session, almost 16 000 potential candidates registered to sit for the NBTs. These registrations were made before the decision was made by universities not to include NBTs as part of their 2021 admission criteria. This indicates that although the tests were subsequently made optional by many universities in South Africa, more than 16 000 candidates were willing to write the NBTs online when this mode of delivery was introduced. Among those who decided to write, at least 4 000 completed the pre-test online simulations and more than 4 000 candidates successfully completed the online NBTs for AQL and MAT and obtained scores for their tests. Tables 1–6 show the breakdown of candidates in each session.

Table 1:   Number of NBT registrations from the website

Session	NBTs registrations
July 2020	16 082
AQL only	2 586
AQL & MAT	13 496
August 2020	6 991
AQL only	1 461
AQL & MAT	5 530
October 2020	1 194
AQL only	337
AQL & MAT	857
Grand Total	24 267

In comparison, during the pre-pandemic years of paper-based test implementation, the NBTs were written by nearly 80 000 candidates annually, as shown in Table 2. In particular, the registrations for the first session in each testing cycle indicate that the number of online registrations is in line with the previous years of paper-based test registrations, as shown in Table 3.

Table 2:   Number of NBTs candidates – paper-based test

	English		Afrikaans	
Year	AQL	MAT	AQL	MAT
2017	76 408	56 160	6 706	4 957
2018	78 559	58 223	6 465	4 825
2019	69 278	53 617	6 291	4 796

Table 3:   Number of NBTs bookings between 2017–2019

Date	Booked seats
2017/05/20	22 380
2018/05/26	24 513
2019/05/11	14 688
Grand Total	61 581

Table 4 shows that there were more than 6 000 candidates who successfully completed the pre-test simulations. This is more than 70% of the candidates who completed the NBTs successfully and suggests the importance of the pre-test simulation in preparation for the actual test.

Table 4: Number of NBTs candidates who did online simulations

Language	July	August	October	Total
SAM01AFR	3 666	1 961	478	5 627
SAM01ENG	466	293	39	759
Total	4 132	2 254	517	6 386

Tables 5 and 6 show the number of candidates who received their scores in the period under review.

**Table 5:** Number of successful AQL online candidates during the 2021 intake cycle

Language	July	August	October	Total
English	4 357	2 522	574	7 476
Afrikaans	565	338	67	970
Total	4 922	2 860	641	8 446

Table 6: Number of successful MAT online candidates during the 2021 intake cycle

Language	July	August	October	Total
English	3 840	1 922	404	6 196
Afrikaans	500	287	41	828
Total	4 340	2 209	445	7 024

While the test data management in the online test delivery model remained the same as in the paper-based model, the data processing in preparation for scoring are slightly different for the two models. Nevertheless, the response matrices from both platforms are subjected to the same scoring process, which uses the three-parameter Item Response Theory (IRT). This is done using Xcalibre software (Guyer & Thompson, 2013).

Looking at descriptive statistics for tests that were written online and as paper-based tests, the preliminary results show better candidate performance for those who took the NBTs online. Figure 7 shows that there is a similar pattern of candidate performance in all the tests for the online and paper-based models. Candidates performed better in the AL test compared to both the QL and MAT, and MAT has the lowest performance in the online and paper-based models.
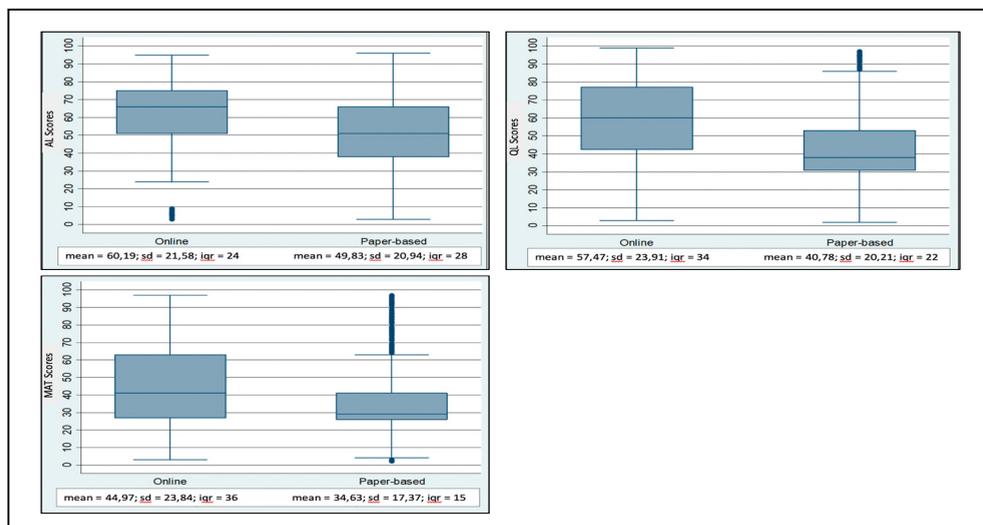
Figure 7: Descriptive statistical analysis: online versus paper-based

