The Dyke

Volume 19 No.1

Hosted & indexed by
Sabinet
African Journals

Research Article

Hosted & indexed by Sabinet African Journals

The Dyke
www.thedyke.msu.ac.zw
https://journals.co.za/loi/dyke

ISSN 1815-9036 (Print)
ISSN 2790-0940 (Online)

Continuous and Summative Assessments in Midwifery Education: Examining Predictive Validity and Independence in Eswatini

Archibald T Nyamayaro[a],Thandeka H Mhlongo[b], Isabel L Zulu[c]

[a-b] Good Shepherd Catholic College of Health Sciences and University of Eswatini.

Abstract

Assessment in nursing and midwifery education is critical for ensuring competence and progression. This study examined the relationship between continuous assessment (CA) and summative examinations among midwifery students at Good Shepherd Catholic College of Health Sciences in Eswatini (N = 30). Using total population sampling and secondary data from multiple modules, the study employed correlation and regression analyses in SPSS v23. Results indicated moderate positive correlations (r = .20–.61) between CA and examination scores, with an average r ≈ .40. Regression models yielded low adjusted R^2 values, suggesting that CA explained only a small portion of the variance in exam performance. Findings highlight inconsistencies between CA and summative outcomes, raising questions about assessment design, academic integrity, and predictive validity. The study recommends strengthening assessment frameworks by aligning them with Bloom's taxonomy, implementing standardised procedures, and providing faculty development. This contributes to debates on balancing formative and summative evaluation in health professions education.

Keywords: continuous assessment, summative assessment, predictive validity, midwifery education, Eswatini

CONTACT: Archibald Nyamayaro archibaldnyamayaro@gmail.com

Introduction

Assessment sits at the heart of health professions education, serving as the mechanism through which competence, readiness, and progression are judged. In nursing and midwifery, where the stakes are exceptionally high, given the potential consequences of inadequate preparation for patient care, the integrity of assessment systems is non-negotiable. Two principal modes of assessment dominate: continuous assessment (CA), which is distributed throughout the semester, and high-stakes summative examinations, which traditionally define outcomes. The challenge for educators and policymakers is ensuring that these modes are not only complementary but also valid, reliable, and fair measures of student competence.

CA has been widely championed as a pedagogical tool that enhances learning by offering iterative feedback loops, sustaining student engagement, and cultivating reflective practice (Wiliam & Leahy, 2024). In principle, it allows instructors to identify weaknesses in real time, intervene, and provide students with opportunities for incremental growth. In practice, however, concerns have emerged regarding the degree to which CA results genuinely reflect underlying competence, especially when juxtaposed with examination performance (Mekonen & Fitiavana, 2021).

Summative examinations, in contrast, serve as standardised evaluators of cumulative knowledge, widely viewed as impartial "gatekeepers" of academic progression (French et al., 2023). However, examinations themselves are not without limitations. They provide a snapshot of knowledge under time pressure, and critics argue they risk over-emphasising rote memorisation at the expense of sustained competence (Shepard, 2021). The tension between formative and summative assessment lies not merely in their format but in the credibility of what they measure: are they aligned to learning objectives, and can they jointly provide a reliable picture of learner capability?

The predictive validity of CA for exam performance remains contested. Some scholars find strong associations, suggesting that CA can function as a robust predictor of exam outcomes (Pudaruth et al., 2013). Others note weak or inconsistent relationships, pointing to misalignment, inflation of CA marks, or integrity issues such as copying, retakes, or inflated grading (Ukwueze, 2012). This divergence calls for context-specific studies that interrogate how these assessments function within different educational and cultural environments.

In Eswatini, midwifery education carries urgency. The health system grapples with maternal and neonatal challenges, and the competence of graduating midwives directly influences health outcomes (Makule & Kibusi, 2020). In this system, CA contributes 40% to the final grade, while summative examinations account for 60%. This weighting reflects the dual imperative of encouraging continuous learning while retaining the rigour of a final evaluative hurdle. However, the extent to which CA can predict or complement exam performance in Eswatini's context has not been empirically established.

The stakes are high: if CA and exam results diverge significantly, it raises questions about fairness, credibility, and whether assessments are fulfilling their intended function. Over-reliance on CA without integrity safeguards risks grade inflation and false assurance of competence. Over-reliance on summative examinations risks neglecting the formative, diagnostic role of assessment in guiding learning. In either case, students, faculty, and ultimately patients bear the consequences.

This study, therefore, interrogates the relationship between CA and examinations in midwifery education in Eswatini, asking whether CA can reliably predict summative outcomes and what this reveals about the design and integrity of assessment frameworks. By doing so, the research not only contributes to the local discourse but also to broader debates in health professions education regarding the balance, validity, and educational purpose of assessments.

Building upon theoretical frameworks of educational assessment, two higher-order propositions informed this study. First, it was hypothesised that continuous assessment (CA) and summative examination scores would demonstrate a significant and positive relationship, reflecting not only statistical correlation but also underlying construct alignment. This position is supported by work on assessment validity, which emphasises that convergence between formative and summative tasks is evidence that both capture core competencies (Messick, 1995; Maki, 2023). From a Bloom's taxonomy perspective, if CA tasks are effectively designed to capture progressive levels of cognitive engagement, from lower-order recall to higher-order synthesis, they should resonate with the competencies assessed in high-stakes examinations (Anderson & Krathwohl, 2001). Studies have shown that strong alignment between formative and summative assessments enhances construct validity and strengthens the interpretive value of assessment outcomes (Chufama & Sithole, 2021; Wiliam & Leahy, 2024). In other words, a positive association between CA and

examination outcomes would affirm that both measures are anchored in the same educational objectives and contribute to a coherent system of evaluation (French, Dickerson, & Mulder, 2023)

Second, it was hypothesised that CA would serve as a meaningful predictor of examination performance across modules, signalling its potential as an authentic measure of learning rather than a mere procedural requirement. This proposition is consistent with theories of formative assessment, which stress its role in scaffolding higher-order thinking and supporting transfer of learning (Black & Wiliam, 2018; Shepard, 2021). If predictive validity is observed, it would confirm CA's capacity not only to monitor but also to forecast summative achievement, thereby reinforcing its role within holistic assessment systems (Morales, Salmerón, Maldonado, Masegosa, & Rumí, 2022; Titova, 2022).

Conversely, weak predictive power would echo findings from other contexts where CA inflation, poor task design, or lack of alignment with Bloom's higher-order objectives undermined its reliability as a forecasting tool (Ukwueze, 2012; Makuvire, Mufanechiya, & Dube, 2023). In such cases, critical questions arise concerning the educational principles underpinning CA, the professional development of lecturers in assessment design, and the institutional safeguards necessary to ensure integrity (Kubiszyn & Borich, 2024; Lazareva & Agostini, 2024).

