

2020

Programmatic assessment condensed: Introducing progress testing approaches to a single semester paramedic subject

James Thompson

Flinders University, Australia, james.thompson@flinders.edu.au

Donald Houston

Flinders University, Australia, don.j.houston@outlook.com

Follow this and additional works at: <https://ro.uow.edu.au/jutlp>

Recommended Citation

Thompson, James and Houston, Donald, Programmatic assessment condensed: Introducing progress testing approaches to a single semester paramedic subject, *Journal of University Teaching & Learning Practice*, 17(3), 2020.

Available at: <https://ro.uow.edu.au/jutlp/vol17/iss3/14>

Programmatic assessment condensed: Introducing progress testing approaches to a single semester paramedic subject

Abstract

The paramedic profession is rapidly evolving and has witnessed significant expansion in the scope of practice and the public expectations of the paramedic role in recent years. Increasing demands for greater knowledge and skills for paramedics has implications for the university programs tasked with their pre-employment training. The certification of paramedic student knowledge typically occurs incrementally across degree programs with aggregate results used to determine student qualification. There are concerns regarding learning sustainability of this approach. The narrowed focus of assessment practices within siloed subjects often neglects the more holistic and integrated paramedic knowledge requirements. Programmatic assessment is becoming increasingly common within medical education, offering more comprehensive, longitudinal information about student knowledge, ability and progress, obtained across an entire program of study. A common instrument of programmatic assessment is the progress test, which evaluates student understanding in line with the full broad expectations of the discipline, and is administered frequently across an entire curriculum, regardless of student year level. Our project explores the development, implementation and evaluation of modified progress testing approaches within a single semester capstone undergraduate paramedic topic. We describe the first reported approaches to interpret the breadth of knowledge requirements for the discipline and prepare and validate this as a multiple-choice test instrument. We examined students at three points across the semester, twice with an identical MCQ test spaced 10 weeks apart, and finally with an oral assessment informed by student's individual results on the second test. The changes in student performance between two MCQ tests were evaluated, as were the results of the final oral assessment. We also analysed student feedback relating to their perceptions and experiences. Mean student correct response increased by 65 percent between test 1 and 2, with substantial declines in numbers of incorrect and don't know responses. Our results demonstrate a substantial increase in correct responses between the two tests, a high mean score in the viva, and broad agreement about the significant impact the approaches have had on learning growth.

Keywords

progress test, programmatic assessment, paramedic education, capstone

Cover Page Footnote

We acknowledge statistical analysis and data management support from Dr Leah Couzner.

Introduction

Since the start of university-based paramedic education in Australia two decades ago, educators have faced challenges in preparing graduates for the highly specified paramedic role when using traditional teaching approaches (O'Brien et al., 2014). The ultimate target of graduate work-readiness within this discipline is measured by a yardstick which is both difficult to quantify and subject to differing interpretations (Thompson et al., 2015). At the time of this study, the Council of Ambulance Authorities (CAA) assumed responsibility as the regulator of the professional standards for the discipline as well as the accreditation of the national university programs, requiring universities to evidence student attainment of knowledge and skills which were seen to align with their set of broad paramedic core competency statements (CAA, 2013). Since this study, the Allied Health Professionals Regulator of Australia (AHPRA) has replaced the CAA as regulator, with these competencies replaced by an interim document of Professional Capabilities for Registered Paramedics (AHPRA, 2019). Ambulance services coordinate their practice and policies to reflect interpretations of these standards, and similarly the specific detail required to inform these broad statements is left to the discretion of local university curriculum designers.

