Journal of the Scholarship of Teaching and Learning, Vol. 24, No. 3, September 2024, pp.52-67.
doi: 10.14434/josotl.v24i3.35672

Assessment Design And Practices Toward Holistic Learning Of Higher Education Students: Empirical Evidence Via Path Analysis Modelling Approach

Ebenezer Afrifa-Yamoah
Edith Cowan University
e.afrifayamoah@ecu.edu.au

Esther Adama
Edith Cowan University

Amanda Graf
University of Notre Dame
Edith Cowan University

Kwadwo Adusei-Asante
Edith Cowan University

Abstract: *COVID-19 has revolutionised assessment design and practices in higher education; however, there has been no shift in the objective of enhancing the relationship between assessment and learning that promote the holistic development of students. In this study, we provide an empirical evaluation of the perceived effects of assessment practices (invigilated examination and alternative assessments) on students' mental wellbeing, learning processes, and academic misconduct. A cross-sectional study design was employed for this study in which a self-reported survey instrument was administered to 380 undergraduate and postgraduate students in a public university in Australia. We explored the correlations within defined networks by path analysis via partial least square structural equation modelling (PLS-SEM) framework of SmartPLS 3. Model assessment indexes indicated acceptable convergent, divergent, and construct validity scores for the instrument used. Compared to invigilated exams, students perceived alternative assessments to have significant positive direct effects on stress levels, research skills, learning process, and time management (*$\boldsymbol{p < 0.05}$*). In relation to academic misconduct, students generally perceived invigilated exams to restrain such practices, however, the perceived effect was not statistically significant when compared with alternative assessments (*$\boldsymbol{p > 0.05}$*). Although, the popularity of alternative assessment practices may have been driven by COVID-19, the pilot findings from this study suggest that these assessment designs and practices have greater potential to promote overall student success and productivity and must be encouraged and utilised in the post-COVID-19 era.*

Keywords: assessments, assessment practices, student wellbeing, SEM, student success, invigilated exams

Assessment in education is a process of ensuring learning and achievement of learning outcomes and relevant competencies within a course or program. For years, there has been an ongoing debate on the use of assessment as a tool for active learning (Hand et al., 1996; Rawlusyk, 2018), with some educators advocating for active learning strategies such as peer and self-assessment, presentations and observations (Gibson & Shaw, 2011; Harrison & Heritage, 2019). For effective assessment designs,

Bloom's taxonomy is recommended in higher education to assess both lower- and higher-order learning (Eber & Parker, 2007; Prasad, 2021; Thompson et al., 2008). There are different assessment designs in higher education, but the most common form of summative assessment design is invigilated examinations or tests because they are fast, easy to manage and can be used in courses with higher enrolments. However, they have been criticised for assessing lower-order learning— such as, remembering facts, understanding and application (Jones et al., 2009). Additionally, invigilated examinations cause stress, which although is perceived to have some positive effect on performance, could negatively affect students' physical and mental wellbeing (Jones et al., 2021). Furthermore, the practice has been described as lacking authenticity as it does not reflect activities at the workplace or real world (Sokhanvar et al., 2021). Therefore, there is an increasing drive to move away from assessment designs that promote rote technicalities to those that enhance problem solving, critical thinking and incorporation of relevant 'soft skills' for employment.

The COVID-19 pandemic occasioned rapid changes in higher education with many universities moving teaching and assessments to online or remote settings (Adama et al., 2023; García-Peñalvo et al., 2021). To maintain assessment integrity and intention, many innovative ways of assessing students within COVID-19 restrictions were introduced with increasing move away from invigilated summative examinations to alternative assessments (García-Peñalvo et al., 2021). Alternative assessments evaluate proficiency levels in a subject rather than students' level of knowledge and are often non-invigilated. Such assessments include take-home open-book assessments, projects, vivas for objective structured clinical assessments for health students, and use of online platforms such as Cadmus, Zoom, and Teams for assessments.

Alternative assessments became popular in higher education during the pandemic. Although these assessment practices have been recommended to continue in the post-Covid era (Benito et al., 2021), few studies have explored the perceptions of students regarding alternative assessments compared with the traditional invigilated summative examinations in their learning process and wellbeing. For instance, Daniels et al., (2021) investigated students' perceptions of the impact of alternative assessments changes attributable to the COVID-19 pandemic. While relevant, the study was limited as the researchers did not develop a conceptual model and practical guides for academics in designing robust assessments, that do not only consider engagement and cheating, but also the wellbeing of students. Given that students' perspectives and learning outcomes are important in shaping assessments in higher education, we sought to fill the gap identified above by developing a framework/model to undertake an assessment analysis of the perceived effects of alternative assessments on students' learning, mental health/wellbeing, and academic integrity, in comparison to invigilated examinations via path analysis modelling approach (see Figure 1).

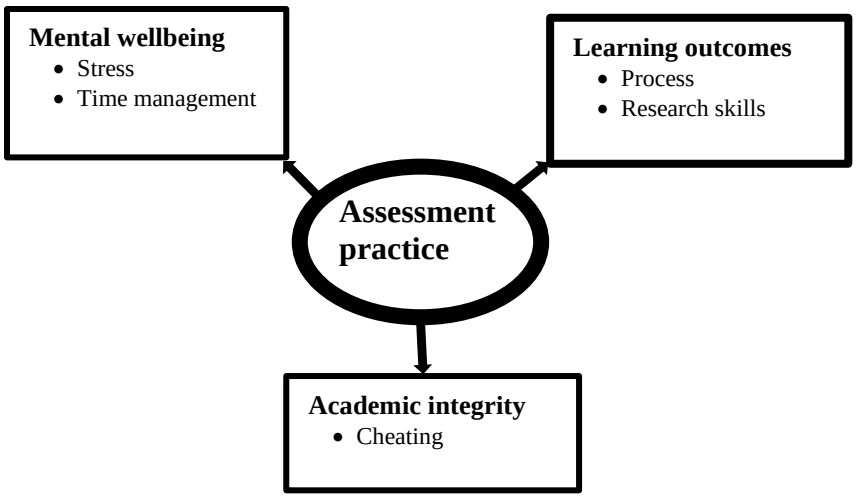

Figure 1*:* Conceptual Model—The Multiplicity of Roles of Assessment Practices in Higher Education

Theoretical framework

Assessment and mental wellbeing

The mental health of students is undoubtedly a high priority in higher education in the post-pandemic era. Positive mental outcomes are associated with effective learning outcomes. The Higher Education Academy acknowledged that the impact of assessments on students' wellbeing stating assessments can cause anxiety and stress (Houghton & Anderson, 2017).

