ŌTAUTAHI CHRISTCHURCH AOTEAROA NEW ZEALAND
35TH AUSTRALASIAN ASSOCIATION FOR ENGINEERING EDUCATION ANNUAL CONFERENCE
8-11 DEC 2024
THE ENGINEER AND THE WORLD

Replacing exams with multiple tests in diploma-level first-year courses.

Hugh Wilson[a]; Kierstin Daviau[b], Joseph Ross[c], Wei Loo[a], Reza Enayatollahi[b],Elena Eskandarymalayery[c], Zaid Al-Tameemi[b], Maurice O'Neill[c], Rachel Beaton [b]
Unitec [a],Toi Ohomai [b] , Wintec[c]
Corresponding Author Email: hwilson@unitec.ac.nz

CONTEXT

High-stakes end-of-semester exams can cause otherwise competent learners to fail a course due to lack of exam experience, stress or cognitive overload at exam time. Exams are especially a problem for learners who are new or returning to tertiary education and may not have developed the academic abilities required to adequately revise and pass such a significant assessment.

Two of the first semester courses in the NZ Diploma in Engineering are DE4101:Engineering Fundamentals and DE4102: Engineering Mathematics 1. These are common to all three engineering strands (civil, mechanical and electrical engineering), and cover a wide range of topics. Some learners report finding it difficult to adequately revise for such broad topics, along with end-of-semester assessments and other exams. This can result in students who have done well in assessments during the semester doing badly in the exams. It was suggested that multiple tests may be a better alternative to exams.

PURPOSE OR GOAL

The purpose of the research was to determine whether replacing one high stakes end-of-semester exam with multiple tests produced better results in first semester engineering maths and science courses while maintaining the same academic standards.

APPROACH OR METHODOLOGY/METHODS

A new assessment regime, including common regular tests, was developed, moderated and approved by the VEENZ Quality Assurance Committee. The revised courses were delivered at Toi Ohomai, Wintec and Unitec polytechnics in Semester 1, 2024.

Research on the effectiveness of the approach consisted of a survey of learner experiences and views, a workshop where tutors reviewed the pilot delivery, a comparison of grades and pass rates with previous years' results and a brief diagnostic test in the successor courses in Semester 2 to determine the retention of the course competencies.

ACTUAL OR ANTICIPATED OUTCOMES

The results of all research indicate the use of multiple tests yields equal or better results than use of one exam. A comparison of the results between the pilot and previous years shows a significant improvement in grades and a decrease in students dropping out. Diagnostic tests in successor classes indicate that knowledge retention is the same for learners who did exams and those who did the pilot. Most learners and tutors prefer the approach citing better pacing of the course, more learner engagement and less stress.

CONCLUSIONS/RECOMMENDATIONS/SUMMARY

The use of multiple tests in place of an end-of-semester exam for first year maths and engineering science courses has yielded significant benefits.

Introduction

Exams have been used as a way of demonstrating learner competencies for centuries. They typically involve the learner in demonstrating their competencies through answering a series of questions covering a complete body of knowledge in a supervised environment in one assessment event. It is considered that this provides an unambiguous and reliable indication of the learners competencies (Bishop, 1998).

There are a number of concerns about the effect of exams on students. One is the stress that high stakes exams impose on learners. While some stress is natural, and even helpful, the stress levels that some learners experience at exam time can be unhealthy and debilitating (Chapell et al, 2005). Related to this is the ability of learners to be successful in exams does vary according to their ability to handle the stress, their exam sitting skills, their self-efficacy, their social and academic support, their level of preparation, the number and timing of exams, their mental and physical health, the stability of their lives in the exam period and many other factors (Rasul & Bukhsh, 2011; Vayre & Vonthron, 2019). These factors can result in learners of equal competence having significantly different results from the exam process (Roney & Woods, 2003). Lastly, the consequences of performing poorly in an exam resulting in failing the course can result in significant cost to a learners finances, mental health and confidence as well delaying them in completing their learning programme.