Paramedic curriculum distributed across subjects is also assessed within these compartmentalised increments of learning, with graduate credentials constructed from the aggregated sum of subject achievements. Assumptions made about student competence which are solely formed by the accumulation of incremental milestones are, however, challenged within medical education, as the atomised testing fails to reflect the more complex and interconnectedness of the content (Schuwirth & Ash, 2013). These concerns are relevant to paramedicine also, notable as undergraduate studies draw towards completion with expectations that students have obtained and retained knowledge from all prior curriculum and are capable of applying it on demand. The consequences of graduate paramedic deficits surfacing in the field of emergency care are clearly potentially devastating. Despite this, it remains usual for university paramedic curricula to teach and assess separate subject components, before moving to the next, seldom revisiting or reassessing student knowledge foundations. Rarely are students assessed on the comprehensive, integrated knowledge required of the discipline, the breadth of an entire curriculum, or through assessments designed in context of the discipline (Thompson et al., 2015; Houston & Thompson, 2017). In addition, there are concerns for the validity and reliability of many existing testing practices. In the cases where there are no marking deterrents imposed for incorrect responses, it is not clear the extent to which student scores faithfully reflect knowledge and not chance (Schuwirth & van der Vleuten, 2012). Similar concerns relate to the effectiveness of assessments in contributing to sustainable student learning (Boud, 2000; Boud & Soler, 2016).

Our capstone paramedic subject was originally introduced to enhance graduate standards through authentic discipline-based and personalised student learning approaches (Thompson et al., 2015). Coordinated use of assessment 'for' learning has been central to all changes to the subject. While many improvements had been made within the subject design, it was evident that there was a significant gap in the expectations of student knowledge between the start of the single semester subject and their graduation. It was also apparent clear representation of the full integrated knowledge expectations of the discipline did not exist to guide students.

Programmatic assessment for learning (PAL), a design commonplace in medical education, features assessment of student knowledge across an entire broad body of curriculum representative of expectations of the field of study (Heeneman et al., 2017). One tool used to achieve this is the progress test; a comprehensive exam designed to evaluate student mastery of knowledge,

administered at regular intervals across all the years of their study (Wrigley et al., 2012). We sought to explore whether progress testing could be effectively introduced to paramedicine, and if an approach which is typified by repeated testing over a whole course, could be effectively applied within a single semester subject. This current project marks a shift in content, delivery and assessment rigour within our capstone paramedic subject. This paper describes the context for our innovation and the collaborative process we used to develop the instrument. We explain how we integrated our progress test into the student learning experience, and our various approaches to evaluate and analyse our findings and their implications.

Capstone paramedic developments

For a decade, we have been developing and evaluating teaching, learning and assessment innovations in a final year, single semester capstone subject within a Paramedic degree in an Australian university. The focus has been on 'bringing it all together' for student learning and making sense of all the material previously covered in the degree in preparation for the transition to paramedic practice (Houston & Thompson, 2017; Thompson et al., 2015). Previous cycles of action research have resulted in multiple modifications to the subject's pedagogy, principally responding to issues relating to the student relationships with assessment and its impacts on learning. Examination-related stress, grade seeking behaviour, student reluctance to accept critical feedback, and poor engagement with learning, had proved constant challenges for teaching staff (Thompson et al., 2016). The incremental changes since the initial redesign have placed much greater emphasis on formative learning, feedback to students and far deeper levels of student understanding (Houston & Thompson, 2017). We had previously introduced a formative pre-test to the start of the subject, sampling questions from subject material students had previously satisfied, before a mid-way summative multiple-choice exam (MCQ), and then explored student knowledge of the content delivered, prior to a final viva interview. While these reforms marked considerable advances to the approaches to student learning, there remained a considerable leap in the content complexity and expectations confronting the students between the pre- and midway tests. Additionally, the content and design of the midway multiple-choice test had yet to be widely validated, and principally reflected student driven learning gaps from within the semester. We made a decision to explore the introduction of a progress testing approach as a means to enhance the quality of the existing suite of assessments within the capstone subject, and trial the suitability and effectiveness of the approach within a single subject ahead of broader program-wide consideration.