The common assessment-related mental health difficulties are stress and anxiety (Pascoe et al., 2020). The level of these difficulties is associated with different assessments designs—invigilated or non-invigilated, with proctored online assessments having a high degree of mental health difficulties during the COVID-19 pandemic (Conijn et al., 2022).

Students who experience test anxiety or stress have reported difficulties in concentration and achievement of learning outcomes (Graf et al., 2023). Test anxiety contributes to overall students' wellbeing and academic underperformance (Rana & Mahmood, 2010). Rana and Mahmood (2010) concluded that test anxiety is associated with cognitive factors that have a negative impact on learning and academic performance. To address this, they suggested training students on management of test anxiety, which is strongly tied to effective time management. Macan et al. (1990) and (Nayak, 2019) reported that effective time management is positively correlated with higher academic performance, work–life satisfaction and less academic stress. Additionally, Stankovska et al. (2018) reported a positive relationship between test anxiety and academic stress (academic workload and assessments). These findings highlight the fact that anxiety and its concomitant impacts on learning can be managed from the curriculum perspective by developing assessments that reduce test anxiety and stress while achieving learning outcomes (Khan & Madden, 2018). Some of the proposed innovative practices to reduce assessment-related stress and anxiety include, the use of active learning methodologies (Cardozo et al., 2020; Khan & Madden, 2018), mindfulness colouring (Carsley & Heath, 2020), gamification-based assessment (Pitoyo, 2019), combination of skill-focused, behavioural and cognitive approaches (Ergene, 2003) and time management strategies, such as task and space organisation (Alhasani et al., 2021).

At the peak of the COVID-19 pandemic, the swift transition to online learning created an additional layer of stress for students (Almossa, 2021; Fitzgerald & Konrad, 2021). Given the impact of COVID-19 on students' mental health, intentional management strategies in students' mental wellbeing must be promoted by embedding the outlined innovative assessment practices and design within the curriculum. To achieve this, post COVID-19 related teaching and assessments designs should consider students' perspectives to understand the impact of these assessment changes on students' mental wellbeing, primarily focusing on the elements of test stress and time management (Graf et al., 2023).

Assessment and learning

Assessments in higher education are generally designed to observe, demonstrate and/or measure the achievement of specified outcomes that are aimed at enabling students to achieve positive personal growth from the learning (Melton, 1996; Munna, 2021). Effectively, enhancing the relationship between assessment and learning is important for the holistic development of students. The design and practice of assessment must focus on goal-oriented learning processes and not merely a means of monitoring students' academic achievement (Vey, 2005). Students' approaches to learning are greatly influenced by the model of assessment (Biggs, 1996; Gijbel & Dochy, 2006). Information processing is the wheel on which learning thrives, leading to three approaches of learning: surface processing (aimed at memorising text), deep processing (aimed at eliciting main message of the text) and strategic processing (focuses on achieving motivation in studying according to a given assessment) (Parpala et al., 2010). For example, Gijbel and Dochy (2006) reported a change in students' approaches to learning (surprisingly to the surface learning approach) after a hands-on experience with a formative mode of assessment. This reinforces the notion that the mode of assessment used can either have a negative or positive 'backwash effect' on learning (Polat, 2020; Bolshakova et al., 2022).

The mode of delivery of an assessment may have unintended side effects that could undermine the intention to encourage student learning. For instance, rote memorisation learning has generally been identified as a common practice for invigilated assessment. In evaluating the overall outcomes of assessment on four tenets (ability to remember, understand, analyse, and apply), Varble (2014) found that students who took non-invigilated assessment scored higher on all four tenets compared with those who undertook invigilated assessment. These findings highlight that a comparatively high degree of deep learning occurred among students who undertook non-invigilated assessments. This is not surprising because a one-off effort in a limited time can place students under extreme pressure in the lead up to the examination and heighten their stress levels, and ultimately interrupt their learning process. Contrary to these findings, Wellman (2005) observed that invigilated examination was more effective in promoting learning compared with non-invigilated learning and found that the constructs of anxiety, self-testing, attitude/interest and motivation were the significant drivers of the findings. Such contradictions highlight the need for ongoing research in this space and to possibly contextualise the narrative of assessment and learning to the several disciplines in higher education.

Assessment design and practice at higher education must be aimed at the development of complex cognitive skills. One of the complex skills expected to be developed by students in higher education is research skills, which may be transferrable or discipline specific. This skill promotes that "knowledge is developed by people and is continuously being developed further" (Stokking, van der Schaaf, Jaspers, & Erkens, 2004). Research skills allow for life-long learning in defined discipline of study and the evolutionary precept of knowledge that is foundational for innovativeness. The mode of assessments can play a role in this skill development and the perspective of students on the effect of invigilated and non-invigilated assessment would provide useful insight.

Assessment and academic integrity

There are different assessment designs for both formative and summative assessments; however, for summative assessments, invigilated assessments appear to be popular because they are generally believed to assure assessment security and academic integrity (Adama et al., 2023). Academic integrity is an important aspect of education because it sets the foundation for training honest and trustworthy graduates in society. Academic misconduct is perceived to be associated with cheating behaviours in the workplace (Alessio & Messinger, 2021).

The COVID-19 pandemic presented challenging situations in higher education with invigilated summative assessments being affected by COVID-19 restrictions (Fuller et al., 2020). As such, to ensure academic integrity, online proctored assessments were undertaken by 51% of higher education institutions, while the remaining used alternative assessments (Reedy et al., 2021). The debate on whether invigilated or non-invigilated assessment reduces academic misconduct continues to exist in higher education prior to and during COVID-19 (Alessio & Messinger, 2021; Reisenwitz, 2020). Institutions that used invigilated online assessments used sophisticated software to maintain assessment integrity (Reedy et al., 2021). Even so, there is evidence to suggest that these technologies were perceived as an intrusion of privacy and a cause of stress and anxiety for students (Alessio & Messinger, 2021).