Whether exams enhance the learning process is also questioned. Williams (2014) notes that exams encourage “strategic learners” to focus on learning how to pass the exam rather than developing the required competencies. At the same time, tutors who are judged on pass rates may “teach to the exam” at the expense of time spent developing the required competencies. He also notes that since exams occur at the end of the semester, the only result is a grade rather than feedback on the learners knowledge.

A number of studies (Hernández ,2012; Lamotte, Izaute & Darnon, 2021; Larsen, Butler & Roediger, 2009; Roediger & Karpicke, 2006; Yang, Luo, Vadillo, Yu & Shanks, 2021) have explored assessment regimes that used frequent quizzes and tests and found these approaches increase learner success and improve knowledge retention. One observation is that continuous assessment supports student learning through more feedback allowing learners to identify and remedy their knowledge gaps during the learning process (Hattie & Timperley, 2007) . It was also suggested that regular testing increases learner motivation and consistent study practices as the learners are continuously focused on doing well in the next test (Roediger & Karpicke, 2006).

This paper presents the results of a pilot study conducted across three polytechnics (Unitec, Wintec and Toi Ohomai) which aimed to evaluate the efficacy of replacing an end-of-semester exam with multiple tests throughout the semester. This approach not only aligns with existing literature advocating for continuous assessment but also addresses specific challenges faced by first-year engineering students in the NZ Diploma in Engineering programme.

Background

The New Zealand Diploma in Engineering provides vocational education in civil, electrical , mechanical and fire engineering. This programme is delivered at 12 polytechnics across New Zealand. It is owned and managed by Vocational Engineering Education New Zealand (VEENZ). This has allowed a significant amount of cooperation and coordination of the programme delivery across New Zealand and is regarded as an example of best practice for vocational education.

The first semester includes DE4101:Engineering Fundamentals and DE4102: Engineering Mathematics 1 courses. These courses are compulsory and cover the basic maths and engineering science required for all disciplines. This means that they have a significant amount of content that many students struggle with. In addition they have exams which carry 50% of the final grade with a 40% thresh-hold.

These courses were chosen for the pilot due to their low success rates, highlighting the need for an urgent intervention. Each semester, 10 to 20% of students failed each course mainly because they performed poorly in the exam even though their marks throughout the semester had been reasonable. There were also a similar number of learners who gave up during the semester and do not attend the exam. These factors resulted in approximately 40% of learners failing these courses. There are a number of reasons for these very poor statistics including –

- Lack of experience in preparing for and sitting exams due to many learners being mature students who have just re-entered education or school leavers who have not had much exposure to exams in secondary school.
- Significant amount of content resulting in learners not being able to learn all of what they need for the exam.
- Excessive stress hindering the learners ability to prepare for and sit the exam.
- Learners not engaging from the start of the course and then continuing to fall behind.
- Learning disabilities, such as dyslexia, autism and ADHD, make it harder for some learners to succeed in exams even though they do well in coursework and tests.

Concern about this high failure rate has resulted in the development of a pilot delivering the two courses using an assessment regime with regular testing throughout the semester replacing an end of semester exam.

Methodology

The research was progressed by developing and delivering alternative delivery and assessment regimes in the DE4101:Engineering Fundamentals and DE4102: Engineering Mathematics 1 courses delivered in Semester 1 (late to February to late June) 2024. The pilot was undertaken by three polytechnics with Wintec and Toi Ohomai piloting both courses and Unitec only piloting the DE4101:Engineering Fundamentals course. This meant that the pilot was undertaken by five separate cohorts in two courses over three polytechnics with a total sample size of 108 individual students although some students enrolled in both courses are present in two cohorts.

The alternative delivery and assessment regimes replaced the 3 hour exam with five 45 minute controlled tests carried out during the semester. The tests were held at the end of each topic and consisted of questions that were of the same difficulty as those that in the exam. Learners were also offered the opportunity to do a resit if they failed the test but they could only get 50% if they passed.