Literature Review

The current study is underpinned by two complementary frameworks: Bloom's taxonomy of educational objectives and the broader construct of assessment validity theory. Together, these frameworks provide a conceptual lens for interrogating the relationship between continuous assessment (CA) and summative examination outcomes in terms of alignment, predictability, and educational value. By situating CA and summative examinations within these theoretical traditions, the study not only investigates statistical relationships but also addresses the deeper question of whether both forms of assessment authentically measure the knowledge, skills, and professional dispositions required in midwifery education.

Bloom's taxonomy, first introduced in 1956 and later revised by Anderson and Krathwohl (2001), remains one of the most influential frameworks for categorising learning outcomes. It positions cognitive processes along a hierarchy, remembering, understanding, applying, analysing, evaluating, and creating, thereby encouraging educators to move beyond rote memorisation towards the

cultivation of higher-order thinking skills. In educational assessment, Bloom's taxonomy has become a tool for ensuring that assessment design aligns with intended learning objectives (Anderson & Krathwohl, 2001; Maki, 2023).

Within nursing and midwifery education, this alignment is critical. Clinical competence cannot be reduced to factual recall; it requires the ability to integrate knowledge, apply it in high-stakes contexts, and exercise sound judgement under pressure (Shepard, 2021). For example, while memorising the stages of labour is important, safe midwifery practice requires the ability to analyse patient data, evaluate risks, and decide appropriate interventions. Thus, CA and summative examinations must both align with Bloom's taxonomy to ensure they measure these competencies (French, Dickerson, & Mulder, 2023).

When CA tasks are anchored primarily at the lower levels of Bloom's taxonomy, such as quizzes testing recall of definitions or short-answer assignments, their predictive capacity for summative examinations, which often target higher-order domains like analysis and evaluation, is limited. This misalignment risks generating a weak or inconsistent relationship between CA and examination results, not because students are ill-prepared, but because the two assessment forms measure different levels of cognition (Maki, 2023; Morris, Perry, & Wardle, 2021). Empirical research reinforces the importance of scaffolding CA tasks across Bloom's levels. Li and Zhang (2021), in a meta-analysis of self-assessment and language testing, showed that when formative tasks emphasised application and critical engagement, students performed better in summative tests. Wiliam and Leahy (2024) similarly found that when CA incorporated problem-solving tasks and collaborative projects, it correlated strongly with examination outcomes. Conversely, when CA was restricted to rote or procedural exercises, it inflated student confidence but showed weak predictive value.

In health sciences education, simulation-based CA has emerged as a powerful tool. Case-based discussions, role-play, and clinical simulations map directly onto Bloom's higher levels, analysis, evaluation, and creation, making them strong predictors of summative outcomes (McCarthy et al., 2018; Makule & Kibusi, 2020). These authentic forms of CA better prepare students for summative assessments that often include applied case scenarios, while also equipping them with competencies essential for clinical practice. Bloom's taxonomy, therefore, provides a crucial interpretive framework: it reminds educators that the quality and level of CA tasks matter as much as their frequency. In Eswatini's midwifery programmes, if CA remains largely recall-based while examinations require critical decision-making, the CA–exam relationship will naturally be weaker.

By contrast, when CA tasks are designed to mirror examination demands across cognitive levels, predictive validity should be strengthened.

While Bloom's taxonomy focuses on cognitive alignment, construct validity theory provides a broader psychometric foundation for evaluating whether assessments measure what they claim to measure. Messick (1995) conceptualises validity not as a property of a test itself, but of the inferences drawn from test scores. Thus, validity requires evidence that the interpretations and uses of scores are appropriate, meaningful, and supported by theory. Applied to the CA-examination relationship, construct validity is demonstrated when both forms of assessment measure the same underlying competencies: integration of knowledge, problem-solving, professional judgement, and the ability to transfer learning to clinical settings. If CA and examination scores align, we can infer that both capture overlapping constructs. Conversely, if correlations are weak, questions arise about whether CA is measuring superficial engagement rather than authentic competence.

Construct validity can be compromised by several factors. First, lenient grading in CA, often driven by institutional pressures to improve pass rates, inflates scores and undermines their comparability with examinations (Makuvire, Mufanechiya, & Dube, 2023). Second, opportunities for retakes or extended deadlines, while pedagogically supportive, dilute the discriminative power of CA (Morales, Salmerón, Maldonado, Masegosa, & Rumí, 2022). Third, lack of standardisation across instructors and modules creates inconsistency in CA practices, further weakening its validity (Lazareva & Agostini, 2024). In contrast, examinations, though criticised for being high-stakes and stress-inducing, offer standardisation and comparability. Examinations are typically administered under controlled conditions, with marking guided by rubrics or moderation, reducing the scope for inflation or inconsistency (French et al., 2023). This explains why, despite their flaws, examinations remain dominant in high-stakes decision-making in health professions education worldwide. The strength of the CA–exam relationship, therefore, hinges on whether CA practices uphold validity standards. If CA tasks are authentic, well-standardised, and aligned with learning outcomes, they can meaningfully predict examination performance. If not, CA risks devolving into a procedural exercise divorced from real competence, undermining its credibility as an evaluative tool (Ukwueze, 2012).

Recent scholarship has emphasised the need to integrate cognitive alignment and construct validity in the design of assessment systems. Morales et al. (2022)

demonstrated that CA in Spanish universities had strong predictive validity only when tasks were deliberately aligned with intended competencies, consistent with Bloom's higher levels. When CA was treated as a procedural requirement, with little attention to validity, correlations with examinations were weak. Similarly, Lazareva and Agostini (2024) warn that, as higher education shifts towards alternative assessments, failure to anchor them in valid constructs risks reducing them to fragmented, unreliable measures.

In Eswatini, midwifery education integration is particularly important. Bloom's taxonomy underscores the need to design CA tasks that develop and assess higher-order competencies, while construct validity theory demands evidence that CA and examinations measure the same constructs. By combining these frameworks, this study frames the CA–exam relationship not simply as a statistical question, but as a matter of educational coherence and professional accountability. The dual frameworks also illuminate the stakes of assessment reform in Eswatini's midwifery education. First, they highlight the risk of misalignment: if CA tasks are simplistic and examinations demand complex decision-making, students may succeed in CA but falter in exams, eroding trust in the fairness of the system. Second, they draw attention to issues of integrity: inflated CA scores may mask deficiencies, leading to graduates who are underprepared for the rigours of professional practice. This has profound implications for patient safety, maternal health outcomes, and the reputation of training institutions (Makule & Kibusi, 2020).