We notice higher mean and median scores for all online tests, a bigger interquartile range and standard deviation for the online MAT and QL, while for AL these measures of dispersion remain pretty much the same for the online and paper-based tests. However, this data needs to be interpreted with caution, because the differences might not be due to a change in the mode of delivery, but may rather be due to the demographic of candidates or the timing of the test administration in the context of the pandemic. Further investigations will be made and studied in the next few months to not only compare the data, but to also investigate the performance at item level.

6. Conclusions

In conclusion, we think that with careful design grounded in methodology, it is possible to successfully implement online tests in high-stakes examinations. However, different solutions are appropriate for different contexts. This article has reported on the first year of implementation for the online delivery of the NBTs. As the year progressed, we have made many adjustments to meet the unanticipated challenges we observed during this test cycle. We expect to continue to do so as we enter a second year of running the online NBTs.

One of the important lessons that we learnt from the process so far is that digital assessment literacy (Eyal, 2012) is required by an assessment team or a lecturer (as assessor in an online environment) that intends to embark on this kind of project. The lecturer or assessor is the designer of the experiences and tools, as well as the data miner and the interpreter of the results. As such, they must engage with all potential data points that would inform their design decisions. This includes reflecting on their own experiences during the development of the instruments and the decisions they made before and during implementation, studying the observations made by those providing support and operational services that form part of the administration of the test and, perhaps most importantly, investigating the student experience from the moment that they register to when they receive their results.

Further research is pointing in the direction of comparability of test experiences and quantitative analysis of test scores due to differences in test administration (i.e., paper-based versus computer-based testing). Looking at Academic Literacies in conjunction with other literacies, such as "functional literacy" (Vágvölgyi, 2016) and "digital literacy" (Tang, 2016) in the South African context would be essential to understanding candidates' readiness, their skills and abilities to take online high-stakes assessments and to inform strategies for training and skills development.

At present, there are still many restrictions that limit our ability to administer the tests in the paper-based format, as we did before the start of the COVID-19 pandemic, and this will continue for the foreseeable future. Even once we return to circumstances that will make paper-based delivery easier, we do not anticipate that we will ever go back to exclusively using the paper-based delivery mode.

7. Acknowledgements

References

Aisbitt, S. & Sangster, A. 2005. Using internet-based on-line assessment: A case study. *Accounting Education: An International Journal*, 14(4): 383-394. https://doi.org/10.1080/06939280500346011

Alsadoon, H. 2021. Challenges of deploying online exams. *Revista Romaneasca pentru Educatie Multidimensionala*, 13(1Sup1): 403-415. https://doi.org/10.18662/rrem/13.1Sup1/403

Black, P., Burkhardt, H., Daro, P., Jones, I., Lappan, G., Pead, D. & Stephens, M. 2012. High-stakes examinations to support policy. Design, development and implementation. *Educational Designer*, 2(5): article 16.

Brookhart, S.M. & Nitko, A.J. 2019. *Educational assessment of students*, eighth edition. Pearson international.

Burnett, T. & Fuentes, S.P. 2020. *Assessment in the time of pandemic: A panic-free guide*. Available at https://doi.org/10.53593/n3298a [Accessed 3 November 2021].