Progress Testing (PT)

An established feature of medical degrees in the Netherlands for over thirty years (Vleuten et al., 1996b), the test enhanced learning approach is now a global phenomenon (Howe et al., 2004). Initially introduced in response to the effects that examinations were having on driving rote learning among students, the progress test (PT) seeks to develop deeper student understanding (Van Berkel et al., 1994). It also answered a call for a suitable assessment strategy to respond to self-directed PBL based curriculum models (Vleuten et al., 1996; Tio et al., 2016). With PTs, conventional single summative tests are replaced by a series of similar repeated tests dispersed across an entire program of study with every enrolled student across all year levels sitting the same test simultaneously (Coombes et al., 2010). Students' broad understanding is tested and retested (Muijtjens et al., 2008). This acknowledges that the outcomes of a single test are likely to be a less reliable indicator of student ability than multiple samples of testing dispersed over time (Schuwirth & van der Vleuten, 2012); a carefully choreographed suite of low stakes assessments providing maximal feedback, enabling the student to be self-aware of their ability levels and development (Hauff et al., 2014,

Muijtjens et al., 2008). Where traditional assessment programs offer insight into student incremental learning steps, they are unable to validate student mastery of the full, inter-related curriculum (Verhoeven et al., 2002).

Progress tests are designed to represent the full breadth of functional knowledge required for the discipline, are not aligned to any one subject or student year level, and commonly involve large samples of questions drawn from large question pools (Heeneman et al., 2017; McHarg et al., 2005). Sampling the whole breadth of curriculum, it is considered near impossible for a student to cram or binge learn as preparation for the test: instead learning is more evenly regulated across a full program (Muijtjens et al., 1998; Van Berkel et al., 1994). This type of test-directed learning is linked to learning enhancement as well as offering educators more reliable indicators of student knowledge retention (Schuwirth & van der Vleuten, 2012). Replacing simple passive measurement of student knowledge with a tool which is an active driver for learning has led many to re-think how they regard assessment (Freeman et al., 2010). Commonly multiple-choice exam formats are used for testing (Ricketts et al., 2009); however, examples of alternative assessment styles such as objective structured clinical examinations (OSCEs) in medicine have been recently emerging (Pugh et al., 2014). The merits of PTs have also seen their broader application to disciplines beyond medicine, such as dentistry, although we found no reported instances of PT use within paramedic education (Bennett et al., 2010). Despite our subject comprising only a single semester of a three year teaching program, we felt that a comprehensive test series which could provide students with rounds of feedback set against the discipline's knowledge requirements, matched the ethos of our capstone approach. Our initial steps were to establish and validate the knowledge expectations of the discipline.

Methods

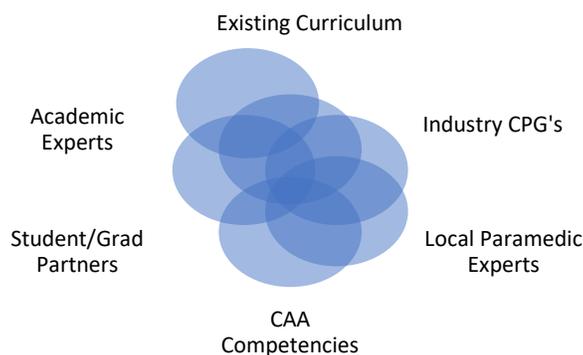
Determining Paramedic Knowledge: Paramedic learning list

Australian universities offering paramedic education have been guided by the 'Paramedic Professional Competency Standards' produced by the Council of Ambulance Authorities (CAA) (CAA, 2013) (This has since this study has been replaced by AHPRA, with a current interim set of Professional Capabilities for Registered Paramedics (AHPRA, 2019)). The CAA document broadly specifies the expectations for paramedic practice within industry, which by inference determines the goals of any underpinning education and training (O'Brien et al., 2014). Broad statements are presented under the headings of 'Professional Expectations of a Paramedic' and 'Knowledge, Understanding and Skills Required for Practice'. These are neither an exhaustive list of knowledge or skill components, nor specific instructions, but represent an equivocal set of points which can be translated for the vastly differing dialects of Australia's ambulance services and education providers.