At the same time, the frameworks point to a pathway forward. By redesigning CA to scaffold Bloom's higher levels, and by instituting validity safeguards, such as moderation, standardised rubrics, and authentic assessment tasks, institutions can strengthen both the predictive validity of CA and the credibility of their programmes. In doing so, they align with global trends towards more holistic, competency-based assessment in health professions education (Gamage, Pradeep, & de Silva, 2022; Wiliam & Leahy, 2024). The theoretical framework thus situates the CA–exam relationship within two powerful traditions: cognitive alignment and construct validity. Bloom's taxonomy provides the pedagogical rationale for expecting CA to scaffold learning towards competencies tested in summative examinations. Validity theory provides the psychometric rationale for testing whether CA and exams authentically measure the same constructs. When combined, they frame the investigation as both a statistical inquiry and an educational evaluation.

For Eswatini's midwifery education system, this dual lens is especially urgent. In a context where midwives are central to reducing maternal and neonatal mortality, the stakes of assessment integrity extend beyond academic progression; they impact lives. By applying Bloom's taxonomy and construct validity theory, this study not only measures correlations but also asks whether current assessment practices serve their ultimate purpose: ensuring competent, reflective, and safe midwifery professionals

Methodology

This study employed a quantitative, correlational design to examine the statistical relationship between continuous assessment (CA) and examination performance among midwifery students. Correlational designs are commonly used in education to determine whether and to what extent variables are related (Creswell & Creswell, 2018). The quantitative approach was chosen for its ability to provide precise, replicable measurements of relationships while minimising researcher bias (Luoma & Hietanen, 2024). The study population comprised all 30 students enrolled at Good Shepherd Catholic College of Health Sciences (Eswatini) during the 2023/24 academic year. Due to the small cohort, a total population sampling method was utilised to enhance representativeness and internal validity, a practice particularly suitable in health sciences education where small cohorts are typical (Etikan & Bala, 2023; Makule & Kibusi, 2020). Data were obtained from official departmental records covering 15 theoretical and clinical modules, including obstetrics, neonatology, community health, and professional practice. Continuous assessment included quizzes, assignments, presentations, and class tests, while summative examination scores were obtained from invigilated end-of-semester examinations.

The independent variable was CA scores, and the dependent variable was examination performance, both measured as percentages. Data analysis, conducted in SPSS v23, included descriptive statistics, Pearson's correlation to assess the strength and direction of relationships, and simple linear regression to test predictive validity. Regression modelling enabled the estimation of the variance in exam scores explained by CA, as reflected in R^2 values. Prior to regression, assumptions of normality, linearity, and homoscedasticity were checked (Field, 2018). Statistical significance was set at $p < .05$. Ethical integrity was upheld by securing approval from the College Principal, anonymising student records, and ensuring confidentiality. Given the sensitive nature of assessment data, findings were framed as constructive insights for institutional

improvement rather than critiques of individual instructors (Cohen, Manion, & Morrison, 2018).

Results

The descriptive statistics reveal a significant pattern: while continuous assessment (CA) scores across all modules are relatively tightly clustered (standard deviations ranging from 3.3 to 5.6), examination scores show much greater variability (standard deviations up to 12.4). This indicates that CA offers a more consistent performance distribution, whereas examinations distinguish students more distinctly, with extremes like MWF410, where exam scores ranged from 19% to 79%.

Table 1: Descriptive Statistics

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
MWF410CA	30	56	74	64.36	4.662
MWF405CA	30	54	77	62.38	4.938
MWF409CA	30	53	70	59.30	3.633
MWF401CA	30	51	66	56.40	3.480
MWF403CA	30	56	76	66.97	5.500
MWF411CA	30	58	84	72.85	5.351
MWF407CA	30	52	76	65.60	5.563
MWF413CA	30	54	74	61.80	4.634
MWF414CA	30	59	77	66.90	4.559
MWF404CA	30	63	76	70.40	3.349
MWF408CA	30	58	73	65.93	3.648
MWF406CA	30	58	75	65.17	4.276
MWF412CA	30	56	74	63.93	4.675
MWF402CA	30	49	65	57.70	3.650
MWF402Exam	30	33	64	50.27	9.329
MWF415CA	30	61	81	71.43	5.083
MWF415Exam	30	52	81	64.50	6.991
MWF412Exam	30	50	78	62.33	6.671
MWF406Exam	30	50	72	62.17	5.983
MWF408Exam	30	46	83	62.70	9.006
MWF404Exam	30	45	74	60.67	7.902
MWF414Exam	30	50	75	64.50	5.976
MWF413Exam	30	51	77	65.03	6.494
MWF407Exam	30	51	79	66.97	6.851
MWF411Exam	30	42	83	70.60	7.837
MWF403Exam	30	52	79	64.87	7.660
MWF401Exam	30	47	73	61.10	7.203
MWF409Exam	30	41	68	55.00	7.799
MWF405Exam	30	46	77	65.33	6.784
MWF410Exam	30	19	79	57.10	12.391
Valid N (listwise)	30				

Table 2 shows that the correlation between continuous assessment (CA) and examination scores for MWF410 is moderate and positive ($r = .392$, $p = .032$).

Table 2: Correlation between continuous assessment (CA) and examination scores

Correlations

		MWF410CA	MWF410Exam
MWF410CA	Pearson Correlation	1	.392*
	Sig. (2-tailed)		.032
	N	30	30
MWF410Exam	Pearson Correlation	.392*	1
	Sig. (2-tailed)	.032	
	N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Table 2 further shows that students who performed better in CA also tended to do better in the examination, although the relationship is not very strong. From the perspective outlined in the introduction, this finding is significant because it offers empirical support for the idea that CA and examinations are not entirely separate but assess overlapping skills. It also strengthens the literature review discussion, which highlighted mixed evidence worldwide: some studies report high predictive validity of CA (Pudaruth et al., 2013; Jena, 2019), while others show weak or inconsistent correlations due to grade inflation or misalignment (Ukwueze, 2012; Makuvire, Mufanechiya, & Dube, 2023). The moderate correlation here reflects this middle ground: CA is linked to exam results, but not enough to be the only predictor of performance.

The correlation results for MWF405 (Table 3) show a weak positive association between continuous assessment (CA) and examination performance ($r = .226$, $p = .230$).

Table 3: The correlation results for MWF405

Correlations

		MWF405CA	MWF405Exam
MWF405CA	Pearson Correlation	1	.226
	Sig. (2-tailed)		.230
	N	30	30
MWF405Exam	Pearson Correlation	.226	1
	Sig. (2-tailed)	.230	
	N	30	30

Unlike the MWF410 results, this correlation is not statistically significant at the 0.05 level, meaning the relationship might be due to chance. From the perspective presented in the introduction, this outcome emphasises the inconsistency of the CA–exam relationship across modules, implying that while CA is meant to reflect ongoing learning and predict final outcomes, its actual predictive power can vary considerably. In this context, the MWF405 findings confirm that CA does not always reliably predict summative outcomes.

The correlation results for MWF409, MWF401, and MWF411 (Table 4) show statistically significant, moderate-to-strong positive relationships between continuous assessment (CA) and examination scores. Table 4 displays the results.