The tutors from the three polytechnics worked together to produce a common set of tests which they checked against existing exam questions to ensure the test questions were at the same level of difficulty as the exam questions. The tests were then moderated by the tutors from other polytechnics who usually wrote the national exams for the two courses, to confirm that the difficulty of the test questions matched those in the exams. The other assessments, such as labs and quizzes, were also prepared by the group. The sequence of topic delivery was aligned although each polytechnic maintained their own times for the class delivery.

The main objective of this research was to determine whether delivery of the courses using an assessment regime without an exam was a better or equivalent approach than one with an exam. The factors that would determine whether the approach was “better or equivalent” were learner preference, tutor preference, learner success, learner retention and knowledge retention. These factors were measured by -

- A Learner Survey which was administered at the end of the semester and sought to identify the learners’ perspectives and experience of the approach.
- A Tutor Workshop which was held in the week after the semester finished. This workshop allowed tutors to reflect on the delivery of the pilots.
- Analysis of the final grades and pass rates from each cohort compared with the results from cohorts who did the courses in the same semester in previous years.

- Diagnostic tests held in the successor classes in Semester 2 2024 where the learners were tested to determine their retention of the knowledge and skills developed in the pilot courses.

Further details of these measurement processes as included in the following section.

Results

This section sets out the results of the research used to assess the effectiveness of the approach. The analysis included all five cohorts (unless otherwise stated) which showed consistent results. Since the results for each cohort were consistent, the results are presented as one combined set of data. The cohort numbers are shown in brackets (e.g. “n = 20” indicates the cohort size for that data is 20 students).

Learner Survey

A learner survey was undertaken in the last week of the semester. The main part of this survey was a series of statements to which participants could agree or disagree. There was also a series of open questions inviting comments on how the course could be improved, benefits and challenges with the new approach, what other courses could have the exam-less approach and any other comments they would like to make.

Figure 1 shows the combined survey results from all five cohorts. There was also no discernible difference to student views between polytechnics, or between genders, age and language.