Although invigilated assessments are believed to improve assessment security (Dawson, 2021), they have been noted to increase mental health difficulties of students which, in part, affects their overall performance (Reisenwitz, 2020). Conversely, non-invigilated assessments have the ability to increase academic performance and enhance problem-solving skills (Daffin Jr & Jones, 2018); however, they have been criticised for their inability to assure academic integrity (Alessio & Messinger, 2021). Regardless of the use of invigilated or non-invigilated assessments, academic integrity must be assured in all assessments (Adama et al., 2023). At the peak of the COVID-19 pandemic, institutions that resorted to alternative assessments had developed innovative strategies to assure assessment integrity (Gamage et al., 2020). Among the strategies suggested to prevent or reduce the risk of academic integrity include, developing questions that cannot be answered with an internet search, otherwise referred to as "non-Googleable questions" (Adama et al., 2023), to increase higher-order learning, timing assessments, and utilising plagiarism detection tools such as Turnitin®. The question as to whether these strategies affected academic integrity in alternative assessments remains unclear. Thus, as the world continues to learn the new norm of living with COVID-19, it is essential to explore students' perspectives on how academic integrity was affected by alternative assessments. There seems to be a paucity of evidence that support academic institutions' decisions on alternative assessments. In this regard, and from the point of view of co-design principles, eliciting these from students will enhance future assessment designs.

Method

Instrument development

The data collection tool was developed through a review of available evidence on alternative assessments, academic integrity, students' wellbeing, learning processes and in consultation with the associate dean of teaching and learning to generate constructs and items of the tool (Adesile et al., 2016; McCabe & Trevino, 1993; Ramdani, 2018). We included both closed-ended and open-ended questions to capture participants' perceptions and experiences. To ensure validity of the tool, we undertook a pilot study with three students (two undergraduate and one postgraduate students) to ensure face validity. Face validity was assessed in terms of readability, consistency, clarity of language

and formatting style. Additionally, the research tool was shared with the associate dean of teaching and learning (ADTL) in the two schools of the host university for content validity. Generally, the pilot showed that students understood the questions as was intended. However, the duration to complete the questionnaire was adjusted to reflect the true time required for the completion. Students who undertook the pilot study were ineligible for the main study. Furthermore, feedback from ADTL was incorporated in the tool to make it more reflective of the constructs and items of the tool.

Demographic data included age and level of education, program/course of study and number of subjects/units enrolled in one semester. Additionally, 13 questionnaire items were used to elicit information on 7 constructs: learning process (4 items), cheating (3 items), researchability (2 items), stress (1 item), time management (1 item), alternative assessment (1 item) and invigilated examination (1 item) in comparison to proctored/invigilated examinations taken in the previous semesters (Adesile et al., 2016; McCabe & Trevino, 1993; Ramdani, 2018). Each item was measured on a five-point Likert scale ranging from 'strongly disagree' to 'strongly agree' with scores from 1 to 5 correspondingly. For the purposes of reducing common method bias (e.g., consistency pattern, response sequence bias and social desirability), we employed common remedies (e.g., avoided ambiguous question items and informed participants of response anonymity and confidentiality) (Podsakoff, MacKenzie & Podsakoff, 2012).

Study design and participants

A cross-sectional study design was employed for this study in which a self-reported survey instrument was administered to 380 Nursing and Social Science students (both undergraduate and postgraduate) in an Australian public university from August to December 2020. Ethics approval was obtained from the prospective university. A purposive sampling scheme was used to select study participants. Nursing and Social Science students were purposively sampled for this study because they mostly use invigilated examinations in their end of semester assessments prior to COVID-19 and their learning outcomes have strong focus on the development of strong evidence-based skills (research skills). Invitation emails were sent to students in these two schools across multiple courses by course/program directors and ADTLs. The inclusion criteria included all students who had undertaken alternative assessments in the 2020 academic year. The exclusion criteria included any student who had not engaged in an alternative assessment in 2020.

Table 1 briefly presents the demographic profile of study participants. Frequencies and percentages highlight the distributions across levels of the demographics for the measures reported. The majority (55%) of the participants were nursing students, mainly undertaking undergraduate (83.9%) studies in their first (29.2%) and second (34.2%) year. Alternative assessments were mostly completed within a 24-hour window (48.2%), followed by those scheduled to be completed between 2 to 4 hours (38.7%).

Table 1: Demographic characteristics of study participants, number of units enrolled and time frame for alternative assessment.

Variable	**N (%)**
Age	
18–24	112 (29.5)
25–34	94 (24.7)
35–44	73 (19.2)
45–54	40 (10.5)
55 & above	11 (2.9)
(Missing)	50 (13.2)

Course of study	
Social sciences	129 (33.9)
Nursing	209 (55.0)
(Missing)	42 (11.1)
Level of study	
Undergraduate	319 (83.9)
Postgraduate	16 (4.2)
(Missing)	45 (11.8)
Stage of study	
Year 1	111 (29.2)
Year 2	130 (34.2)
Year 3	75 (19.7)
Postgrad	22 (5.8)
(Missing)	42 (11.1)
Number of units enrolled	
1	62 (16.3)
2	80 (21.1)
3	81 (21.3)
4 or more	113 (29.7)
(Missing)	44 (11.6)
Time frame for alternative assessment	
2–4 hrs	147 (38.7)
24 hours	183 (48.2)
48 hours	48 (12.6)
1–2 week(s)	19 (5.0)
Other	61 (16.1)