Table 4: The correlation results for MWF409, MWF401, and MWF411

Correlations

		MWF409CA	MWF409Exam
MWF409CA	Pearson Correlation	1	.571**
	Sig. (2-tailed)		.001
	N	30	30
MWF409Exam	Pearson Correlation	.571**	1
	Sig. (2-tailed)	.001	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		MWF401CA	MWF401Exam
MWF401CA	Pearson Correlation	1	.503**
	Sig. (2-tailed)		.005
	N	30	30
MWF401Exam	Pearson Correlation	.503**	1
	Sig. (2-tailed)	.005	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		MWF411CA	MWF411Exam
MWF411CA	Pearson Correlation	1	.591**
	Sig. (2-tailed)		.001
	N	30	30
MWF411Exam	Pearson Correlation	.591**	1
	Sig. (2-tailed)	.001	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Based on the results shown in Table 4, MWF409 exhibits a correlation of r = .571 (p = .001), MWF401 shows r = .503 (p = .005), and MWF411 records the strongest association at r = .591 (p = .001). These findings suggest that students who perform well in CA are consistently more likely to excel in the corresponding examinations. The comparative analysis across modules (as illustrated in Table 4) highlights notable variation in the strength and significance of the relationship between continuous assessment (CA) and examination performance. MWF405 demonstrated only a weak, non-significant correlation (r = .226, p = .230), indicating that CA in this module is a poor predictor of examination outcomes. MWF410 displayed a moderate but significant association (r = .392, p = .032), reflecting partial predictive validity. Conversely, MWF409 (r = .571, p = .001), MWF401 (r = .503, p = .005), and MWF411 (r = .591, p = .001) showed moderate-to-strong, statistically significant correlations, pointing to a stronger alignment between CA and examinations in these modules. Overall, these results imply that while CA can be a reliable predictor of exam performance, its effectiveness varies across modules, possibly due to differences in assessment design, alignment, and integrity.

For MWF407 (Table 5), the correlation between continuous assessment (CA) and examination scores is weak and not statistically significant (r = .286, p = .126), indicating that CA in this module does not reliably predict examination performance. The results showed contrasting patterns across the two modules.

Table 5: The correlation results for MWF407 and MWF-413

Correlations

		MWF407CA	MWF407Exam
MWF407CA	Pearson Correlation	1	.286
	Sig. (2-tailed)		.126
	N	30	30
MWF407Exam	Pearson Correlation	.286	1
	Sig. (2-tailed)	.126	
	N	30	30

Correlations

		MWF413CA	MWF413Exam
MWF413CA	Pearson Correlation	1	.616**
	Sig. (2-tailed)		.000
	N	30	30
MWF413Exam	Pearson Correlation	.616**	1
	Sig. (2-tailed)	.000	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

In contrast, MWF413 exhibits a strong and highly significant positive correlation ($r = .616$, $p < .001$), indicating that students' CA performance closely matches their examination results. As outlined in the introduction, these differing findings reflect the contentious nature of the CA–exam relationship, and as highlighted in the literature review, they align with global debates where CA sometimes predicts summative outcomes effectively (Jena, 2019; Kubiszyn & Borich, 2024) but often falls short due to grade inflation or misalignment (Ukwueze, 2012; Makuvire, Mufanechiya, & Dube, 2023). Bloom's taxonomy offers a theoretical perspective for understanding these differences: where CA tasks in MWF413 likely engaged higher-order thinking consistent with examinations, stronger correlations emerged, whereas in MWF407, weaker task design or limited alignment may have resulted in the poor association.

Table 6 shows the correlation results for MWF414, MWF404, and MWF408.

Correlations

		MWF414CA	MWF414Exam
MWF414CA	Pearson Correlation	1	.385*
	Sig. (2-tailed)		.035
	N	30	30
MWF414Exam	Pearson Correlation	.385*	1
	Sig. (2-tailed)	.035	
	N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

		MWF404CA	MWF404Exam
MWF404CA	Pearson Correlation	1	.508**
	Sig. (2-tailed)		.004
	N	30	30
MWF404Exam	Pearson Correlation	.508**	1
	Sig. (2-tailed)	.004	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		MWF408CA	MWF408Exam
MWF408CA	Pearson Correlation	1	.363*
	Sig. (2-tailed)		.049
	N	30	30
MWF408Exam	Pearson Correlation	.363*	1
	Sig. (2-tailed)	.049	
	N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Table 6 demonstrates moderate, though statistically significant, relationships between continuous assessment (CA) and examination scores, with varying levels of strength. MWF414 produced a correlation of $r = .385$ ($p = .035$), indicating a modest but significant link, while MWF404 showed a stronger association at $r = .508$ ($p = .004$), suggesting that CA in this module is a reasonably reliable predictor of exam outcomes. MWF408 recorded $r = .363$ ($p = .049$), again statistically significant but weaker in magnitude. Taken together, these results reinforce the point raised in the introduction that the predictive validity of CA is not uniform across modules, and, as highlighted in the literature, effectiveness depends on how well CA is designed and aligned with summative assessments (Jena, 2019; Kubiszyn & Borich, 2024).

In modules like MWF404, where CA may have incorporated higher-order skills (application, analysis, evaluation), the correlation with examinations was stronger, whereas in MWF414 and MWF408, where CA tasks may have focused more narrowly on recall or comprehension, predictive validity was weaker. In this case, the construct validity theory (Messick, 1995) may explain that the meaningfulness of the scores rests on whether CA and exams truly capture the same construct of competence.

Results for MWF406, MWF412, and MWF402 (Table 7) demonstrate statistically significant positive associations between continuous assessment (CA) and examination scores, though with varying strengths. Figure 6 is the summary.

Table 7: The correlation results for MWF406, MWF412, and MWF402

Correlations

		MWF406CA	MWF406Exam
MWF406CA	Pearson Correlation	1	.465**
	Sig. (2-tailed)		.010
	N	30	30
MWF406Exam	Pearson Correlation	.465**	1
	Sig. (2-tailed)	.010	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		MWF412CA	MWF412Exam
MWF412CA	Pearson Correlation	1	.444*
	Sig. (2-tailed)		.014
	N	30	30
MWF412Exam	Pearson Correlation	.444*	1
	Sig. (2-tailed)	.014	
	N	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

		MWF402CA	MWF402Exam
MWF402CA	Pearson Correlation	1	.571**
	Sig. (2-tailed)		.001
	N	30	30
MWF402Exam	Pearson Correlation	.571**	1
	Sig. (2-tailed)	.001	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

MWF406 shows a moderate relationship (r = .465, p = .010), MWF412 exhibits a slightly weaker but still significant association (r = .444, p = .014), while MWF402 demonstrates the strongest correlation among the three (r = .571, p = .001). These findings reinforce the idea from the introduction that CA and examinations are not independent but are rather complementary measures of student competence, although the strength of the relationship varies across modules. As highlighted in the literature, such variability reflects global trends where CA sometimes reliably predicts summative outcomes but, in other contexts, shows only moderate or weak associations (Jena, 2019; Morales et al., 2022; Kubiszyn & Borich, 2024).