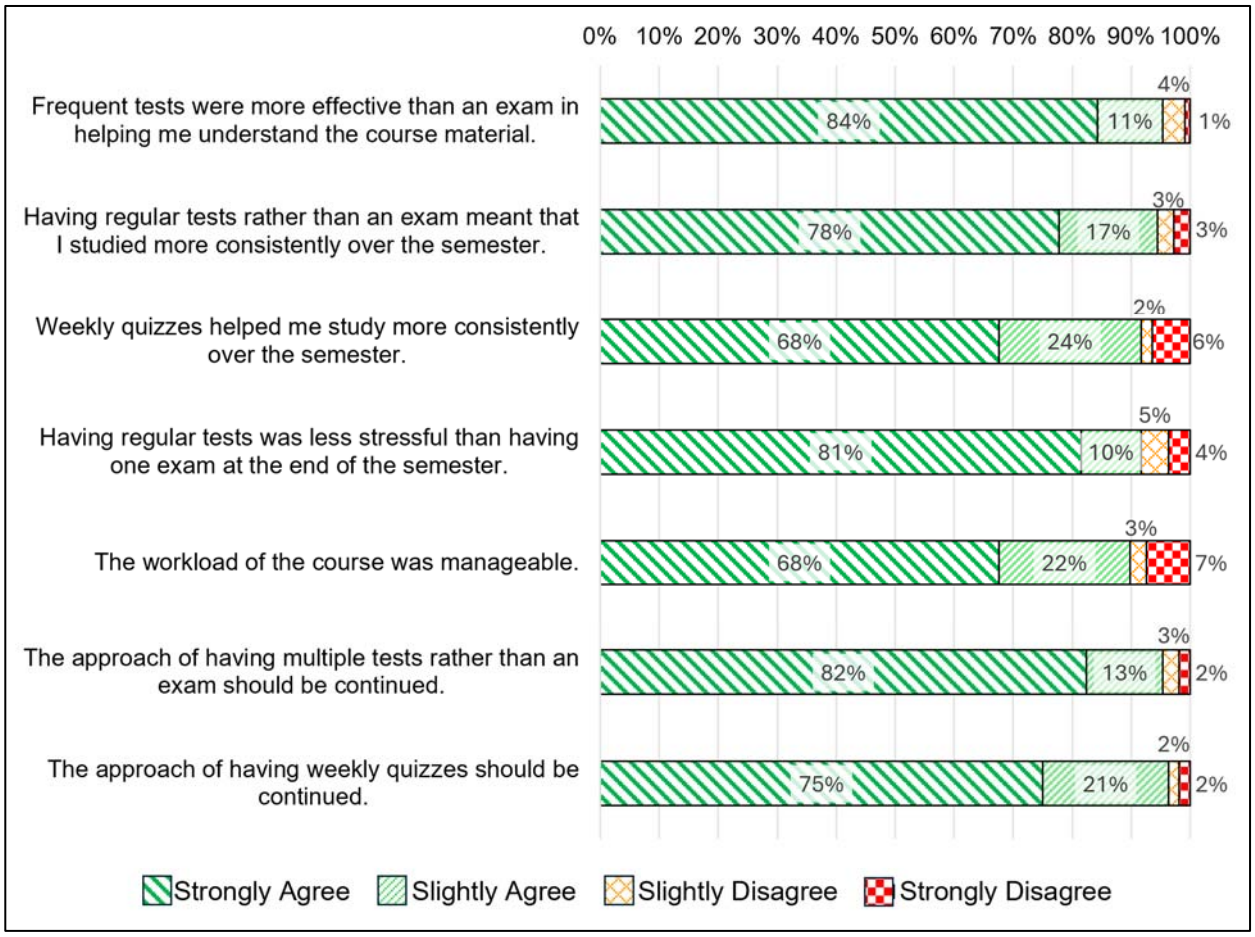

Figure 1: Learner Survey Responses (n=108)

Figure 1 indicates that almost all respondents either strongly agreed or slightly agreed with each statement. It shows that students prefer regular short tests (approximately 45 minutes long) instead of a three-hour exam at the end of the semester. Students found that this encouraged them to study regularly during the semester, as there was always a test within the next few weeks for them to prepare for. It also helped them to track their progress as they received regular feedback on their understanding of the topics covered. Conversely with courses that have a final exam some students don't feel motivated to study in the early part of the semester when the final exam is still months away. One respondent noted that "*Being able to study throughout the semester instead of all at once at the end I retained the information better"* Another noted that they found "*Not having a huge exam to procrastinate on*" was beneficial was for their learning.

Learners also agreed with the statement that regular tests were less stressful than having exams. One reason for this may be the anxiety that comes with high stakes assessments like exams, but another reason may be that the tests encourage them to work consistently throughout the semester which meant they did not fall behind. Lastly, most of the learners agreed that the approach of having multiple tests instead of exams should continue.

Tutor Reflections

The tutors involved in the delivery of the courses meet for a workshop in the week after the courses finished to reflect on the workload associated with the approach, student engagement and learning, class dynamics and whether they considered the approach should be continued.

One observation was that the shift to tests required more time for development and delivery. Tutors spent significant effort creating materials and systems, but this will reduce workload in future semesters due to reusability. Initial iterations will still demand significant time for refinement.

Student engagement increased notably. The continuous assessment format encouraged students to stay on top of the material and actively participate. Higher attendance rates and more interaction during classes were observed, indicating improved engagement.

The new format appeared to provide less stress and a more manageable workload for learners. This is backed up by responses and comments from the learner survey as well as observations from managers that there was less pastoral care required for the pilot cohort. The learners in these classes are in their first semester of study and some struggle to keep up and seek help. However, there were no learners in this position during the pilot study.