Data analysis

The Shapiro–Wilk and Henze–Zirkler tests were used to test univariate and multivariate normality respectively. Sample adequacy and homogeneity of variance across samples were assessed using Kaiser–Meyer–Olkin (KMO) statistic and Bartlett's test of sphericity. Cronbach's alpha was used to investigate the reliability of the items for the various constructs considered. The average variance extract (AVE), Heterotrait–Monotrait Ratio (HTMT), and Fornell–Larcker criterion (LFC) were used to assess the convergent and discriminant validity of the instrument. The partial least square structural equation model (PLS-SEM) was used to estimate the conceptual model. SEM was applicable in establishing cause-and-effect outcomes (Afrifa-Yamoah, 2016). Model fit was reported using the chi-square test, comparative fit index (CFI), root mean square error of approximation (RMSEA) and Tucker–Lewis index (TLI). We further calculated the composite reliability (CR) statistics to establish the construct validity or otherwise of the instrument used (Nunfam et al., 2022[a,b]). A probability value (p-value) of less than 0.05 was deemed statistically significant. Statistical analysis was conducted using SmartPLS 3 (Ringle, Wende & Becker, 2015). Non-parametric bootstrapping analyses were deployed to test the mediational models based on estimates obtained from 1,000 bootstrapped samples.

Results

Procedural Remedies, Reliability and Validity

The sample adequacy was confirmed based on the KMO score 0.905 > 0.5 and the dataset diverged significantly from identity matrix to enabling data reduction because Bartlett's test of sphericity had a *p*-value < 0.001. The Shapiro–Wilk test (all p-value > 0.05) revealed that the measurement items were normality distributed. Treating the measurement items as a multivariate dataset, Henze–Zirkler test (p-value > 0.05) established that they are multivariate normally distributed. The maximum likelihood estimation approach was employed for the model parameters according to the results of both univariate and multivariate tests of normality. Herman's single factor test revealed that the maximum variance explained by a single factor was 28.97% (which is less than 50%); hence, there exist no presence of common variance bias in the dataset.

Table 2 presents the standardised factor loadings for the confirmatory factor analysis model, all of which were statistically significant ($p < 0.001$). The Cronbach's α for the reliability of the instrument was 0.869. The internal consistency of the extracted domains was good with Cronbach's α and CR statistics lying between $0.70 < \alpha < 0.95$ (see Table 2). Convergent validity was achieved across all constructs with more than one item; their observed AVE scores were greater than 0.5. The Fornell–Larcker criterion (FLC) scores indicate that the inter-construct correlation estimates were all lower than the square root of the respective AVE for the domains (see main diagonal of FLC matrix in Table 3). Also, in Table 3, the HTMT criteria scores for each pair of constructs based on the item correlations are all less than 0.85 (Henseler, Ringle, and Sarstedt, 2015). The assessment indexes indicated that appropriate levels were achieved by the measuring instrument in terms of reliability and construct validity. Additionally, the measurement model had a good fit ($\chi^2(343) = 138.68$, p-value = 0.248), as indicated by the CFI (CFI = 0.894), TLI (TLI = 0.901), RMSEA (RMSEA = 0.04) and standardised root mean square residual (SRMR = 0.07).

Table 2: Bootstrapped Measurement Items Average Loadings and Construct Validity and Reliability Assessments.

Constructs and their respective items	**Loadings**
Learning process (LP) (Cronbach's alpha = 0.922, CR = 0.944, AVE = 0.809)	
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? I feel that my critical thinking skills improved more with the alternative assessment, than invigilated exam	0.918***
To what extent do you agree or disagree with the following statements? The alternative assessment increased my confidence in solving real-world problems	0.908***
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? I feel that I learned more in the process of preparing and executing the alternative assessment in comparison to an invigilated exam	0.878***
To what extent do you agree or disagree with the following statements? I believe I learned more than I would normally learn for an invigilated exam	0.898***
Cheating (CH) (Cronbach's alpha = 0.737, CR = 0.682, AVE = 0.512)	
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? I seek help from my peers during alternative assessment	0.563*
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? I collaborate with others during alternative assessment	0.884**
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? I seek help from experts/professional/senior colleagues	0.694**

Researchability (RS) (Cronbach's alpha = 0.713, CR = 0.856, AVE = 0.753)	
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? Alternative assessment challenged my ability to research relevant materials within a short time	0.753***
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? The alternative assessment enhanced my research skills	0.963***
Stress (ST)	
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? Invigilated exams cause more stress than an alternative assessment	1
Time management (TM)	
Compared with an invigilated examination, to what extent do you agree or disagree with the following statements? I noticed improvement in my time management skills undertaking alternative assessment	1
Alternative assessment	
To what extent do you agree or disagree with the following statements? Alternative assessments should be encouraged as a form of end of semester assessment	1
Invigilated examination	
To what extent do you agree or disagree with the following statements? - Invigilated exams should be encouraged as a form of end of semester assessment	1

*** p-value < 0.001; * p-value < 0.05 - indicating significance of item loading; CR – Composite Reliability

Table 3: Discriminant Validity Analysis.

	Fornell–Larcker Criterion					Heterotrait–Monotrait Ratio				
	LP	CH	RS	ST	TM	LP	CH	RS	ST	TM
LP	0.899									
CH	0.410	0.716				0.551				
RS	0.377	0.369	0.868			0.372	0.527			
ST	0.326	0.478	0.335	1		0.499	0.404	0.457		
TM	0.326	0.325	0.261	0.308	1	0.511	0.435	0.383	0.396	

LP = Learning process; CH = Cheating; RS = Researchability; ST = Stress; TM = Time management

Testing the Effects of Modes of Assessment on Student Outcomes

According to the SEM model in Figure 2, invigilated examinations were found to have negative direct effects on researchability, learning process and stress. Of these effects, invigilated examination had a significant direct effect on perceived stress (critical ratio [CR] = 22.192, $p < 0.001$). Conversely, invigilated examinations had a positive direct effect on time management and cheating, although the relationships were not statistically significant ($p > 0.05$) (see Table 3).

Conversely, alternative assessment had significant positive direct effects on researchability (CR = 10.025, $p < 0.001$), learning process (CR = 48.482, $p < 0.001$), perceived stress (CR = 9.740, $p < 0.001$) and time management (CR = 17.414, $p < 0.001$). However, the positive direct effect on cheating (CR = 1.330, $p = 0.251$) was not significant (see Table 3). Apart from cheating, alternative assessment had a greater perceived positive direct effect on learning process (0.895 v. –0.002), researchability (0.599 v. –0.034), stress (0.563 v. –0.183) and time management (0.741 v. 0.040) compared with invigilated examination.