CETAP. 2019. *The National Benchmark Tests National Report* (p.14). Available at https://nbt.uct.ac.za/content/reports [Accessed 2 August 2021].

Cliff, A. & Yeld, N. 2006. Test domains and constructs: Academic literacy. In H. Griesel (Ed.). *Access and entry level benchmarks: The national benchmark tests project* (pp.19-25). Pretoria: Higher Education South Africa.

Cliff, A. 2015. The national benchmark test in academic literacy: How might it be used to support teaching in higher education. *Language Matters*, 46(1): 3-21. https://doi.org/10.1080/10228195.2015.1027505

Chinembiri, T. 2020. *Despite reduction in mobile data tariffs, data still expensive in South Africa.* Available at https://researchictafrica.net/wp/wp-content/uploads/2020/06/Tapiwa-Chinembiri-Mobile-Data-Pricing-Policy-Brief2-2020-FINAL.pdf [Accessed 2 August 2021].

Coombe, G., Lester, A. & Moores, L. 2020. *Ofqual report: Online and on-screen assessment in high stakes, sessional qualifications. A review of the barriers to greater adoption and how these might be overcome.* Coventry: Ofqual.

Cramp, J., Medlin, J.F., Lake, P. & Sharp, C. 2019. Lessons learned from implementing remotely invigilated online exams. *Journal of University Teaching & Learning Practice*, 16(1): article 10. https://doi.org/10.53761/1.16.1.10

Eyal, L. 2012. Digital assessment literacy – The core role of the teacher in a digital environment. *Journal of Educational Technology & Society*, 15(2): 37-49.

Frankl, G. & Schratt-Bitter, S. 2012. Online exams: Practical implications and future directions. *Proceedings of the European Conference on e-Learning*: 158-164.

Frith, V. & Prince, R. 2006. Quantitative literacy. In H. Griesel (Ed.). *Access and entry level benchmarks: the national benchmark tests project* (pp. 28-34; 47-54). Pretoria: Higher Education South Africa.

Griesel, H. 2006. The context of the national benchmark project. In H. Griesel (Ed.). *Access and entry level benchmarks: The national benchmark tests project* (pp. 1-6). Pretoria: Higher Education South Africa.

Guyer, R. & Thompson, N.A. 2013. *User's manual for Xcalibre™ item response theory calibration software, version 4.2*. Woodbury MN: Assessment Systems Corporation Press.

Kemp, S. 2021. *Digital 2021: South Africa*. Available at https://datareportal.com/reports/digital-2021-south-africa [Accessed 30 July 2021].

Khan, S. & Khan, R.A. 2019. Online assessments: Exploring perspectives of university students. *Education and Information Technologies*, 24: 661-677. https://doi.org/10.1007/s10639-018-9797-0

Kubiszyn, T. & Borich, G. 2013. *Educational testing and measurement: Classroom application and practice,* tenth edition. John Wiley & Sons, Inc.

Leonard, W.P. 2020. *Why standardised testing is necessary to select students*. Available at https://www.universityworldnews.com/post.php?story=20200629110332217 [Accessed 6 November 2021].

Luna-Bazaldua, D., Liberman, J. & Levin, V. 2020. *Moving high-stakes exams online: Five points to consider.* Available at https://blogs.worldbank.org/education [Accessed 3 November 2021].

Mabizela, S.E. & George, A.Z. 2020. Predictive validity of the national benchmark test and national senior certificate for the academic success of first-year medical students at one South African university. *BMC Medical Education*, 20: 1-10. https://doi.org/10.1186/s12909-020-02059-8

Magno, C. 2020. *High-stakes examinations and large-scale learning assessments in times of emergencies and crises. NEQMAP 2020 thematic review.* Bangkok: UNESCO.

Morrow, W.E. 2009. *Bounds of democracy: Epistemological access in higher education*. Cape Town: HSRC Press.

Moss, P.G. 2020. *Training students for online exams reduces cognitive overload.* Available at https://paulgmoss.com/2020/05/22/training-students-for-online-exams-reduces-cognitive-overload/ [Accessed 2 August 2021].

