A Comparison of Selected Assessment Results before and during the COVID-19 Pandemic in One University in South Africa

Tafirenyika Mafugu[1]

[1] School of Mathematics, Science and Technology Education, University of the Free State, Phuthaditjhaba, South Africa

Correspondence: Tafirenyika Mafugu, School of Mathematics, Science and Technology Education, University of the Free State, Phuthaditjhaba, South Africa.

Abstract

The study examined the impact of coursework-only assessment, as made necessary at the onset of the COVID-19 pandemic, adopting a quantitative research approach with 1013 students. The data obtained were analysed using SPSS version 27.0 to obtain descriptive and inferential statistics. The results revealed significant differences between the 2019- and 2020 marks for the same courses. In two of the science courses (T2 and T3), the mean scores for 2019 were significantly higher than the mean scores for 2020. In the mathematics course, the 2020 marks were significantly higher than the marks for 2019. While a normal distribution was assumed for the science courses, the mathematic course showed marks that were skewed to the right. A higher number of distinctions in the F1 course and a significant decline in the mean scores for T1 and T2 implies that there is a need for professional development of lecturers teaching in the online space. It is, therefore, recommended that higher education lecturers need adequate professional development on setting and administering online assessments. The assessment should test adequate lower- and higher-order cognitive skills for sufficient testing of student knowledge during online assessments. Furthermore, a variety of assessment methods and a diversity of tasks may be used to ensure the reliability of the assessment outcomes.

Keywords: online assessment, coursework, examination, higher education, COVID-19 pandemic

1. Introduction

The onset of the COVID-19 pandemic brought several challenges to the teaching and learning fraternity. The coronavirus pandemic brought abrupt changes in instructional practices as well as assessment methods. Guangul, Suhail, Khalit, and Khidhir (2020) observed challenges of academic dishonesty, coverage of learning outcomes and lack of commitment by students to submit assessment tasks. Mafugu (2020) noted challenges of poor attendance as well as network challenges resulting in posting PowerPoint presentations with audio. However, monitoring the learning process after posting the audio was not possible. Poor attendance and failure to interact with the learning material has a significant impact on the assessment outcome.

Assessment is a process that follows careful planning and teaching; a process to evaluate the student's performance. It is an orderly process that is done during teaching, at the end of a chapter, end of a term and end of a year. According to Black and William (1998), assessment refers to activities undertaken by both teachers and students that provide feedback that can be used for student learning, and to modify teaching. Assessments are usually grade-based and they include exams, portfolios, final projects, and standardised tests.

Two broad groups of assessment include formative and summative assessment. Formative assessments give insights into student learning and shape the learning process (Kampen, 2020). Formative assessments also help teachers to meet the learning needs of students and to provide remedial measures for those who are struggling with content comprehension. It also gives an indication whether what the student has learnt, is in line with the curriculum (Kampen, 2020). Formative assessment is done to check learners' understanding of concepts, while summative assessment can be used to report learners' performance, and to progress students into the next level (LaDue, Libarkin, & Thomas, 2015). In addition, formative assessments provide essential feedback to improve one's instructional strategy, and help to teach critical thinking skills as well as problem-solving skills (Kampen, 2020).

In a study by Bakhru and Mehta (2020), students affirmed that group work assignment and presentations assisted them in understanding the topic because they were able to get effective corrections and guidance during the presentation. Students also echoed that the presentations assisted them in developing confidence. Presentations and online-based computer testing can help to develop a better learning environment at institutions of higher learning (Bakhru & Mehta, 2020). According to several studies (Ellis & Calvo, 2004 and 2006; Ellis, Goodyear, Calvo & Prosser, 2008), learning through discussion deepens understanding of concepts.

Even when instruction is planned with great care, delivered effectively, and in a way that engages students, the learning outcomes often bear little or no relation to what was intended. According to (William, 2011), careful planning of the lesson delivery strategy, effective lesson delivery, and student engagement, does not guarantee the total grasp of all the concepts taught. Students comprehend the concepts at different rates, hence, the need for assessment. Assessment will assist in identifying areas that have not been understood by students, and assist in identifying misconceptions and areas that need remediation (Bakhru & Mehta, 2020; Ghaica, 2016).

The study is underpinned by the social cognitive theory which views learning as an active process of knowledge construction based on social interaction and the environment (Kay & Kibble, 2016). In the current study the construction of knowledge occurs through the interaction among the lecturer, students, and the environment (including learning platforms). During the learning process, Murray (2019) views assessment as a key pillar of evaluating the effectiveness of the pedagogical instructional strategies and learners' comprehension of the content. The interactions and feedback on the individual assessment activities facilitate the learning process. According to Ghaica (2016) pedagogical instruction and assessment are complementary processes that are critical in providing knowledge and support. Continuous assessment which involves both formative and summative assessment can be used to get feedback necessary to track students at risk, monitoring progress, modifying pedagogical practices and certification (Ghaica, 2016). Assessment assists the lecturer to reflect and engage in student-lecturer discussion on how to improve teaching and learning and achievement scores. According to Jimaa (2011), the methods used to assess students influence the learning process. The assessment methods influence how learners study the content in preparation for the assessment.