The amount of variability in the endogenous variables explained by the exogeneous variables (modes of assessment) were assessed to highlight the perceived combined predictive effect of invigilated examination and alternative towards students' overall learning process, researchability, stress, time management and cheating. The modes of examination explained 80.5% of the variability in students' learning process, 52.1% of variability in time management, 45.9% of variability in stress, 33.2% of variability in researchability and 1.4% of variability in cheating.

Table 4: Bootstrapped Regression Weights and Standard Deviations for the Fit of Path Model Across Study Constructs.

Path	Standard coefficients (SD)	Bootstrapped 95% CI	p-value
Alternative assessment			
⟶ Learning process	0.895 (0.019); t = 48.429	[0.858, 0.931]	**< 0.001**
⟶ Cheating	0.085 (0.090); t = 1.1500	[–0.153, 0.204]	0.251
⟶ Researchability	0.599 (0.053); t = 10.448	[0.440, 0.649]	**< 0.001**
⟶ Stress	0.563 (0.053); t = 10.558	[0.451, 0.661]	**< 0.001**
⟶ Time management	0.741 (0.042); t = 17.518	[0.642, 0.812]	**< 0.001**
Invigilated examination			
⟶ Learning process	–0.002 (0.027); t = 0.083	[–0.063, 0.045]	0.934
⟶ Cheating	0.101 (0.121); t = 1.081	[–0.236, 0.245]	0.280
⟶ Researchability	–0.034 (0.053); t = 0.653	[–0.145, 0.068]	0.514
⟶ Stress	–0.183 (0.052); t = 3.591	[–0.290, –0.088]	**< 0.001**
⟶ Time management	0.040 (0.044); t = 0.938	[–0.051, 0.118]	0.349

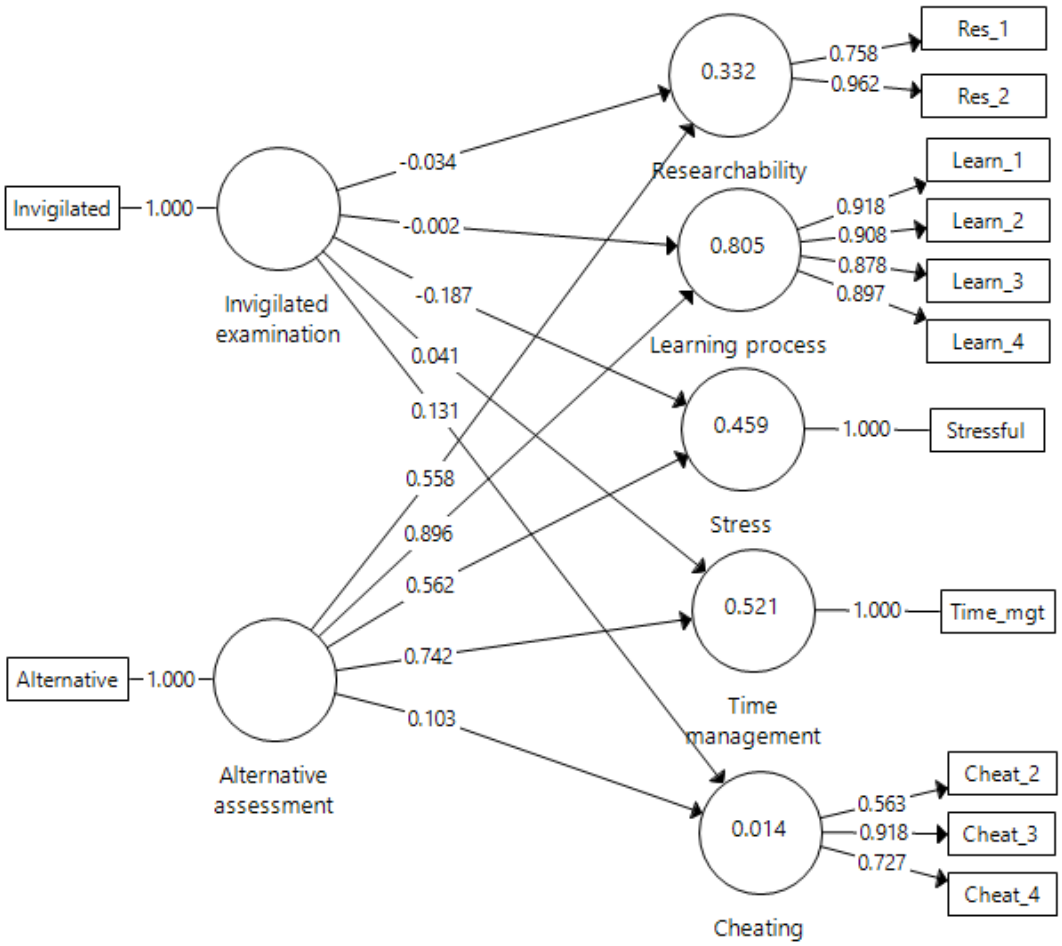

Figure 2: Standardised Estimates of the Path Analysis Model Evaluating the Perceived Relationships Between Two Different Modes of Assessment, Academic Integrity, Mental Wellbeing and Learning Outcomes

Discussion

Assessment is a necessary aspect of higher education aimed at evaluating prescribed learning outcomes and relevant competencies (Munna, 2021). Consequently, the design and practice of assessment need to be meaningful and effective to demonstrate these outcomes. This study aimed to provide an empirical evaluation of students' perceived impact of assessment practices (invigilated examination and alternative assessments) on mental wellbeing (measured by stress and time management), learning outcomes (measured by learning process and researchability) and academic integrity (measured by cheating). Path analysis modelling framework was employed to establish the significance or otherwise of the effects of invigilated examination and alternative assessments on the study parameters. Overall, the study demonstrates students' overwhelming preference for alternative assessments, highlighting the significant positive effects of these assessment practices compared with invigilated examinations on stress levels, research skills, learning process and time management ($p < 0.05$). Although, invigilated examinations were perceived more to curb cheating. However, the direct effects of both assessment practices on academic misconduct were non-significant ($p > 0.05$). As the assessment landscape of higher education has changed in recent times (greatly driven by the COVID-19 pandemic), alternative assessments have been perceived and lend as viable assessment practices with great potential to promote overall student success and productivity.