In the absence of definitive detail to inform specific graduate knowledge, we set about compiling these elements. Starting with the existing undergraduate curriculum, each learning outcome and all teaching and assessment artefacts were reviewed, itemised and paired alongside the clinical practice guidelines (CPGs) which represent instructions for practicing local paramedics. A decision was made to restrict the parameters of content to the core paramedic science and practice topics, reflecting the more qualitative nature of much of the professional stream subjects and consideration of the suitability of a MCQ to effectively assess this knowledge. Academic staff responsible for teaching design and delivery reviewed the lists in relation to their own teaching and curriculum priorities. In addition, several senior paramedic clinicians from the local industry were invited to review items, offering opinions regarding the significance of items to the practice of paramedics. We also included several recent graduates within the item review process. The process of identifying

the elements for the list is illustrated in Figure 1. The approach represented a modified Delphi method of building consensus around the inclusion of items through repeated phases of expert item review (Gordon, 1994).

Figure 1. Learning list item contributions



The next phase involved mapping and linking each of these items through a process which drew together concepts normally the domain of a single subject with those from others and with features of practice requirements drawn from the CPGs. This scaffolding process integrated concept themes, such as anatomy, pathophysiology, pharmacology, clinical skills and field instructions. Despite these broad subject areas usually representing pre-requisite requirements, student knowledge usually had been evaluated in isolation. An exhaustive process of itemising, accounting and organising key items of learning and paramedic practice enabled us to produce a template comprising primary items and four connected subsidiary items. Mapping, distilling and aligning concepts generated a framework which underpinned our progress test. This was again considered by our review team for legitimacy and perceived relevance to both university curriculum and paramedic practice. The product which resulted was a prioritised and validated list of 100 primary concepts, each aligned to 4 sub-concepts (400 in total).

The test design: Test Questions

Once the list had been established and validated, constructing test questions commenced, with the list providing the framework for both questions and four potential responses (one correct and three distractors). Throughout the design process a goal was to ensure that assessments represented a faithful measurement of student knowledge. We deliberately sought to reduce student results which were obtained through chance or through only partial topic knowledge used to eliminate obvious incorrect distractors. We aimed to create items which a student 'who knows' would be able to get correct but a student 'who doesn't know' would be unlikely to get correct. Consequently, test outcomes would be less likely to reflect false positive or false negative performances (Schuwirth, 2004).

Literature and resources on optimal assessment design were consulted, and the revised taxonomy of multiple-choice item writing guidelines tool was applied as a filter during question composition and editing phases (Haladyna et al., 2002). Consistency in response item length and opening wording were carefully considered to ensure that item structure was unlikely to be a factor influencing the student response decision. Our group of academic staff, recent graduates and senior paramedics then

participated in a series of question review sessions to ensure relevance, non-ambiguity, fairness and balance, as a final validation of the question set with particular attention to content, format, style and writing. The result was a set of questions with validated discriminators which correlated directly to a learning list of items which integrated broad curriculum and industry expectations and represented consensus between all contributors. An example of the learning list items and MCQ question relationship is presented in Table 1.

Table 1. Learning list and MCQ relationship example

Learning list item: <i>Myocardial Infarction</i>	Subject Domains
<ul style="list-style-type: none"> • Hospital referral criteria for code STEMI 	Clinical Practice Guidelines
<ul style="list-style-type: none"> • Identify coronary arteries and the regions they perfuse 	Anatomy
<ul style="list-style-type: none"> • Explain pharmacodynamics of aspirin 	Pharmacology
<ul style="list-style-type: none"> • Describe the ECG changes consistent with MI 	Assessment skills
<p>MCQ Question: <i>Which one of the following statements regarding Myocardial Infarctions is correct?</i></p> <ul style="list-style-type: none"> a) The patient must be experiencing moderate to severe chest pain in order to satisfy code STEMI criteria b) The left anterior descending artery supplies blood to the AV node and posterior myocardium c) Aspirin promotes prostaglandin release to create less clot formation d) ST segment elevation in leads II, III and aVF is suggestive of an inferior MI e) Don't know 	