The environmental factors that influence the pedagogical practices and assessment methods include the availability of training for the online platforms used, contextual constraints for the organisation, lecturers and students (Tinoca, Oliveira & Pereira, 2014). The constraints may include availability of resources such as laptops, internet, electricity, class size, and other key result areas in which the lecturer is expected to perform by the institution. Palm (2008) and Pereira, Tinoca and Oliveira (2010) indicate the need to align the assessment design with competencies intended to be developed and the learning situations experienced by the student during the teaching process. Furthermore, Dierick and Dochy (2001), and Herrington and Herrington (1998) emphasise the need to employ a variety of assessment methods and diversity of tasks to ensure the reliability of the assessment outcome.

When setting assessments, it is important to consider Bloom's taxonomy of cognitive domains to ensure that learning outcomes are aligned with the assessment's methods. The revised Bloom's taxonomy has four knowledge dimensions namely factual, conceptual, procedural, and metacognitive; and six cognitive process dimensions as: to remember, understand, apply, analyse, evaluate, and create (Krathwohl, 2002; Dalton, 2003).

Table 1. Revised Bloom's Taxonomy knowledge dimensions and cognitive process dimensions

| | Cognitive Process Dimension | | | | | |
| | Knowledge | | Skill | | Ability | |
Knowledge dimension	1. Remember	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
Factual	Remember facts.	Understand facts.	Apply facts.	Analyse using facts.	Evaluate using facts.	Create using facts.
Conceptual	Remember concepts.	Understand concepts.	Apply concepts.	Analyse using concepts.	Evaluate using concepts.	Create using concepts.
Procedural	Remember procedures.	Understand procedures.	Apply procedures.	Analyse using procedures.	Evaluate using procedures.	Create using procedures.
Metacognitive	Remember metacognitive strategies.	Understand metacognitive. strategies.	Apply metacognitive strategies.	Analyse metacognitive strategies.	Evaluate metacognitive strategies.	Create metacognitive strategies.

Source: (Dalton, 2003)

Bloom's taxonomy include lower order cognitive skills development, which require simple recall and a minimum level of understanding (Jideani & Jideani, 2012). The lower order cognitive skills focus on understanding and remembering facts, concepts, procedures, and metacognitive strategies. The revised Bloom's taxonomy also focuses on higher-order skills, where students are tested on their ability to apply the knowledge gained to analyse-, evaluate-, and create data. The higher-order cognitive skills require deep conceptual understanding, and therefore, most students struggle to perform well in the latter (Crowe, Dirks, & Wenderoth, 2008). Tests must focus on both lower- and higher-order cognitive skills. For students to perform well in higher-order cognitive skills, they must be afforded enough practice to develop a deep conceptual understanding of the material (Crowe, Dirks, & Wenderoth, 2008; Jideani & Jideani, 2012).

According to LaDue et al. (2015), scientific and engineering practices emphasise the importance of developing and using models, analysing and interpreting data, using mathematics and computational thinking, and obtaining, evaluating, and communicating information. LaDue et al. (2015) also claim that assessment in science should incorporate disciplinary core ideas, cross cutting concepts, and scientific and engineering practices. Assessment should contain multiple components to adequately assess the standards of scientific knowledge and process skills components. Visual representations such as graphs and tables are highly rated in science education (LaDue, Libarkin, & Thomas, 2015). Due to the value ascribed to visuals, they often form part of science assessment activities. The ACT science tests provide a good example of science tests because they measure science process skills, which include higher cognitive order skills data interpretation, analysis, evaluation, and problem solving (ACT, 2021). The 40-question ACT science test measures science process skills, which include data interpretation, analysis, evaluation, reasoning, and problem solving (ACT, 2021). About 40% of the test questions measure graph reading skills, interpretation of scatterplots, and interpretation of information presented in tables, diagrams, and figures. An analysis of the Biology and Chemistry New York State Regent examinations indicates that it contain a notable number of diagrams, tables, and graphs (LaDue, Libarkin, & Thomas, 2015). Test items need to undergo a rigorous test-development process that involves pilot testing items and reliability studies via classical test theory and item response theory (LaDue, Libarkin, & Thomas, 2015).

Previous studies focused on assessment where examinations were written under controlled conditions, with invigilators monitoring to ensure that there are no reference books or access to the internet. The current study explores a unique and unusual situation where students complete assessments with all books, notes, and internet sources available for reference, with no physical invigilation. This requires new questioning techniques to ensure that the desired learning outcomes are achieved, and that learning is reflected by the achievement scores. For many lecturers in tertiary institutions, this was not an effortless process.