The current study emphasises the important role of assessment practices in affecting students' engagement in learning. Assessment practices significantly drive students' approaches to learning (Agarwal 2019; Jones et al., 2009; Vey, 2005). The results suggest that students perceive invigilated examination to have a negative backwash effect on learning, as a negative direct effect was observed. Assessments are generally designed to emphasise understanding of concepts; however, their form-related dimensions such as assessment type, practice and task types can have a countereffect on learning. For instance, timed invigilated assessments (irrespective of the question types used) are generally perceived to promote rote learning, which is unproductive for students' overall academic and personal growth. The design and practice of assessments should be aimed at developing students' critical argumentation, research and dialogic skills, which translates into critical thinking and deep understanding of complex issues (Deane & Song, 2014). These are crucial elements that support students' autonomy including their propensity to take meaningful initiatives and the broader skills of self-regulation (William and Flora Hewlett Foundation Assessment for Learning Working Group, 2018). Alternative assessment was perceived by students to have a positive backwash effect on their learning process; we observed positive direct effects on learning process and research skills. Effectively, alternative assessments are more likely to positively affect students' dispositions and mindsets towards learning because these practices provide greater flexibility and opportunities to create systems of assessments that students may perceive as useful and valuable, thereby stimulating the use of deeper learning approaches (Gijbel & Dochy, 2006).

Tellingly, the study showed that higher-order learning skills comes a reduction in students' stress levels. The positive impact of alternative assessment on mental wellbeing was confirmed in our study, conducted at a time when the world was experiencing and continuing to experience high levels of mental health difficulties, such as stress and anxiety. Our findings theoretically resonate with Nsor-Ambala (2020) and Graf et al., (2023), who reported the impact of assessment practices on the mental wellbeing of students. Additionally, the reduced stress could be explained in the work of Lynam and Cachia (2018), who found that students derive greater pleasure in completing assessments when assessments require the application of skills and knowledge to solve real-world problems. Conversely, students perceived invigilated/proctored assessment as having a negative direct effect on their mental wellbeing (predominantly stress). The findings of this study are timely for higher education in the post-

COVID-19 era, in which students' mental health remains a priority. The dual benefit of higher-order learning and enhanced students' mental wellbeing suggests that alternative assessment could be the best practice in higher education assessment designs.

Furthermore, the data from our study suggest increased time management skills with alternative assessment. The perceived impact of alternative assessment on students' time management skills may be owing to the longer assessment time (four hours to one week) or the fact that students undertook the assessments in online settings in the comfort of their homes instead of travelling to the university examination hall. Additionally, the observed time management skills could be a result of students' increasing self-directed learning stimulation associated with alternative assessments.

We found no statistical significance of the direct effects of both invigilated examination and alternative assessments on students' propensity to cheat. Academic integrity is fundamental to students' achievement of learning objectives and the award of degrees (Gamage et al., 2020). Literature lends to the idea that academic misconduct is not focused on whether an assessment is proctored or non-proctored but to the assessment itself and cheating is less likely to occur with more authentic assessments (Reedy, 2021). Additionally, integrity is more likely to be upheld when there are consequences for cheating (Adama et al., 2023). Universities are moving towards increasing education on academic integrity and investing in technologies for detecting cheating and misconducts. Technological supports are also available to help establish most forms of academic misconduct, such as, plagiarism, collusion, contract cheating and fabrication. Furthermore, universities are imposing more severe penalties for cases of academic misconduct, including cheating. Ultimately, it becomes a worthless risk (Adama et al., 2023). Walsh et al. (2021) advocated for the replacement of closed-book examinations with open-book exams to provide the opportunity to administer higher-order questions. Open-book exams are synonymous to alternative assessments, which comes in variety of forms.

Limitations

The design and scope of our research had some limitations as with most empirical studies. Our sample was less than 400 and respondents were from a given university context. Thus, our study has limited generalisability and the results must be interpreted with caution. Furthermore, the respondents were from the disciplines of social science and nursing. While the findings from our study are theoretically transferable, the claims of alternative assessments in this paper may not reflect the perspectives of students from other disciplines and academic institutions. Further studies are required to examine the views of students from other disciplines.

Conclusion

Based on the findings in this study, the key take-home messages are:

- Students perceive to learn best with alternative assessments because there were observed positive significant associations between alternative assessments and students' learning process and research skills.
- Students perceive mental wellbeing to be positively associated with alternative assessment. Compared with invigilated examinations, alternative assessments are perceived to be less stressful and promote good time management.
- Non-significant positive associations were observed between cheating and both alternative assessments and invigilated examination.

These findings highlight that, although the application and popularity of alternative assessment practices were mainly driven by the COVID-19 pandemic, these assessment practices are well perceived by students and showcase overwhelming preference in the extension of their utilisation in the post-COVID era.