The tutors agreed that student learning improved with the new assessment model. Continuous assessments provided a better measure of students' knowledge over time. The quality and volume of student questions indicated better comprehension and deeper engagement. In addition, class delivery involved more recorded sessions to ensure access to material due to frequent assessments. This change allowed students to review sessions for tests. The regular assessments allowed tutors to stay closely connected to student progress, making teaching more dynamic. It was later noted that this engagement has carried over into Semester 2 classes.

Tutors found teaching more enjoyable due to increased student engagement. The continuous feedback from regular assessments made the teaching process more interactive and rewarding. Real-time insights into student learning added to the fulfillment of teaching. All tutors preferred to continue with the test-based assessment model due to continuous engagement and effective tracking of student progress.

Cohort Performance

The results from each cohort were collected and analysed. Data considered were the whole class final grades and course retention which were compared to the results from previous years cohorts at the same polytechnics.

Figure 2 shows the grades from both the DE4101:Engineering Fundamentals and DE4102: Engineering Mathematics 1 classes at Wintec and Toi Ohomai for Semester 1 of each year from 2021 to 2024. Unitec was excluded as it uses a slightly different method of recording withdrawals.

The figure shows a significant increase in the percentage of learners getting A and B grades while the percentage of student failing also shows a significant reduction. There was a significant drop in withdrawals (learners who dropped out of the course during the semester) which indicates that learners were better able to keep up with the coursework and may also indicate better engagement in the classroom.

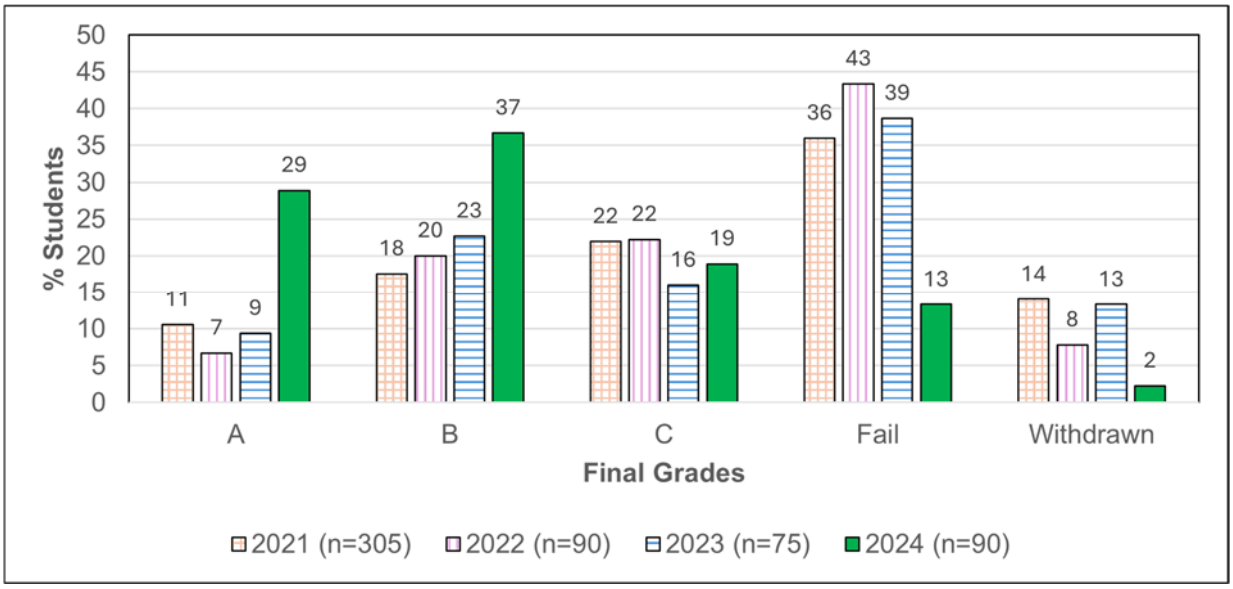

Figure 2 – Final Grades

Student performance improved under the test-based assessment regime, as shown in Figure 3. The mean and median course results show an upwards shift in the final grades with a smaller spread about the median.