The study, therefore, intended to answer the research question:

How do the achievement scores of selected undergraduate courses in which there were no final examinations during the COVID-19 period in 2020 compare with the results of the pre-COVID-19 period in 2019?

The following hypotheses were tested at 5% probability levels:

1) H_0: There is no significant difference in the student mean achievement scores of T1 course between 2019 when both coursework and examination were written and 2020 when only coursework was considered.

2) H_0: There is no significant difference in the student mean achievement scores of T2 course between 2019 when both coursework and examination were written and 2020 when only coursework was considered.

3) H_0: There is no significant difference in the student mean achievement scores of T3 course between 2019 when both coursework and examination were written and 2020 when only coursework was considered.

4) H_0: There is no significant difference in the student mean achievement scores of F1 course between 2019 when both coursework and examination were written and 2020 when only coursework was considered.

2. Research Methodology

The study was underpinned by a quantitative research approach. The marks compared were only for lecturers who gave consent for the marks to be analysed. Consent was sought from two lecturers; one with three science courses, and one with one mathematics course.

The final year marks for three science courses (T1, T2 & T3) and one mathematics course (F1) for 2019 and 2020 were downloaded onto an excel sheet from the gradebook. The marks were transferred into SPSS version 27.0. The marks were categorised into the following sections: 0-39; 40-49; 50-59; 60-69; 70-74; 75-100. The frequencies of

marks in each category were obtained in a table in the SPSS output (Table 2). Furthermore, descriptive statistics (mean and standard deviation, Table 3) were obtained for each set of results for 2019 and 2020.

Inferential statistics: The Independent sample t-test was used to determine if the difference in the mean scores for 2019 and 2020 were significantly different. This was only done for the set of data that met the condition of normality after the Kolmogorov-Smirnov test. For the set of data where the condition for normality was not met (F1), the non-parametric independent samples t-test (Mann-Whitney U test) was used in the comparison. Levene's test for equality of variances was also performed using SPSS. Levene's test for equality of variances was used to determine whether equality of variance should be assumed or not, in the independent samples t-test. Where the Levene's test indicated no significant difference ($p > 0.05$), equal variance was assumed in the independent samples t-test, while equal variance was not assumed where the Levene's test indicated significant difference ($p < 0.05$). All tests were conducted at 5% significant level.

The researchers obtained ethical clearance from the University Ethical Clearance Committe, and the lecturers who taught the modules from 2019 to 2020 gave informed consent. The course data were coded T1, T2, T3 and F1 to ensure anonymity and confidentiality.

3. Results

This section presents the frequencies of marks that were obtained in four courses during 2019 and 2020. In 2019, the students were assessed using the coursework marks obtained through mini-assessment tasks during the term, and an examination that students wrote at the end of the semester. The coursework marks contributed 50% to the final mark, with the other 50% reflecting the final semester examination. In 2020, only coursework marks were considered. The COVID-19 pandemic's restrictions required social distance to minimise the spread of the coronavirus, and therefore dictated that students could not write venue-based examinations with physical invigilation.

Table 2. Distribution of marks for different courses in 2019 (preCOVID-19) and 2020 (during the COVID-19 period)

| | Courses | | | | | | | |
| | T1 | | T2 | | T3 | | F1 | |
Mark category	Frequency & % 2020	Frequency & % 2019	Frequency & % 2020	Frequency & % 2019	Frequency & % 2020	Frequency & % 2019	Frequency & % 2020	Frequency & % 2019
75-100	12(10.3)	6(4.8)	3(7.1)	5(31.3)	6(4.5)	11(18.3)	114(43.8)	5(1.9)
70-74	17(14.5)	8(6.4)	8(19.0)	6(37.5)	7(5.3)	8(13.3)	86(33.1)	20(7.5)
60-69	59(50.4)	71(56.6)	13(31.0)	4(25.0)	61(45.9)	30(50.0)	43(16.5)	128(48.3)
50-59	18(15.4)	36(28.8)	14(33.31)	1(6.3)	50(37.7)	9(15.0)	9(3.5)	71(26.8)
40-49	4(3.4)	1(0.8)	1(2.4)	0	5(3.8)	2(3.3)	1(0.4)	17(6.4)
0-39	7(6.0)	3(2.4)	3(7.1)	0	4(3.0)	0	7(2.7)	19(7.2)
Total	117(100)	125(100)	42(100)	16(100)	133(100)	60(100)	260(100)	260(100)

The values in brackets (…) represent %.