Acknowledgements

The authors would like to acknowledge the contribution of Dr. Fiona Foxall, Associate Dean Teaching and Learning, School of Nursing and Midwifery for her valuable feedback on the initial survey questionnaire and Prof Clive Barstow, Associate Dean Teaching and Learning, School of Arts and Humanities, for his support. They would also like to express their gratitude towards the students who participated in the study. We are also and grateful to the reviewers for their time and feedback which helped us to improve the quality of the manuscript.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

References

Adama, E., Graf, A., Adusei-Asante, K., Afrifa-Yamoah, E. (2023). Covid-19 and alternative assessments in higher education: implications for academic integrity among Nursing and Social Science Students. *International Journal for Educational Integrity*. 19(8), 1-19. https://doi.org/10.1007/s40979-023-00129-0

Adesile, I., Nordin, M. S., Kazmi, Y., & Hussien, S. (2016). Validating academic integrity survey (AIS): An application of exploratory and confirmatory factor analytic procedures. *Journal of Academic Ethics*, 14(2), 149-167. https://doi.org/10.1007/s10805-016-9253-y

Afrifa-Yamoah, E. (2016). Achievement motivation as a function of participation, strive, willingness to work and maintaining work: application of structural equation modelling (SEM). *International Journal of Psychology and Behavioural Sciences*, 6 (3), 133-138. https://doi.org/10.5923/j.ijpbs.20160603.06

Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology, 111*(2), 189. https://doi.org/10.1037/edu0000282

Alessio, H. M., & Messinger, J. D. (2021). Faculty and student perceptions of academic integrity in technology-assisted learning and testing. *Frontiers in Education*. 6:629220. https://doi.org/10.3389/feduc.2021.629220

Alhasani, M., Alkhawaji, A., & Orji, R. (2022). Mental Health and Time Management Behavior among Students During COVID-19 Pandemic: Towards Persuasive Technology Design. *Human Behavior and Emerging Technologies*, https://doi.org/10.1155/2022/7376748

Benito, Á., Dogan Yenisey, K., Khanna, K., Masis, M. F., Monge, R. M., Tugtan, M. A., Vega Araya, L. D., & Vig, R. (2021). Changes that should remain in higher education post COVID-19: A mixed-methods analysis of the experiences at three universities. *Higher Learning Research Communications, 11*, 4. https://doi.org/ 10.18870/hlrc.v11i0.1195

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education,* 32(3), 347-364. https://doi.org/10.1007/BF00138871

Bolshakova, K., Shmidt, E., Bolina, M., Novikova, V., & Naumenko, L. (2022). The Backwash Effect Of EGE Examination On English Language Teaching In Public Schools In Russia. In INTED2022 Proceedings (pp. 256-260). IATED.

Cardozo, L. T., Azevedo, M. A. R. d., Carvalho, M. S. M., Costa, R., de Lima, P. O., & Marcondes, F. K. (2020). Effect of an active learning methodology combined with formative assessments on performance, test anxiety, and stress of university students. *Advances in Physiology Education, 44*(4), 744-751. https://doi.org/10.1152/advan.00075.2020

Conijn, R., Kleingeld, A., Matzat, U., & Snijders, C. (2022). The fear of big brother: The potential negative side-effects of proctored exams. *Journal of Computer Assisted Learning*, 1-14. https://doi.org/10.1111/jcal.12651

Daffin Jr, L. W., & Jones, A. A. (2018). Comparing Student Performance on Proctored and Non- Proctored Exams in Online Psychology Courses. *Online learning*, 22(1), 131-145. https://doi.org/ 10.24059/olj.v22i1.1079

Daniels, L. M., Goegan, L. D., & Parker, P. C. (2021). The impact of COVID-19 triggered changes to instruction and assessment on university students' self-reported motivation, engagement and perceptions. *Social Psychology of Education, 24*(1), 299-318. https://doi.org/10.1007/s11218-021-09612-3

Dawson, P. (2021). Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education. Routledge

Deane, P. and Song, Y. (2014). A case study in principled assessment design: Designing assessments to measure and support the development of argumentative reading and writing skills. *Psicología Educativa*, 20, 99-108. https://doi.org/10.1016/j.pse.2014.10.001

Ergene, T. (2003). Effective interventions on test anxiety reduction: A meta-analysis. *School Psychology International, 24*(3), 313-328. https://doi.org/10.1177/01430343030243004

Fitzgerald, A., & Konrad, S. (2021). Transition in learning during COVID-19: Student nurse anxiety, stress, and resource support. *Nursing Forum*, 56(2), 298-304. https://doi.org/ 10.1111/nuf.12547

Gamage, K. A., Silva, E. K. d., & Gunawardhana, N. (2020). Online delivery and assessment during COVID-19: Safeguarding academic integrity. *Education Sciences*, 10(11), 301. https://doi.org/10.3390/educsci10110301

García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande-de-Prado, M. (2021). Recommendations for Mandatory Online Assessment in Higher Education During the COVID-19 Pandemic. In *Radical Solutions for Education in a Crisis Context* (pp. 85-98). Springer.

Gibson, K., & Shaw, C. M. (2011). Assessment of active learning. In *Oxford Research Encyclopedia of International Studies*.

Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: can formative assessment make a difference?. *Educational Studies*, 32(4), 399-409. https://doi.org/10.1080/03055690600850354

Graf, A., Adama, E., Afrifa-Yamoah, E., Adusei-Asante, K. (2023). Perceived nexus between non-proctored summative assessment and mental health difficulties: a cross-sectional study. *Journal of Academic Ethics*. 21, 609-623. https://doi.org/10.1007/s10805-023-09472-w

Hand, L., Sanderson, P., & O'Neil, M. (1996). Fostering deep and active learning through assessment. *Accounting Education, 5*(2), 103-119. https://doi.org/10.1080/09639289600000013

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Houghton, A.-M., & Anderson, J. (2017). Embedding mental wellbeing in the curriculum: maximising success in higher education. *Higher Education Academy,(forthcoming), 68*.

Hussein, M. J., Yusuf, J., Deb, A. S., Fong, L., & Naidu, S. (2020). An evaluation of online proctoring tools. *Open Praxis*, 12(4), 509-525. http://doi.org/10.5944/openpraxis.12.4.1113

Jones, E., Priestley, M., Brewster, L., Wilbraham, S. J., Hughes, G., & Spanner, L. (2021). Student wellbeing and assessment in higher education: the balancing act. *Assessment & Evaluation in Higher Education, 46*(3), 438-450. https://doi.org/10.1080/02602938.2020.1782344

Jones, K. O., Harland, J., Reid, J. M., & Bartlett, R. (2009). Relationship between examination questions and bloom's taxonomy. 2009 39th IEEE frontiers in education conference, 1-6. https://doi.org/ 10.1109/FIE.2009.5350598

Khan, A., & Madden, J. (2018). Active learning: a new assessment model that boost confidence and learning while reducing test anxiety. *International Journal of Modern Education and Computer Science,* 12, 1-9. https://doi.org/ 10.5815/ijmecs.2018.12.01