Marking & Grading Decisions

Considerable disagreement exists on optimal test marking approaches. Central to the debate is the capacity of differing approaches to provide a true account of student knowledge (Lesage et al., 2013; Burton, 2005). In the case of simple 'marks for correct answer' approach, criticism relates to assessors being unaware of the extent a final score is achieved from chance (Burton, 2001). Alternatively, negative marking approaches which seek to discourage students guessing through penalising incorrect answers attract criticism for the additional test-related anxiety these create for some students, while others suggest that it is infrequent that students entirely guess a response, but instead use deduction informed by some knowledge (Lesage et al., 2013). One point of consensus is that there is no one optimal measure, but instead a need for assessment design to consider local need and context (Burton, 2004). The specific context of our discipline ultimately informed our grading decisions. Paramount to the practice of paramedics is the requirement that all clinical decisions are founded on effective knowledge for practice, with a high degree of risk aversion and clinician recognition of their own limitations (CAA, 2013). We wanted test practices to echo this philosophy. Conscious of the critical negative marking rhetoric it was still felt that reducing chance results and encouraging students to self-identify material they had not yet mastered was consistent with our wider learning intentions. The construct of our test distractors involving differing domain knowledge was intended to counter deductive elimination based on partial knowledge guesses. In the case of a student who was unsure of the correct answer, our preference was that they choose the 'don't know' option and received the structured learning support featured within our subject

pedagogy. Our final student test scores were designed to reflect a summary of correct minus incorrect responses.

Test Implementation

Progress Test 1

Progress test 1 was administered on the first day of the semester. Typically progress testing is introduced with no prior exposure to material being examined. By contrast our students had previously covered most of the content across two and a half years of the teaching program. While they had previously satisfied the assessment milestones, their knowledge had been examined solely within the boundaries of individual subjects and not the broader context of the pre-hospital setting requirements.

Students were required to select a single correct answer from four possible options or 'don't know', with three options being distractors. The first test was entirely formative, introducing students to the PT experience and offering early performance benchmarking and self-reflection opportunities. Negative marking applied to incorrect answers and students received a zero mark for each unanswered or declared unknown answer. As each question shared a direct relationship to knowledge expectations for practice, we wanted incorrect answers to show that there would be foreseeable consequences associated with judgement or practice errors, while also considering areas of strengths and weaknesses in their understanding of the curriculum.

A common practice with PTs is to provide students with a copy of their exam questions, to encourage students to continue to reflect beyond the test, noting problems encountered when tests feedback is withheld (Wade et al., 2012). We decided to deviate from this and provided students access to their results and a copy of the learning list which corresponded directly with each individual question. This directed student learning towards identified knowledge gaps (incorrect answers) with a corresponding learning list, supporting learning while also preserving the question set for subsequent test use and enabling us to make direct comparisons on the two tests. Replacing the exam questions with the learning list was intended to encourage the development of broader student understanding and discourage students from being distracted by debating question semantics rather than investing effort in learning. We offered students the opportunity to seek additional clarification in a face to face meeting with staff, where additional feedback or concerns could be explored.

Learning List in Teaching

In addition to providing the framework for exam questions, the learning list permeated all other areas of teaching. Classroom problem-based learning (PBL) sessions constructed around authentic cases steered students through selections of items from the learning list presented in context of actual patient cases and reasoning challenges. The PBL encourages students to recognise the context and deeper understanding around material through collaborative problem solving (Vleuten et al., 1996; Wood, 2003). At the close of each class, students were required to self-nominate a list item to research before reporting back to a group-shared wiki platform. Like the collaborative peer learning experience of the PBL, the wiki offers students a vehicle to continue co-constructing understanding (Notari, 2006; Parker and Chao, 2007; Cole, 2009). Students assembled a collective body of information and sourced links supporting the learning of the group. Over the semester each group compiled entries for all the items on the learning list producing a comprehensive database of shared study resources which corresponded to the PT content.

Practical classes were also mapped to the learning list to encourage a hands-on application of required knowledge. Simulated scenarios mimicking ‘on-road’ events, required students to work through a defined discipline-specific paramedic process of care (Carter & Thompson, 2015). Student responses, performance and judgement formed the basis of these events, which were calibrated with the guidance of their paramedic tutors in a consensus-based assessment approach, a largely self-regulated approach to learning requiring students to self-critique their efforts (Thompson et al., 2017). This format of alternating PBL, practical classes and online wikis connected through the learning list was used for a ten-week cycle prior to students repeating the identical progress test for the second time.