The majority the students for T1, 59(50.4%) and 71(56.6%) in 2020 and 2019 respectively, obtained marks between 60-69, while the majority 128(48.3%) in F1 course obtained marks between 60-69 in 2019 with a few students in the extremes (Table 2). In 2019 most of the students (eleven, therefore 68.8%) of the class undertaking T2, had marks above 70%. For T3, the majority of the students had marks between 50-69 in 2020 and 2019, with few students in the extremes. In F1, the majority of students obtained distinctions (marks above 75%). The absence of examinations may have contributed to the reflected performance difference.

Table 3. Mean and standard deviations of marks of different courses obtained in 2019 (pre-COVID-19 period) and 2020 (during the COVID-19 period)

Course	year	N	Mean mark.	Std. deviation.
T1	2020	117	62.85	12.06
	2019	125	61.34	8.60
T2	2020	42	60.81	13.51
	2019	16	71.38	9.32
T3	2020	133	59.16	10.89
	2019	60	65.07	11.88
F1	2020	260	76.0	14.63
	2019	260	57.7	13.50

For T1 and F1, the mean marks for 2020 were higher than the mean for 2019 (Table 3). However, in courses T2 and T3, the mean marks for 2019 were higher compared to the marks for 2020, where the number of learners was more than double.

Table 4. Levene's test for equality of variances and independent sample t-test for T1 2019 and 2020 marks

	Levene's Test for Equality of Variances		Independent sample t-test		
	F	Sig.	t	Df	Sig. (2-tailed)
Equal variances assumed.	6.39	.012	1.13	240	.258
Equal variances not assumed.			1.12	208.59	.263

Levene's test for equality of variances indicates that the difference in variance for the T1 2019 and 2020 marks was significant (Table 4). Equal variance was not assumed in the independent sample t-test at 5 % level of significance. The was no significant difference in student mean marks for T1, $t(209) = 1.12$, $p = 0.26$ despite students in 2020 (M = 62.85, SD 12.06) attaining higher scores than 2019 students (M = 61.34, SD = 8.60).

Table 5. Levene's test for equality of variances and independent sample t-test for T2 2019 and 2020 marks

	Levene's Test for Equality of Variances				Independent sample t-test
	F	Sig.	t	Df	Sig. (2-tailed)
Equal variances assumed.	1.29	.262	-2.87	56	.006
Equal variances not assumed.			-3.38	39.37	.002

Levene's test for equality of variances indicates that the difference in variance for the T2 2019 and 2020 marks was not significant (p = 0.26) at 5 % level of significance, hence, equal variance was assumed in the independent sample t-test (Table 5). The 16 students for T2 in 2019 (M = 71.38, SD = 9.32) compared to 42 students in 2020 (M = 60.81, SD 13.51) obtained significantly higher scores, $t(56) = -2.87$, $p = 0.006$.

Table 6. Levene's test for equality of variances and independent sample t-test for T3 2019 and 2020 marks

	Levene's Test for Equality of Variances		Independent sample t-test		
	F	Sig.	t	df	Sig. (2-tailed)
Equal variances assumed	0.17	.900	-3.39	191	.000
Equal variances not assumed.			-3.28	105.42	.000

Similarly, Levene's test for equality of variances indicate that the difference in variance for the T3 2019 and 2020 marks was not significant (p =0.90) at 5 % level of significance, hence, equal variance was assumed in the independent sample t-test (Table 6). The 60 students for T2 in 2019 (M = 65.07, SD = 11.88) compared to 133 students in 2020 (M = 59.16, SD 10.89 obtained significantly higher scores, $t(191) = -3.39$, $p = 0.00$.

Table 7. Mann-Whitney Test and the Ranks for 2020 and 2019 F1 marks

Ranks				Test Statistics	
Year	N	Mean Rank	Sum of Ranks	Mann-Whitney U	7478.00
2020	260	359.29	3216.00	Wilcoxon W	41408.00
2019	260	159.26	1834.00	Z	-15.24
Total	520			Asymp. Sig. (2-tailed)	.000

The achievement scores for 2020 (Mdn = 359.3) were higher than those for 2019 (Mdn = 159.3). A Mann-Whitney test indicated that this difference was statistically significant, $U(N_{2020mark} = 260, N_{2019mark} = 260) = 7478.00$, $z = -15.24$, $p < 0.00$ (Table 7).