Table 7: The correlation results for MWF415

Correlations (r)			
		MWF415CA	MWF415Exam
MWF415CA	Pearson Correlation	1	.236
	Sig. (2-tailed)		.209
	N	30	30
MWF415Exam	Pearson Correlation	.236	1
	Sig. (2-tailed)	.209	
	N	30	30

The results for MWF415 (Figure 7) indicate a weak and statistically non-significant correlation between continuous assessment (CA) and examination scores (r = .236, p = .209). This suggests that student performance in CA within this module does not reliably predict examination outcomes, aligning with earlier findings in modules such as MWF405 and MWF407, where similar weak associations were observed. As noted earlier, weak correlations highlight ongoing concerns about the consistency and predictive validity of CA in specific contexts, particularly when tasks may be inflated, insufficiently rigorous, or misaligned with the higher-order competencies assessed in summative examinations (Ukwueze, 2012; Makuvire, Mufanechiya, & Dube, 2023).

The correlation analysis across the 15 modules reveals varying levels of association between continuous assessment (CA) and examination performance, ranging from weak to moderately strong. At the lower end, correlations around r = 0.2 indicate a weak positive relationship, where students who perform well in CA tend to perform only slightly better in examinations. In such cases, high-achieving students may become complacent after strong CA performance and reduce their study effort, while those who perform poorly in CA may compensate by increasing their examination preparation, thereby narrowing the gap. This explains why weak correlations sometimes mask cases of "surprise" outcomes, where low CA performers excel in examinations or vice versa.

Moderate correlations in the range of r = 0.3–0.4 suggest that CA has some predictive validity, with students who perform well in CA tending to maintain similar momentum into examinations. However, the moderate strength also reveals limitations: students who struggle in CA generally continue to perform poorly in examinations, indicating that CA outcomes may "lock in" performance trajectories with little room for dramatic change. By contrast, stronger correlations in the range of r = 0.5–0.6 reflect a moderately strong

positive relationship, where CA becomes a clearer predictor of examination performance. In such instances, high CA performers almost always succeed in examinations, while low CA performers are consistently at risk, leaving little room for unexpected outcomes.

Overall, the average correlation across all modules was $r = 0.4$, a moderate relationship. This finding suggests that while CA and examinations are related, CA performance does not consistently mirror examination outcomes. In practice, this means a student who performs poorly in CA can still succeed in examinations, while a student who excels in CA may underperform in final assessments. Such variability raises critical concerns for the validity of CA as a predictor of summative performance, echoing global debates on the uneven reliability of formative measures (Morales et al., 2022; Makuvire, Mufanechiya, & Dube, 2023).

Regression analysis results

The regression analysis for MWF410 shows that continuous assessment (CA) is a statistically significant but weak predictor of examination performance (See Table 8). The model yielded a correlation coefficient of R = .392, with an R^2 value of .154, indicating that about 15.4% of the variance in exam scores can be explained by CA performance. The adjusted R^2 of .123 suggests that, after considering sample size and predictor adjustment, the explanatory power is slightly lower but remains modest. The standard error of estimate (11.601) reveals a considerable amount of unexplained variation in exam scores.

Table 8: Regression analysis for MWF410

Regression (R)
Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	MWF410CA[b]	.	Enter

a. Dependent Variable: MWF410Exam
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.392[a]	.154	.123	11.601

a. Predictors: (Constant), MWF410CA

Table 9 is the ANOVA table for MWF410.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	684.115	1	684.115	5.083	.032[b]
	Residual	3768.585	28	134.592		
	Total	4452.700	29			

a. Dependent Variable: MWF410Exam
b. Predictors: (Constant), MWF410CA

Table 9 confirms that the regression model is statistically significant, with $F(1,28) = 5.083$, $p = .032$. This result shows that continuous assessment (CA) contributes meaningfully to predicting examination performance, even though its explanatory power is modest. Practically, the model indicates that CA offers some predictive validity, but a large portion of exam score variability remains unexplained by CA alone.

The coefficients table for MWF410 shows that continuous assessment (CA) is a significant predictor of examination performance, with a standardised Beta of .392 ($t = 2.255$, $p = .032$). The unstandardized coefficient ($B = 1.042$) indicates that for every one-unit increase in CA score, the examination score is expected to increase by approximately 1.04 units, holding other factors constant. The constant term (-9.948, $p = .741$) is not statistically significant, indicating that, without CA input, the model does not reliably predict exam scores.

Table 10: The regression results for MWF405

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	MWF405CA[b]	.	Enter

a. Dependent Variable: MWF405Exam
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.226[a]	.051	.017	6.726

a. Predictors: (Constant), MWF405CA

The regression results for MWF405 (Table 10) show that continuous assessment (CA) is a very weak predictor of examination performance. The model yielded a correlation coefficient of R = .226, with an R^2 value of just .051. This indicates that CA explains only 5.1% of the variance in examination scores, leaving over 94% unexplained by CA. The adjusted R^2 of .017 also suggests negligible explanatory power, given the sample size. The relatively high standard error of the estimate (6.726) reflects significant unexplained variability in student exam performance.

Table 11 shows the ANOVA results for MWF405.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	68.075	1	68.075	1.505	.230[b]
	Residual	1266.592	28	45.235		
	Total	1334.667	29			

a. Dependent Variable: MWF405Exam
b. Predictors: (Constant), MWF405CA

The ANOVA results for MWF405 confirm that the regression model is not statistically significant, F (1,28) = 1.505, p = .230. This means that continuous assessment (CA) does not significantly predict examination outcomes in this module, supporting the earlier finding that the model explained only about 5% of the variance (R^2 = .051). Essentially, students' performance in CA within MWF405 has little to no explanatory power for their performance in the final examination.

Figure 12 shows the coefficients table for MWF405.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	45.979	15.825		2.905	.007
	MWF405CA	.310	.253	.226	1.227	.230

a. Dependent Variable: MWF405Exam

The coefficients table for MWF405 confirms that continuous assessment (CA) is not a significant predictor of examination scores in this module. The standardised Beta value is .226, with t = 1.227 and p = .230, indicating the observed relationship could easily be due to chance. The unstandardised coefficient (B = 0.310) suggests that for every one-unit increase in CA, exam scores would rise

by 0.31 points, but this effect is not statistically significant. The constant term (45.979, p = .007) is significant, indicating that baseline exam performance is independent of CA input, further underlining the weak predictive contribution of CA in this module.

Table 13 examines the relationship between continuous assessment scores (MWF401CA) and final exam results (MWF401Exam).

Table 13: CS scores (MWF401CA) and final exam results (MWF401Exam).