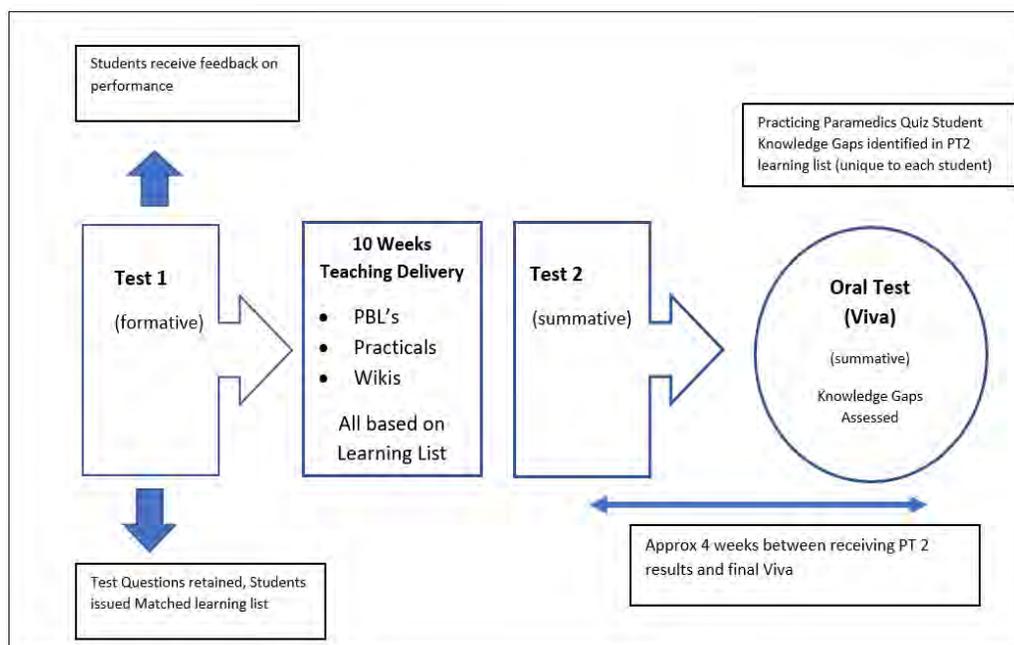
Progress Test 2

The identical test was re-administered at week 11, with student marks this time contributing to their final grade for the subject. Questions were again retained by staff at the close of the exam and feedback on performance was channelled through the learning list. This time the test was also used as a diagnostic tool with results informing a personalised oral exam, unique to the gaps identified for each student.

Viva/ Oral Exam (Test 3)

Ambulance industries routinely use a viva approach to determine knowledge or competence particularly during recruitment (Guss & Posluszny, 1984; Thompson et al., 2015). Despite the importance placed on a graduate’s ability to respond well, students had previously not been exposed to these within their paramedic program, meaning they were unprepared for these events prior to recruitment. This influenced our inclusion of vivas within our assessment strategies. Vivas are noted for enabling face to face judgements of student competence beyond what is achievable within a written exam (Torke et al., 2010). For many of our graduates these also represent one of the next major hurdles they will encounter, potentially with high stakes attached to their performance (Thompson et al., 2017). Students were made aware from the start they were to sit the two identical progress tests and that the question items they had been unable to answer correctly in Progress Test 2 (PT2) would contribute to a pool of list items they could potentially be asked to discuss during the viva. Students had approximately four weeks following the PT2 to target remaining gaps in their knowledge. The strategy intended to direct maximal learning efforts towards students’ weakest areas of understanding. Each viva was assessed by two tutors who were also practicing paramedics, in a deliberate effort to calibrate the quality of student responses against the expectations of local industry. Many of these paramedic tutors were already familiar with the academic motives of the capstone subject. Assessors selected three items from the student’s unique results profile and during a 15-minute interview the student shared their understanding of these items. Summative scores were awarded for accuracy, depth and breadth of information provided. The viva marked the final step of the interrelated test-driven learning experience illustrated in Figure 2.