4. Discussion

Both 2019 and 2020 results for the T1 to T3 courses and FI course results for 2019 indicate clear discrimination and a good distribution of students' marks with few student scores in the extremes, and the majority (60-69%) of the scores clustered around the mean. T2 results for 2019 and 2020 show a similar trend with the highest frequencies recorded in the categories (70-74%) and (50-59%) respectively. Attainment of a normal distribution curve in students' performance scores, according to Krathwohl (2002), can be achieved when the assessment tasks are carefully designed to include discriminatory activities or questions that focus on judging, critiquing, constructing, designing, and hypothesising (Figure 1). In designing assessments, it is essential to focus on cognitive process dimensions. The diagram (Figure 1) can assist one in designing appropriate questions that fall into specific cognitive process dimensions.
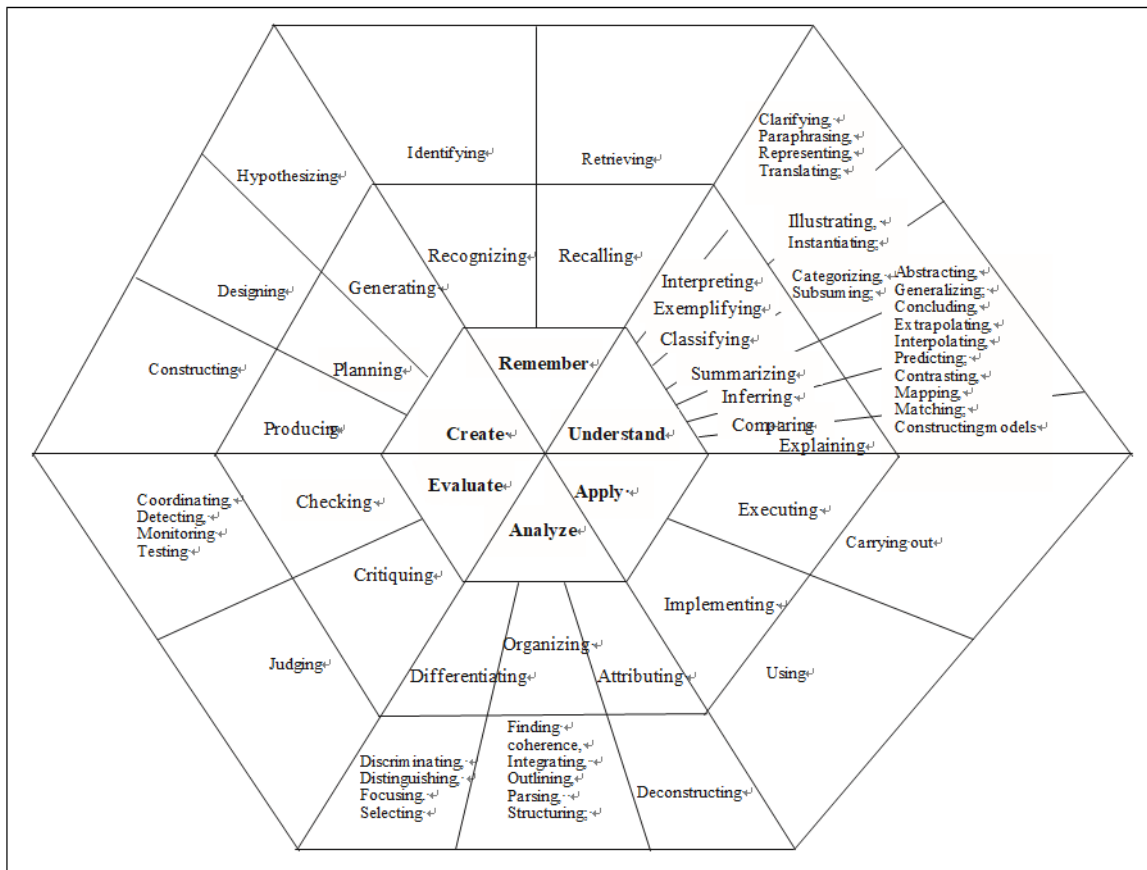

Figure 1. The cognitive process dimensions (Krathwohl, 2002)

The pattern of the T1, T2 and T3 results reflect that the learners were heterogeneous in ability and the assessment activities adequately discriminated the gifted students from the average students in both 2019 and 2020. The pattern

of results displayed by both 2019 and 2020 T1 and T3 courses and F1 course results for 2019, could also be explained by possible good alignment of the assessment design with competences intended to be developed, and the learning situations experienced by the student during the teaching process (Palm, 2008; Pereira, Tinoca & Oliveira, 2010).The other dimension that could explain the results, according to Dierick and Dochy (2001) and Herrington and Herrington (1998), was a possible use of a variety of assessment methods and diversity of tasks to ensure the reliability of the assessment outcome.

The significantly higher mean scores of the 2019 T2 and T3 courses than similar courses in 2020 could have been attributed by the challenges in the novel online teaching and assessment strategies that were abruptly introduced in the tertiary institutions. Guangul et al. (2020) observed challenges of coverage of learning outcomes and lack of commitment by students to submit assessment tasks. Mafugu (2020) noted challenges of poor attendance as well as network challenges which negatively affect assessment outcomes. The onset of the pandemic could have resulted in limited interaction due to challenges such as network. According to the social cognitive theory, learning is more effective when learners are actively involved in the process of knowledge construction through peer-peer interaction or student- lecture interaction as well as interaction with the environment (Kay & Kibble, 2016; Mafugu, 2021).