Nel, B.P. 2020. Implications of the quantitative literacies test results of the National Benchmark Test Project (NBTP) for teachers. *South African Journal of Education*, 40(1): 1-8. https://doi.org/10.15700/saje.v40n1a1792

Ngqondi, T., Maoneke, P.B. & Mauwa, H. 2021. A secure online exams conceptual framework for South African universities. *Social Sciences & Humanities Open*, 3(1): 100132. https://doi.org/10.1016/j.ssaho.2021.100132

Nitko, A.J. 2001. *Educational assessment of students*. New Jersey: Merrill Prentice Hall.

Nusche, D. 2013. Student assessment: Putting the learner at the centre. In H.D. Herzog (Ed.). *Synergies for better learning* (pp. 139-269). Paris, France: OECD. https://doi.org/10.1787/9789264190658-7-en

Prince, R. & Frith, V. 2020. An investigation of the relationship between academic numeracy of university students in South Africa and their mathematical and language ability. *ZDM: The International Journal on Mathematics Education,* 52(3): 433-445. https://doi.org/10.1007/s11858-019-01063-7

Prince, R.N., Frith, V., Steyn, S. & Cliff, A. 2021. Academic and quantitative literacy in higher education: Relationship with cognate school-leaving subjects. *South African Journal of Higher Education*, 35(3): 163-181. https://doi.org/10.20853/35-3-3943

Prisacari, A.A. & Danielson, J. 2017. Computer-based versus paper-based testing: investigating testing mode with cognitive load and scratch paper use. *Computers in Human Behavior*, 77: 1-10. https://doi.org/10.1016/j.chb.2017.07.044

Richardson, M. & Clesham, R. 2021. Rise of the machines? The evolving role of AI technologies in high stakes assessment. *London Review of Education*, 19(1): 1-13. https://doi.org/10.14324/LRE.19.1.09

SPARC. 2021. *Higher education reckons with concerns over online proctoring and harm to students.* Available at https://sparcopen.org/news/2021/higher-education-reckons-with-concerns-over-online-proctoring-and-harm-to-students/ [Accessed 2 November 2021].

Steyn, S., Prince, R., Sango, T. & Mudavanhu, P. 2020. Enhancing pedagogical responsiveness post-COVID 19. *Paper presented at the TLC2020 Conference*. University of Cape Town, 17-23 September.

Tang, C.M. & Lee Yen Chaw, L.Y. 2016. Digital literacy: A prerequisite for effective learning in a blended learning environment? *The Electronic Journal of e-Learning*, 14(1): 54-65.

McGrath, C.H., Henham, M.L., Corbett, A., Durazzi, N., Frearson, M., Janta, B., Kamphuis, B.W., Katashiro, E., Brankovic, N., Guerin, B., Manville, C., Schwartz, I., & Schweppenstedde. D. 2014. Higher Education entrance qualifications and exams in Europe: a comparison study. The European Parliament. https://doi.org/10.7249/RR574

The World Bank. 2020. *High-stakes school exams during COVID-19 (Coronavirus): What is the best approach?* Available at https://blogs.worldbank.org/education/high-stakes-school-exams-during-covid-19-coronavirus-what-best-approach [Accessed 2 November 2021].

UNESCO. 2020a. COVID-19 – An overview of national coping strategies on high-stakes examinations and assessments. Paris, France: UNESCO.

UNESCO. 2020b. Managing high-stakes assessments and exams during crisis. *UNESCO Covid-19 Education Response. Education Sector.* Issue Notes: Issue Note N°4.3.

Vágvölgyi, R., Coldea, A., Dresler, T., Schrader, J. & Nuerk, H.C. 2016. A review about functional illiteracy: Definition, cognitive, linguistic, and numerical aspects. *Frontiers in Psychology*, 7: 1617. https://doi.org/10.3389/fpsyg.2016.01617

Wise, S.L. 2019. Controlling construct-irrelevant factors through computer-based testing: disengagement, anxiety, & cheating. *Education Inquiry*, 10(1): 21-33. https://doi.org/10.1080/20004508.2018.1490127

Zaharias, P. 2004. Usability and e-Learning: The road towards integration. *eLearn Magazine*, (6): 4. https://doi.org/10.1145/998337.998345