Lynam, S., & Cachia, M. (2018). Students' perceptions of the role of assessments at higher education. *Assessment & Evaluation in Higher Education*, 43(2), 223-234. https://doi.org/10.1080/02602938.2017.1329928

Macan, T. H., Shahani, C., Dipboye, R. L., & Phillips, A. P. (1990). College students' time management: Correlations with academic performance and stress. *Journal of Educational Psychology*, 82(4), 760-768. https://doi.org/10.1037/0022-0663.82.4.760

McCabe, D. L., & Trevino, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education*, 64(5), 522-538. https://doi.org/10.2307/2959991

Melton, R. (1996), Learning Outcomes for Higher Education: Some key issues, *British Journal of Educational Studies*, 44(4), 409-425. https://doi.org/10.2307/3121912

Munna, A. S. (2021). Assessment and Verification: A Higher Education Perspective. *Journal of education and learning (EduLearn)*, *15*(3), 425-431. https://doi.org/10.11591/edulearn.v15i3.20080

Nayak, S. G. (2019). Impact of Procrastination and Time-Management on Academic Stress among Undergraduate Nursing Students: A Cross Sectional Study. *International Journal of Caring Sciences*, 12(3), 1480

Nsor-Ambala, R. (2020). Impact of exam type on exam scores, anxiety, and knowledge retention in a cost and management accounting course. *Accounting Education*, 29(1), 32-56. https://doi.org/10.1080/09639284.2019.1683871

Nunfam, V., Afrifa-Yamoah, E., Akanganngang , JA. (2022[a]). Mediation roles of pedagogical approaches and personality traits in entrepreneurial curriculum design and entrepreneurial intention nexus. *Entrepreneurship Education*, 1-19. https://doi.org/10.1007/s41959-022-00072-0

Nunfam, V. F., Asitik, A. J., and Afrifa-Yamoah, E. (2022). Personality, Entrepreneurship Education and Entrepreneurial Intention Among Ghanaian Students. *Entrepreneurship Education and Pedagogy*, 5(1), 65-88. https://doi.org/10.1177/2515127420961040

Parpala, A., Lindblom-Ylänne, S., Komulainen, E., Litmanen, T., & Hirsto, L. (2010). Students' approaches to learning and their experiences of teaching-learning environment in different disciplines. *British Journal of Educational Psychology*, 80(2), 269-282. https://doi.org/ 10.1348/000709909X476946

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2020). The impact of stress on students in secondary school and higher education. *International Journal of Adolescence and Youth, 25*(1), 104-112. https://doi.org/10.1080/02673843.2019.1596823

Pitoyo, M. D. (2019). Gamification based assessment: A test anxiety reduction through game elements in Quizizz platform. *IJER (Indonesian Journal of Educational Research), 4*(1), 22-32.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539-569. https://doi.org/ 10.1146/annurev-psych-120710-100452

Polat, M. (2020). Investigating the Backwash Effect of Higher Education Exam (YGS) on University Students' Attitudes. International Journal of Psychology and Educational Studies, 7(3), 152-163.

Prasad, G. (2021). Evaluating student performance based on bloom's taxonomy levels. *Journal of Physics: Conference Series*, 1797 012063. https://doi.org/ 10.1088/1742-6596/1797/1/012063

Ramdani, Z. (2018). Construction of academic integrity scale. *International Journal of Research Studies in Psychology*, 7(1), 87-97. https://doi.org/10.5861/ijrsp.2018.3003

Rana, R., & Mahmood, N. (2010). The relationship between test anxiety and academic achievement. *Bulletin of Education and research, 32*(2), 63-74.

Rawlusyk, P. (2018). Assessment in higher education and student learning. *Journal of Instructional Pedagogies, 21*.

Reedy, A., Pfitzner, D., Rook, L., & Ellis, L. (2021). Responding to the COVID-19 emergency: student and academic staff perceptions of academic integrity in the transition to online exams at three Australian universities. *International Journal for Educational Integrity*, 17(1), 1-32. https://doi.org/10.1007/s40979-021-00075-9

Reisenwitz, T. H. (2020). Examining the necessity of proctoring online exams. *Journal of Higher Education Theory and Practice*, 20(1), 118-124.

Ringle, C. M., Wende, S., & Becker, J-M. (2015). SmartPLS. Retrieved from https://www.smartpls.com

Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation, 70*, 101030. https://doi.org/10.1016/j.stueduc.2021.101030

Stankovska, G., Dimitrovski, D., Angelkoska, S., Ibraimi, Z., & Uka, V. (2018). Emotional Intelligence, Test Anxiety and Academic Stress among University Students. *Bulgarian Comparative Education Society*.

Stokking, K., van der Schaaf, M., Jaspers, J. & Erkens, G. (2004). Teachers' assessment of students' research skills. *Bristish Educarional Research Journal.* 30 (1). 93-116.

Varble, D. (2014). Reducing Cheating Opportunities in Online Test. *Atlantic Marketing Journal*, 3(3) , 131-149. https://digitalcommons.kennesaw.edu/amj/vol3/iss3/9

Vey, L. D. (2005). *Enhancing the relationship between learning and assessment*. Doctoral dissertation, University of Canberra.

Walsh, L. L., Lichti, D. A., Zambrano-Varghese, C. M., Borgaonkar, A. D., Sodhi, J. S., Moon, S., ... & Callis-Duehl, K. L. (2021). Why and how science students in the United States think their peers cheat more frequently online: perspectives during the COVID-19 pandemic. *International Journal for Educational Integrity*, 17(1), 1-18. https://doi.org/10.1007/s40979-021-00089-3

Wellman, G. S. (2005). Comparing learning style to performance in online teaching: Impact of Proctored v, Un-Proctored Testing. *Journal of Interactive Online Learning,* 4(1), 1-20.

William and Flora Hewlett Foundation Assessment for Learning Working Group (2018). *Five Elements for Assessment Design and Use to Support Student Autonomy*. Students at the Center: Deeper Learning Research Series. Boston, MA: Jobs for the Future.