Li and Konsantoppoulos, (2017) observed no significant effect of class size on student performance. Hence, we can assume that the class size had no effect on performance. In the course F1, results for 2020 student scores show the majority 114(43.8%) of students in the 75-100% category in a course with a total of 260 students. Furthermore, the scores for F1 in 2020 were significantly higher than the scores for 2019. The 2020 F1 results indicate a lack of adequate discrimination questions, resulting in all students' final scores clustered in the upper extreme. This could have been the effect of the rapid change to online, making it difficult for some lecturers to produce relevant strategies that allow adequate discrimination of student scores. The skewed results could have been a resulting lack of higher-order questions that demanded application of knowledge under conditions, where students have access to all resources. According to Crowe, Dirks and Wenderoth (2008), the higher-order cognitive skills require deep conceptual understanding; hence most students struggle to perform well. According to Rush, Rankin and White (2016), an increase in item complexity increases the index of difficulty and discrimination value of the item. Guangul et al. (2020) observed challenges of academic dishonesty among students when writing assessments where there is no one to monitor any form of cheating. In the F1 course, the assessor who was used to giving a final year examination that would be written under strict conditions, possibly had difficulties for setting challenging activities that were written by students who were not monitored.

Lemons and Lemons (2013) highlight that it is important to consider Bloom's Taxonomy, as well as the amount of the time required to complete the task and the student experience, in practising similar assessment tasks to determine the index of question difficulty (Figure 2). Higher-order questions may require more time to complete, while increased student experience with a particular type of assessment makes it easier for students to complete it (Kotovsky, Hayes, & Simon, 1985; Sweller & Chandler, 1994). This study further adds another principal factor to consider when conducting assessments online: the availability of resources. A question that is classified under higher-order critical skills may be easy when students have access to resources where they can get examples that guide them to the answer. Such guidelines may be on the internet, or in other reference resources.
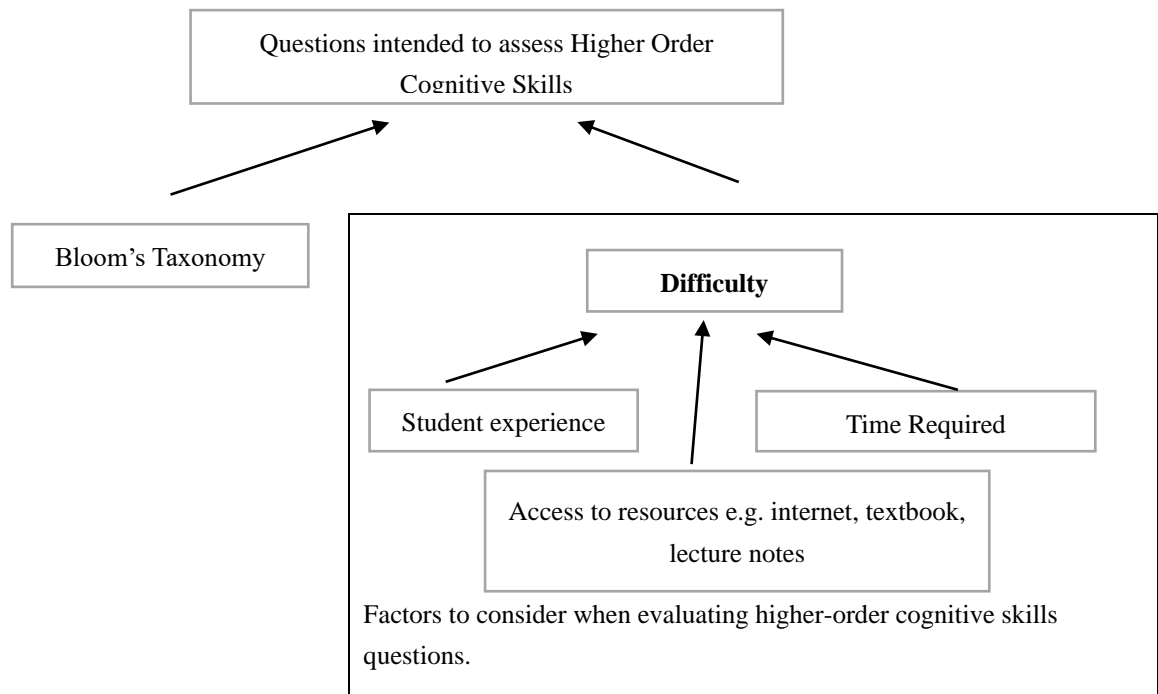
Figure 2. Modified Framework for research showing how one can conceptualise questions intended to assess higher-order critical skills during online teaching and learning. Adapted from: (Lemons & Lemons, 2013)