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	MWF401CA[b]	.	Enter

a. Dependent Variable: MWF401Exam
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.503[a]	.253	.227	6.335

a. Predictors: (Constant), MWF401CA

Based on the provided regression analysis (Table 13), which examines the relationship between continuous assessment scores (MWF401CA) and final exam results (MWF401Exam), a moderate positive relationship exists (R = 0.503), indicating that students who perform better on coursework tend to do better on the exam. Statistically, the coursework score explains 25.3% of the variance in exam scores (R Square = 0.253), a figure that remains meaningful even after adjustment (Adjusted R Square = 0.227). However, this also means that the vast majority (about 75%) of what determines a student's exam performance is influenced by other factors not included in this model, such as exam anxiety, final revision effectiveness, or topic-specific understanding. With a standard error of the estimate of 6.335, indicating the typical prediction error, coursework is a relevant indicator but far from a definitive predictor, and relying on it alone would lead to many inaccurate forecasts of final exam success.

Table 14 shows the coefficients table for MWF405.

Table 14: The coefficients table for MWF405

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	381.007	1	381.007	9.494	.005[b]
	Residual	1123.693	28	40.132		
	Total	1504.700	29			

a. Dependent Variable: MWF401Exam
b. Predictors: (Constant), MWF401CA

Table 14 demonstrates that the coefficients table for MWF405 confirms that continuous assessment (CA) is not a significant predictor of examination scores in this module. The standardised Beta value is .226, with t = 1.227 and p = .230, indicating the observed relationship could easily be due to chance. The unstandardized coefficient (B = 0.310) suggests that for every one-unit increase in CA, exam scores would rise by only 0.31 points, but this effect is not statistically significant. The constant term (45.979, p = .007) is significant, indicating that baseline exam performance is independent of CA input, further underlining the weak predictive contribution of CA in this module.

Table 15 presents the regression coefficients for MWF401.

Table 15: Regression coefficients for MWF401

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.355	19.100		.123	.903
	MWF401CA	1.042	.338	.503	3.081	.005

a. Dependent Variable: MWF401Exam

Table 15 shows that the regression coefficients for MWF401 indicate that continuous assessment (CA) is a significant and moderately strong predictor of examination performance. The standardised Beta value is .503, with t = 3.081 and p = .005, indicating that the relationship is both statistically significant and educationally meaningful. The unstandardised coefficient (B = 1.042) suggests that for every one-unit increase in CA, examination scores are expected to increase by approximately 1.04 points, holding other factors constant. The constant (2.355, p = .903) is not significant, meaning baseline exam performance

without CA input cannot be meaningfully interpreted, but the strength of CA as a predictor remains clear.

Table 16 shows the regression model summary for MWF407.

Table 16: The regression model summary for MWF407

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	MWF407CA[b]	.	Enter

a. Dependent Variable: MWF407Exam
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.286[a]	.082	.049	6.681

a. Predictors: (Constant), MWF407CA

The regression model summary for MWF407 indicates that continuous assessment (CA) is a weak predictor of examination performance. The correlation coefficient is $R = .286$, with an R^2 of .082, meaning CA explains only about 8.2% of the variance in examination scores. The adjusted R^2 drops further to .049, underscoring the model's limited explanatory power after accounting for sample size. The standard error of estimate (6.681) indicates substantial unexplained variation in exam results.

This outcome reflects the uneven role of CA as a predictor of summative performance. Unlike modules such as MWF401 (where CA was a strong predictor), MWF407 shows that CA contributes very little to explaining exam outcomes, which aligns with literature that highlights problems of grade inflation, inconsistent assessment design, and misalignment between CA and exam tasks (Ukwueze, 2012; Makuvire, Mufanechiya, & Dube, 2023).

Table 17 are the ANOVA results for MWF407.

Table 17: The ANOVA results for MWF407

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	110.985	1	110.985	2.486	.126[b]
	Residual	1249.982	28	44.642		
	Total	1360.967	29			

a. Dependent Variable: MWF407Exam
b. Predictors: (Constant), MWF407CA

The ANOVA results for MWF407 confirm that the regression model is not statistically significant, $F(1,28) = 2.486$, $p = .126$. This indicates that continuous assessment (CA) does not make a meaningful contribution to predicting examination performance in this module. Although the model summary earlier showed a weak correlation ($R = .286$) and explained about 8.2% of the variance, the lack of statistical significance here means the observed relationship could be due to chance rather than a genuine predictive effect.

The results highlight the inconsistent role of CA as a predictor across different modules, echoing global findings that CA can be unreliable when misaligned with summative assessments (Morales et al., 2022; Makuvire, Mufanechiya, & Dube, 2023).

Table 18 presents the coefficients table for MWF407.

Table 18: The coefficients table for MWF407

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	43.898	14.682		2.990	.006
	MWF407CA	.352	.223	.286	1.577	.126

a. Dependent Variable: MWF407Exam

The coefficients table for MWF407 confirms the earlier regression and ANOVA results, showing that continuous assessment (CA) is not a significant predictor of examination performance in this module. The standardised Beta value is .286, with a t statistic of 1.577 and a non-significant p-value of .126, suggesting that the relationship between CA and exam outcomes is weak and not statistically meaningful. The unstandardised coefficient (B = .352) indicates that for each one-unit increase in CA, exam scores might increase by approximately 0.35 points; however, this effect is too small and statistically insignificant to support firm conclusions. Interestingly, the constant term (43.898, p = .006) is significant, implying that students' baseline exam performance is largely independent of their CA scores.

Discussion

This study examines whether CA and examinations in midwifery education are complementary, valid indicators of competence, or merely loosely related signals. Across fifteen modules, the pattern is consistent but uneven. The average correlation (≈ .40) indicates a moderate overall association, though module-level effects vary from non-significant and weak (e.g., MWF405, MWF407, MWF415) to modest (e.g., MWF410, MWF414, MWF408) and moderate-to-strong with high significance (e.g., MWF401, MWF404, MWF409, MWF411, MWF402).

Regression findings converge; CA explains a small to moderate proportion of variance (e.g., $R^2 \approx .15$ in MWF410; ≈ .25 in MWF401), but its influence is negligible in others (e.g., $R^2 \approx .05$ in MWF405). Simply put, CA can predict overall performance but not consistently. These results support the claim that the relationship between CA and exams is contested and dependent on context, reflecting the literature's mixed record. It is stronger when CA is rigorous and well-aligned, and weaker when it is inflated, inconsistent, or narrowly focused on cognitive demand (Jena, 2019; Morales et al., 2022; Kubiszyn & Borich, 2024; Ukwueze, 2012; Makuvire, Mufanechiya, & Dube, 2023).