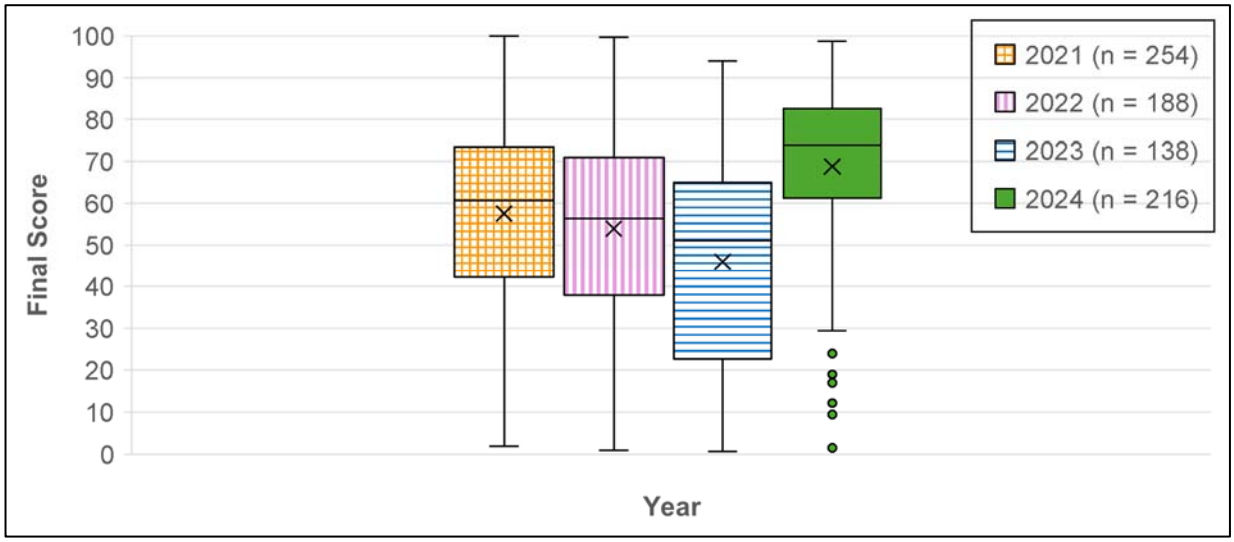

Figure 3 – Combined Score Distributions

Content Retention

Diagnostic tests were administered in the first week of the Semester 2 classes that succeed the two courses. This test consisted of 10 multichoice questions related to the basic concepts and techniques taught in the two courses. The questions ranged from basic knowledge to calculations. There were different tests for each discipline as they all have different successor courses. For example, the civil learners this test was administered in the Structures 1 course and the learners were tested on the science and maths concepts related to that course.

All students in the successor classes were tested so there was a mix of learners who have been involved in the pilot and those that had not. Those who were not involved in the pilot are the control group for the analysis.

Figure 4 shows very little difference between the average marks in each cohort. This indicates that the retention of knowledge between learners who did the pilot and those that did exams is about the same. Put another way, the data indicates that the use of tests provides the same level of knowledge retention as assessment regimes that include exams.

It is acknowledged that the diagnostic tests needed to be more rigorous to provide a more accurate indication of learner competencies. In addition, the control groups are composed of learners with a wider range of histories in the topic with some having not done the Maths and Engineering Fundamentals in the previous semesters and others repeating the Structures 1 course. This may have resulted in the control groups being weighted toward lower performing learners. Further research is required on this measure.