An understanding of the subject-specific skills is also essential and should be considered during the process of assessment. For example, in science subjects, specific skills that are linked to the subjects should be assessed. The assessment should include application of knowledge, data analysis and interpretation, converting one form of data into another, and communication, which are important during online teaching and learning, as they assist in discriminating the average students from the gifted. This is in line with LaDue, Libarkin and Thomas (2015) who emphasise that scientific practices should highlight the importance of developing and using models, analysing, and interpreting data, using mathematics and computational thinking, and obtaining, evaluating, and communicating information. Ensuring that lecturers understand the critical skills will assist in setting questions that meet the subject standards. Higher order questions provide students the opportunity to develop one of the 21st century skills, namely critical thinking (Tyas, Nurkamto, Marmanto, & Laksani, 2019). This, in turn, results in ensuring that the students have developed the relevant skills needed in the job market.

5. Implications for Teaching and Professional Development

Findings presented in this study have implications for designing teaching activities and assessment tasks suitable when the face-to-face examinations are not possible. Assessment must include higher-order cognitive skills. Bloom's taxonomy, the amount of time required to complete the task, students' experience with the particular type of assessment, as well as the availability of resources must be considered when determining the difficulty index of a question. Assessment should discriminate gifted students from mediocre students by including discriminatory activities or questions that focus on judging, critiquing, constructing, and designing. A well-set formal task should produce a normal distribution curve with two extremes, with the majority getting closer to the mean mark. This rule is particularly important for large classes. It is critical to ensure that lecturers are aware of Bloom's taxonomy, as well as other factors that can be used to evaluate whether a question elicits higher-order cognitive skills. More research needs to focus on the type of questions that are administered by lecturers when students are assessed under conditions where they have access to all resources, and lack the test or examination type-controlled conditions.

6. Limitations of the Study Results

The assessments made in 2020 and 2019 were of a different nature due to the COVID-19 pandemic. While there were examinations and continuous assessment marks in 2019, there were no examination in 2020 due to the pandemic that called for social distancing worldwide. However, the comparison was facilitated by the need to see if the lack of examinations did not result in either exaggerated high marks, or too low in student performance. Another limitation is that the study did not focus on the specific type of questions-, and the knowledge dimensions and

cognitive process dimensions of the questions asked. The focus was only on the outcome scores for the students. The lecturer sample population was very limited and future studies will need a statistically significant lecturer sample size to strengthen the generalisability of the inferences and conclusions of the study.

Conflict of Interest

The author declares no conflict of interest.

Source of funding

NIL

References

ACT. (2021). The *ACT Test.* Retrieved from content covered by the ACT test. Retrieved from https://www.act.org/content/act/en/products-and-services/the-act/test-preparation/science-practice-test-questions.html-?page=0&chapter=0

Bakhru, S. A., & Mehta, R. P. (2020). Assignment and Project Activity Based Learning Systems as an Alternative to Continuous Internal Assessment. *Procedia Computer Science, 172*, 397-405. https://doi.org/10.1016/j.procs.2020.05.073

Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80(2)*, 139-148.

Crowe, A., Dirks, C., & Wenderoth, M. (2008). Biology in Bloom: implementing. *CBE-Life Science Education, 7*, 368-81. https://doi.org/10.1187/cbe.08-05-0024

Dalton, E. (2003). *The "New Bloom's taxonomy," objectives and assessments*. Retrieved from: http://gaeacoop.org/dalton/publications/new_bloom.pdf

Dierick, S., & Dochy, F. J. R. C. (2001). New lines in edumetrics: new forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation, 27*, 307-329. https://doi.org/10.1016/S0191-491X(01)00032-3

Ellis, R. A., & Calvo, R. A. (2004). Learning through discussions in blended environments. *Educational Media International, 41*, 263-274. https://doi.org/10.1080/09523980410001680879

Ellis, R. A., & Calvo, R. A. (2006). Discontinuities in university student experiences of learning through discussions. *British Journal of Educational Technology, 7*, 55-68. https://doi.org/10.1111/j.1467-8535.2005.00519.x

Ellis, R. A., Goodyear, P., Calvo, R. A., & Prosser, M. (2008). Engineering students' conceptions of and approaches to learning through discussions in face-to-face and on-line. *Learning and Instruction, 18*, 267-282. https://doi.org/10.1016/j.learninstruc.2007.06.001

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. *Education Assessment Evaluation and Accountability, 32*, 519-535. https://doi.org/10.1007/s11092-020-09340-w

Ghaicha, A. (2016). Theoretical Framework for Educational Assessment: A Synoptic Review. *Journal of Education and Practice, 7*(24), 212-231. Retrieved from https://files.eric.ed.gov/fulltext/EJ1112912.pdf