The findings from this study sound an alarm for midwifery education in Eswatini and similar contexts. The inconsistent and often weak predictive validity of continuous assessment (CA) for summative examination performance, with R^2 values as low as .05 (MWF405) and an average explanatory power of only ~16%, is not merely a statistical observation; it is a critical indicator of a system failing to achieve construct validity. The moderate average correlation ($r \approx .40$) indicates that while CA and examinations are related, they do not measure the same underlying competencies with sufficient consistency. This misalignment, viewed through the lens of Bloom's taxonomy (Anderson & Krathwohl, 2001) and construct validity theory (Messick, 1995), suggests that CA tasks are often not scaffolding the higher-order cognitive processes, analysis, evaluation, and creation, required for safe clinical practice and robustly assessed in final examinations. The stark divergence in score variability, with exams exposing dramatic performance gaps (e.g., scores from 19% to 79%) that CA fails to predict, points to a dangerous disconnect. This is not an academic exercise; it is a patient safety issue.

When a student performs adequately on low-order CA tasks but fails catastrophically in an exam simulating clinical decision-making, the system has provided a false sense of competence. This inconsistency across modules,

where predictive strength varies wildly from negligible (MWF405, MWF407) to moderate (MWF401, MWF411), echoes global concerns about CA inflation and poor task design (Ukwueze, 2012; Makuvire et al., 2023). It reveals an ad hoc, unstandardised assessment culture in which the integrity and cognitive demands of CA are left to individual instructor discretion. Modules with stronger correlations likely feature CA better aligned with examination demands, incorporating case-based problems or simulations that target application and judgment (McCarthy et al., 2018). Those with weak correlations may rely on recall-based quizzes or assignments that are vulnerable to academic integrity issues, thereby severing the link between formative effort and the summative demonstration of skill.

Why the heterogeneity?

The theoretical framework offers a crisp explanation. Bloom's taxonomy predicts stronger CA–exam concordance when CA tasks ascend the hierarchy, application, analysis, evaluation, and creation, because those same levels are typically stressed in summative examinations (Anderson & Krathwohl, 2001; Shepard, 2021). The modules with the largest coefficients (e.g., MWF401, MWF404, MWF409, MWF411, MWF402) plausibly embedded authentic, higher-order work, case discussions, structured clinical judgments, and problem-solving assignments, thereby cultivating the exact cognitive moves exams later sampled.

Conversely, weak modules likely relied on lower-order checks (recall/comprehension quizzes), resulting in compressed CA distributions and attenuated predictions. Construct validity theory (Messick, 1995) sharpens the point: validity lies in the interpretations we make from scores. If CA scores arise from leniency, multiple retakes, or poorly moderated tasks, they do not warrant the same inference ("competence achieved") that exam scores warrant; the correlation weakens because the underlying construct differs. The descriptive statistics support this: CA scores were tightly clustered (low SDs) while exam scores spread widely (higher SDs), a classic signature of restricted range and differential rigour. The result is not merely a smaller r; it is a validity warning.

Methodologically, the study's total-population sampling ($N = 30$) ensured that the full cohort signal was observed rather than a sample artefact, appropriate for a small programme (Etikan & Bala, 2023). The analytic strategy, descriptives, Pearson correlations, and simple regressions in SPSS, after checking core assumptions (Field, 2018), was proportionate to the research questions and

the data available. That CA explained only ~15–25% of variance in its best-performing modules is not a failure of method; it reflects the reality that examinations also capture variance due to anxiety regulation, time-pressured reasoning, item format familiarity, and last-mile revision, factors CA often under-samples. Equally, the non-significant models (e.g., MWF405, MWF407, MWF415) are not statistical noise; they are diagnostic signals that CA practices in those modules are either misaligned or lack integrity safeguards.

What do these results mean educationally?

Firstly, CA is necessary but insufficient. It functions well as a learning driver and as an early indicator of authenticity, cognitive demand, and consistent grading (Wiliam & Leahy, 2024; Chufama & Sithole, 2021). Where those conditions fail, CA becomes non-informative for predicting exam performance and may even mislead students and staff about readiness. Secondly, the programme's 60/40 weighting (exam/CA) is not inherently problematic, but the quality of the 40% is decisive. The stronger modules demonstrate that when CA tasks map to the same constructs as the exam, via scenario-based reasoning, structured reflection, OSCE-style checklists, or viva-style defences, the signal strengthens.

Thirdly, the findings support the caution raised earlier in the introduction regarding equity and fairness. Weak CA and exam alignment can penalise students who rely on CA feedback to calibrate study effort; if CA is soft or off-target, students enter high-stakes exams under-prepared. Conversely, students with poor CA who then excel in the exam are telling us that CA did not measure what mattered for the culminating assessment.

The results also map closely to the broader literature. Studies reporting robust CA typically feature clear standards, moderation, and higher-order task design (Jena, 2019; Kubiszyn & Borich, 2024). Where links are weak or unstable, authors cite grade inflation, recycled tasks, and inconsistent rubrics (Ukwueze, 2012; Makuvire et al., 2023). Morales et al. (2022) show that predictive validity is conditional on alignment; our module-by-module pattern is an almost textbook replication of that claim. For a midwifery context, where clinical judgment and maternal-neonatal safety are at stake, the cost of misalignment is not just statistical; it is professional and ethical.

Taken together, the implications are immediate and actionable. There is a need to re-engineer CA around higher-order outcomes by requiring every module to map tasks to *analyse* levels and by replacing recall-heavy items with case analyses, SBAR handovers, drug-calculation justifications, and brief OSCE-style

micro-stations, steps that increase cognitive isomorphism with examinations and should lift both *r* and R^2 (Anderson & Krathwohl, 2001; Shepard, 2021).

There is a need to standardise and moderate through common rubrics, double-marking of key CA tasks, and routine module-level moderation meetings to counter validity threats and integrity risks (Messick, 1995; French, Dickerson, & Mulder, 2023; Makuvire, Mufanechiya, & Dube, 2023). Tighten assessment security and attempt policies by limiting retakes, rotating item banks, and requiring original artefacts (e.g., reflective links to clinical logs) to curb inflationary practices (Ukwueze, 2012).

CA should be used as an *early-warning system* rather than a score buffer. In other words, the urge should be to deploy analytics to trigger timely support (peer instruction, supplemental coaching) so that early dips become actionable signals when CA aligns with exam constructs (Wiliam & Leahy, 2024; Chufama & Sithole, 2021). The 60/40 exam should also be reconsidered; CA weighting must only be done after quality uplift. Once CA design and moderation are strengthened, evaluation of a competency-based blend (e.g., adding OSCEs as summative components) that better reflects programme outcomes may be undertaken.

Limitations matter

With N = 30, confidence intervals around module-specific coefficients are broader than desired and using simple (one-predictor) regressions excludes other relevant covariates (e.g., prior GPA, attendance, clinical hours). Nevertheless, the consistent pattern, strong where alignment is probable, weak where it is not, indicates the signal is meaningful rather than an artefact. Future research should therefore employ multivariate models (e.g., hierarchical linear models by module), including task-level coding of Bloom levels, and test interventions (rubric standardisation, simulation-anchored CA) using pre–post or stepped-wedge designs. This would shift the discussion from description to causal improvement.