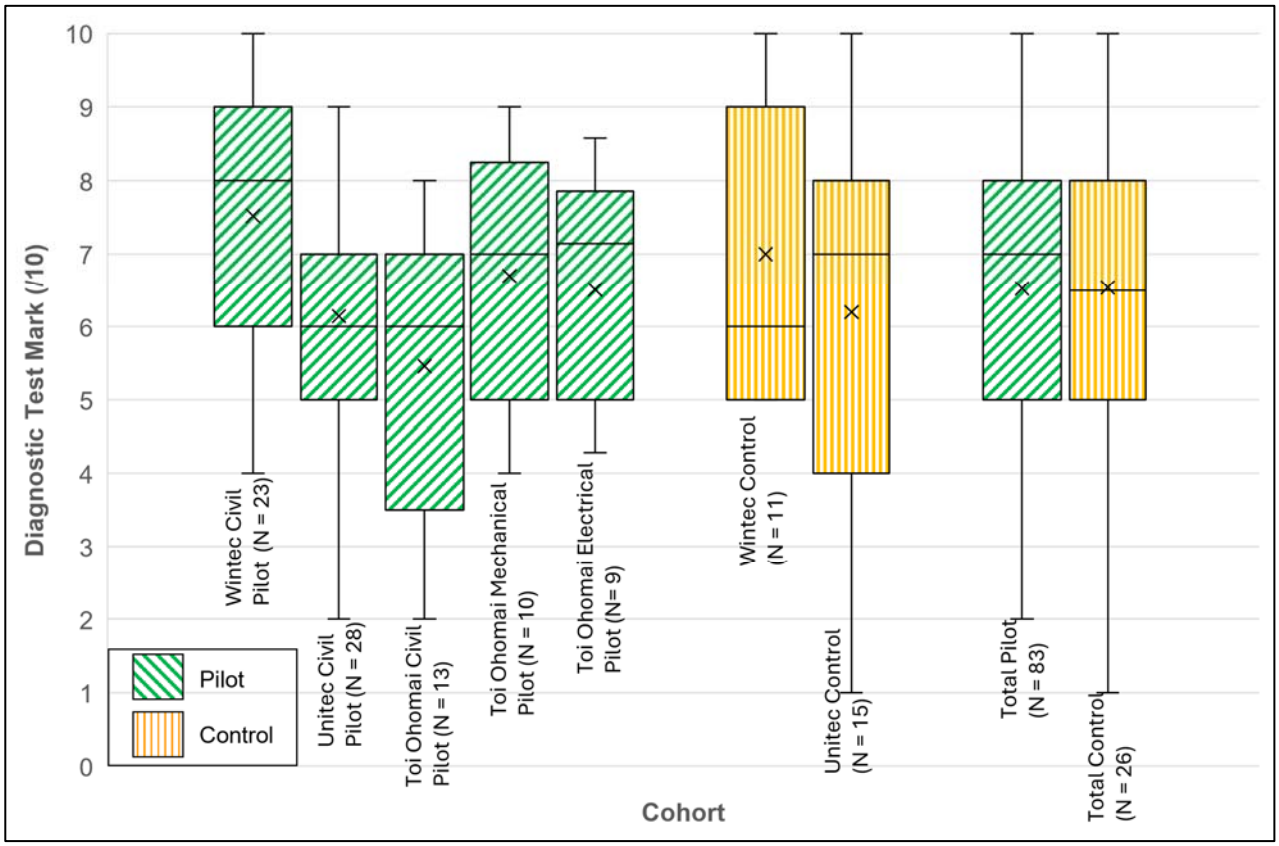

Figure 4 – Diagnostic Tests

Discussion

All polytechnics observed improved student results and retention rates during the pilot trail for both courses. The tests were designed to be the same level of rigor as the exam and the testing conditions were comparable to exam conditions for each assessment. The only difference was the timing and duration of the tests. The results of the diagnostic testing demonstrate that student knowledge and retention of the material is identical between cohorts who took the exam and those participating in the pilot programme. The consistency observed between the cohorts indicates that the quality of learning was maintained. Higher pass rates in the pilot programme mean that more students were able to achieve this same quality and extent of learning.

The tutors noted that there was more engagement in class as the learners were aware that a test was always just a week or two away and were in preparation mode as noted in previous research. However, the learner survey contained no comments about being overly stressed by this assessment regime and there was a strong preference for regular tests. This would indicate that regular tests provide a healthy level of pressure on the learners to learn but not enough to overwhelm them.