Herrington, J., & Herrington, A. (1998). Authentic Assessment and Multimedia: how university students respond to a model of authentic assessment. *Higher Education Research & Development, 17*(3), 305-322. https://doi.org/10.1080/0729436980170304

Jideani, V. A., & Jideani, I. A. (2012). Alignment of Assessment Objectives with Instructional Objectives Using Revised Bloom's Taxonomy-The Case for Food Science and Technology Education. *Journal of Food Science Education, 11(3)*, 34-42. https://doi.org/10.1111/j.1541-4329.2012.00141.x

Jimaa, S. (2011). The Impact of Assessment on Students Learning. *Procedia – Social and Behavioral Sciences, 28*, 718-721. https://doi.org/10.1016/j.sbspro.2011.11.133

Kampen, M. (2020, June 23). *The 6 Types of Assessment [+ How to Use Them].* Retrieved from https://www.prodigygame.com/main-en/blog/types-of-assessment/

Kay, D., & Kibble, J. (2016). Learning Theories 101: Application to Everyday Teaching and Scholarship. *Advances in Physiological Education, 40*, 17-25. https://doi.org/10.1152/advan.00132.2015

Kotovsky, K., Hayes, J., & Simon, H. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 248-294. https://doi.org/10.1016/0010-0285(85)90009-X

Krathwohl, D. (2002). A revision of Bloom's taxonomy: an overview. *Theory into practice, 41*, 212-218. https://doi.org/10.1207/s15430421tip4104_2

Kurt, S. (2019). "Social Learning Theory: Albert Bandura," in *Educational Technology*, December 26, 2019. Retrieved from https://educationaltechnology.net/social-learning-theory-albert-bandura/

LaDue, N. D., Libarkin, J. C., & Thomas, S. R. (2015). Visual representations on high school biology, chemistry, earth science, and physics assessments. *Journal of Science Education and Technology, 24*(6), 818-834. https://doi.org/10.1007/s10956-015-9566-4

Lemons, P., & Lemons, J. (2013). Questions for Assessing Higher-Order Cognitive Skills: It's Not Just Bloom's. *Life Sciences Education, 12*(1), 47-58. https://doi.org/10.1187/cbe.12-03-0024

Li, W., & Konsantoppoulos, S. (2017). Does class-size reduction close the achievement gap? Evidence from TIMSS 2011. *School Effectiveness and School Improvement, 28*(2), 292-313. https://doi.org/10.1080/09243453.2017.1280062

Mafugu, T. (2020). Opportunities, challenges, and experiences with digital teaching during the COVID-19 pandemic. *Digital 2020 International Conference Proceedings on Teaching Assessment and Learning in the Digital Age*, 277-289. https://doi.org/10.13140/RG.2.2.24757.70888

Mafugu, T. (2021). The impact of peer-peer learning and student-lecturer interaction on Biology pre-service teachers' achievements. *Cypriot Journal of Educational Sciences*, 1-10. https://doi.org/10.18844/cjes.v16i2.5629

Murray, A. (2019). The role of practical assessment in the delivery of successful enterprise education. *Education+Training, 61*(4), 413-431. https://doi.org/10.1108/ET-10-2018-0216

Palm, T. (2008). Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature. *Practical Assessment Research & Evaluation, 13*(4). Retrieved from http://pareonline.net/getvn.asp?v=13&n=4

Pereira, A., Tinoca, L., & Oliveira, I. (2010). Authentic assessment contribution to competence based education at Universidade Aberta: Questions and challenges. In Siran Mukerji and Purnendu Tripathi (Eds.) *Cases on Technological Adaptability and Transnational Learning: Issues and Challenges* (pp. 266-287). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-61520-749-7.ch015

Rush, B., Rankin, D., & White, B. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Med Educ*, 250. https://doi.org/10.1186/s12909-016-0773-3

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognitive and Instruction, 12*, 185-233. https://doi.org/10.1207/s1532690xci1203_1

Tinoca, L., Oliveira, I., & Pereira, A. (2014). A conceptual framework for e-assessment in Higher Education – authenticity, consistency, transparency and practicability. In Siran Mukerji and Purnedu Tripathi (Eds.) *Handbook of Research on Transnational Higher Education Management* (pp. 652-673). IGI Global. https://doi.org/10.4018/978-1-4666-4458-8.ch033

Tyas, M., Nurkamto, J., Marmanto, S., & Laksani, H. (2019). Developing Higher Order Thinking Skills (HOTS)–Based Questions: Indonesian EFL Teachers' Challenges. *Proceeding of the 2nd International Conference on Future of Education, 2(1),* 52-63. https://doi.org/10.17501/26307413.2019.2106

William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-11. https://doi.org/10.1016/j.stueduc.2011.03.001