In sum, this study selectively validates both hypotheses. H_1 (positive CA–exam relationship) holds on average and robustly in several modules; H_2 (predictive utility) holds where CA is higher-order, authentic, and standardised, and collapses where CA is soft, narrow, or inconsistently graded. The message for Eswatini's midwifery education, and other contexts exhibiting similar outcomes, is urgent but practical: CA should not be abandoned but be upgraded. Ground tasks in Bloom's higher levels will enforce validity safeguards while CA will

be used to teach and to tell the truth about readiness. Done well, CA becomes an engine for learning and a credible forecast of summative performance; done poorly, it is noise. The pathway is thus clear, the tools are known, and the stakes, maternal and neonatal safety, demand that assessment design meet the moment (French et al., 2023; Wiliam & Leahy, 2024; Maki, 2023).

Conclusion

This study set out to interrogate the relationship between continuous assessment (CA) and summative examinations in midwifery education in Eswatini. Across fifteen modules, the results revealed that while CA and examinations are positively correlated, the strength of this relationship is inconsistent, ranging from weak, non-significant associations to moderately strong and significant predictive validity. On average, CA explained around 40% of the variance in examination outcomes, but regression analyses confirmed that its predictive power is modest and uneven across modules. This variability reveals a fundamental truth: the effectiveness of CA as a predictor is not inherent in the practice itself, but contingent on how it is designed, aligned, and implemented.

From a theoretical perspective, Bloom's taxonomy explained why modules with higher-order, application-oriented CA tasks (e.g., case-based or simulation exercises) demonstrated stronger predictive relationships, while construct validity theory highlighted how inconsistencies in grading, lack of standardisation, or inflationary practices undermined the interpretive value of CA scores. Methodologically, the study's use of total population sampling and quantitative analysis provided reliable evidence to support global debates: CA has potential as both a learning driver and a predictor, but only if its integrity is safeguarded and its design is anchored in higher-order competencies.

The implications are urgent for midwifery education in Eswatini and other contexts. Over-reliance on poorly aligned CA risks inflating student achievement and eroding trust, while over-reliance on summative examinations neglects the formative role of assessment in shaping learning. A balanced approach, anchored in rigorous CA design, standardisation, moderation, and cognitive alignment, offers the most credible path forward. Strengthening CA practices would not only enhance predictive validity but also improve fairness, transparency, and the quality of graduates entering the health system. Ultimately, in a context where midwifery competence is directly tied to maternal and neonatal outcomes, ensuring that assessment systems measure what matters is both an educational and a moral imperative.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Black, P., & Wiliam, D. (2018). *Classroom assessment and pedagogy*. Routledge. https://doi.org/10.4324/9781315659312

Chufama, M., & Sithole, F. (2021). The pivotal role of diagnostic, formative and summative assessment in higher education institutions' teaching and student learning. *International Journal of Multidisciplinary Research and Publications, 4*(2), 5–15.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Etikan, I., & Bala, K. (2023). Sampling methods in educational research: A methodological review. *International Journal of Education and Research, 11*(3), 45–56.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.

French, S., Dickerson, A., & Mulder, R. A. (2023). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education, 86*(2), 189–207. https://doi.org/10.1007/s10734-022-00979-6

Gamage, D., Pradeep, U., & de Silva, K. (2022). Competency-based assessment in health professions education: Emerging trends and implications. *Medical Teacher, 44*(6), 655–663. https://doi.org/10.1080/0142159X.2022.2057339

Jena, P. K. (2019). Academic assessment system of learners in IGNOU. *International Journal of Advanced Research, 7*(5), 1124–1133. https://doi.org/10.21474/IJAR01/9020

Kubiszyn, T., & Borich, G. D. (2024). *Educational testing and measurement* (12th ed.). John Wiley & Sons.

Lazareva, E., & Agostini, F. (2024). Alternative assessments in higher education: Challenges for validity and reliability. *Assessment & Evaluation in Higher Education, 49*(1), 56–72. https://doi.org/10.1080/02602938.2023.2201214

Li, J., & Zhang, Y. (2021). Formative assessment and student achievement: A meta-analysis of international studies. *Educational Assessment, Evaluation and Accountability, 33*(4), 613–639. https://doi.org/10.1007/s11092-021-09388-2

Luoma, P., & Hietanen, L. (2024). The role of quantitative methods in modern educational research. *Journal of Educational Methodology, 14*(2), 145–160.

Makule, D., & Kibusi, S. M. (2020). The role of simulation-based education in enhancing midwifery competence: Evidence from sub-Saharan Africa. *BMC Nursing, 19*(1), 1–10. https://doi.org/10.1186/s12912-020-00486-1

Makuvire, T., Mufanechiya, A., & Dube, T. (2023). Grade inflation and the credibility of continuous assessment in African higher education. *Zimbabwe Journal of Educational Research, 35*(1), 1–18.

Maki, P. L. (2023). *Assessing for learning: Building a sustainable commitment across the institution* (3rd ed.). Routledge.

McCarthy, B., Murphy, S., Larkin, P., & Hegarty, J. (2018). Simulation in midwifery education: A systematic review. *Nurse Education Today, 62*, 77–83. https://doi.org/10.1016/j.nedt.2017.12.009

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Morales, M., Salmerón, A., Maldonado, A. D., Masegosa, A. R., & Rumí, R. (2022). An empirical analysis of the impact of continuous assessment on the final exam mark. *Mathematics, 10*(21), 3994. https://doi.org/10.3390/math10213994

Morris, R., Perry, T., & Wardle, L. (2021). The impact of formative assessment on student outcomes in higher education: A review of the evidence. *Assessment & Evaluation in Higher Education, 46*(6), 843–857. https://doi.org/10.1080/02602938.2020.1810505

Pudaruth, S., Juwaheer, T. D., & Devi, R. (2013). An evaluation of continuous assessment and final examinations in tertiary education: A case study. *International Journal of Educational Management, 27*(7), 699–714. https://doi.org/10.1108/IJEM-02-2013-0018

Shepard, L. A. (2021). Classroom assessment and accountability. In E. Baker & P. Peterson (Eds.), *International encyclopedia of education* (4th ed., pp. 495–502). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.04005-4

Titova, E. V. (2022). Continuous assessment as a means of learning outcomes enhancement (Master's thesis). University of Eastern Finland.

Ukwueze, A. C. (2012). Correlation between web-based continuous assessment and examination scores in open and distance education: Implications for academic counselling. *Malaysian Journal of Distance Education, 14*(2), 71–87.

Wiliam, D., & Leahy, S. (2024). *Embedding formative assessment: Practical techniques for K–12 classrooms* (2nd ed.). Routledge.