Frequent testing gave students the opportunity to have feedback on their work and, in turn, take more responsibility for their successes and mistakes. This is a stark contrast to previous years where students would often blame the exam as being ‘unfair’ if a poor mark was obtained, particularly since students do not get significant feedback on their exam performance due to the exams being held at the end of the course. The increased responsibility of the learners in the pilot programme led to better test preparation which in turn improved their success.

Resit test opportunities were available during the programme, but it is important to note that no learners passed solely because of resits. The maximum mark for resits was 50%. The knowledge that a resit was an option appeared to make learners more relaxed and better able to focus on the task at hand.

Better supporting learners and improving their academic success is a critical objective in any educational setting. One of the primary objectives of this research was to address learner stress, particularly among priority learners who are more susceptible to the pressures of high-stakes assessments. Initial concerns arose that increasing the frequency of assessments might exacerbate stress levels among students. However, these concerns did not eventuate, and it appears that the tests reduce stress through encouraging learners to keep up as well as not having a high-stakes exam to worry about.

This method also provides better alignment of the programme with industry expectations. In the professional world, graduates often need to draw on specific pieces of knowledge rather than recalling extensive content all at once. An exam does not replicate any processes in industry, whereas regular testing more closely mimics the ongoing, practical assessments that professionals face.

Conclusion

This paper has presented the results, analysis and findings of a pilot trial using multiple tests to replace an end of semester exam with other assessments also being adapted to the alternative assessment regime.

Various factors such as effects on learners, learning and tutors were measured through a learner survey, tutor workshop, grade analysis and diagnostic tests in following courses. All of these measures indicated that the use of multiple tests produced at least the same level of learning as existing assessment regimes that use exams.

There were indications that the approach had some benefits over the use of exams. One benefit was that many learners were more attentive in class because they always had a test in the near

future. However, their stress levels remained at healthy levels compared to the more extreme stress experienced with high stakes end-of semester exams. Another benefit was the ability of tutors to provide feedback after each test so learners understood where they needed to improve in a particular topic which is an improvement over what is possible with an end-of-semester exam. Both learners and tutors considered the approach to be better. Learners liked the more consistent workload and lower stress levels while tutors appreciated better learner engagement and performance.

It is recommended to continue this approach and explore its potential benefits for other levels beyond the pilot, while ensuring the positive impacts on learner engagement, stress levels, tutor feedback, and the consistency of assessing are maintained.

References

Bishop, J. H. (1998). The effect of curriculum-based external exit exam systems on student achievement. *The Journal of Economic Education*,*29*(2), 171-182.

Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of educational Psychology*, *97*(2), 268.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Hernández, R. (2012). Does continuous assessment in higher education support student learning? *Higher education*, *64*, 489-502.

Lamotte, M., Izaute, M., & Darnon, C. (2021). Can tests improve learning in real university classrooms?. *Journal of Cognitive Psychology*, *33*(8), 974-992.

Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Medical education*, *43*(12), 1174-1181.

Rasul, S., & Bukhsh, Q. (2011). A study of factors affecting students' performance in examination at university level. *Procedia-Social and Behavioural Sciences*, *15*, 2042-2047.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*,*17*(3), 249-255.

Roney, S. D., & Woods, D. R. (2003). Ideas to minimize exam anxiety. *Journal of Engineering Education,92*(3), 249.

Vayre, E., & Vonthron, A. (2019). Relational and psychological factors affecting exam participation and student achievement in online college courses. The Internet and Higher Education, 43, 100671. https://doi.org/10.1016/J.IHEDUC.2018.07.001.

Williams, P. (2014). Squaring the circle: a new alternative to alternative-assessment. *Teaching in Higher Education*,19(5), 565–577. https://doi.org/10.1080/13562517.2014.882894

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological bulletin*,*147*(4), 399.

Copyright statement