Figure 2. Summary of the assessment design



Evaluation of the innovation

We analysed student performance in the identical tests administered 10 weeks apart as well as performance in the final viva. Additionally, a questionnaire was administered to participants recruited from the student cohort Ethics Approval was obtained through the Social and Behavioural Research Ethics Committee, project approval: 8034. Students were notified of the study by email prior to commencing the subject and advised their participation was entirely voluntary and assured their responses would not be identifiable. Participant responses were obtained via a paper questionnaire administered to students during the last contact day of the subject. Students were requested to rate their level of agreement with statements as well as being provided an option for free text to provide additional comment. All responses were de-identified, and responses entered to a spreadsheet for analysis.

All 103 students (101 internal and 2 distance education) attempted both progress tests. Item response analysis was conducted on the progress test questions using RASCH modelling. The statistical parameters for difficulty and discrimination had previously been established and validated for use with the university's medical degree and are presented alongside the two test results in Table 2.

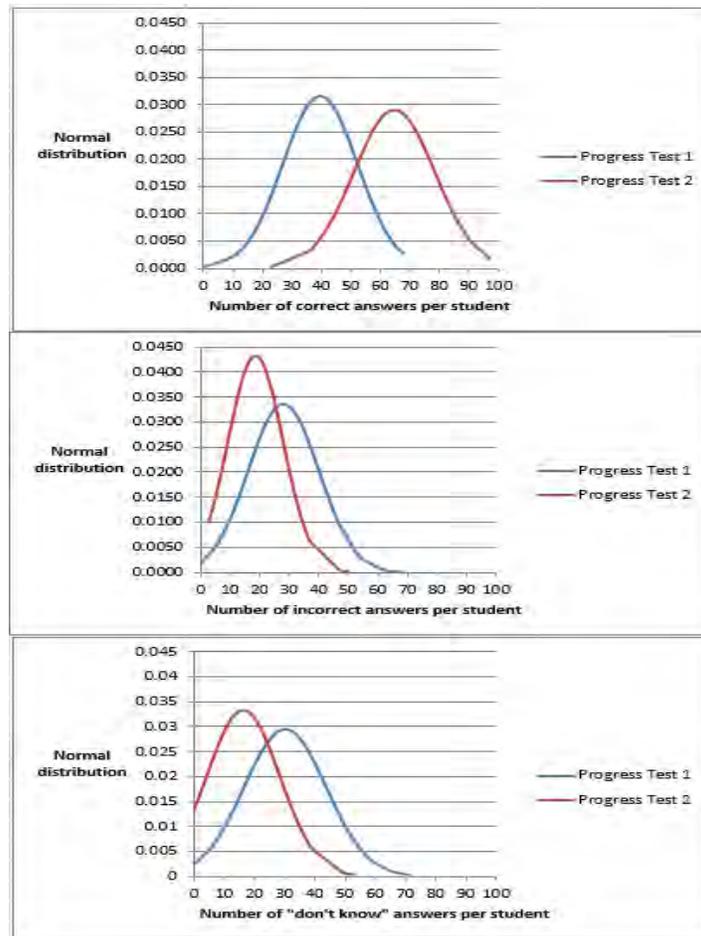
Calculation of discrimination was calculated using percentile for discrimination of 25.

Table 2. Statistical parameters for question difficulty and discrimination

$$D = \left(\frac{\text{item count (upper)}}{\text{number students (upper)}} \right) - \left(\frac{\text{item count (lower)}}{\text{number students (lower)}} \right)$$

	Difficulty				Discrimination		
	>=80 Easy	>=50 Good	>=30 Hard	<30 V Hard	>=0.1 9 Good	>=0.1 Moderate/OK	< 0.1 Non Discrim
PT 1	4	34	32	30	66	15	19
PT 2	26	53	18	3	79	15	6

Figure 3. Distribution Curves Student Responses for Progress Tests 1 & 2



Initial measures considered the differences in student responses over the 2 tests (‘correct’, ‘incorrect’ and ‘don’t know’ responses), analysed using a paired t-test analysis in order to establish the

statistical significance of the variance between the two tests using SPSS version 25. The eta squared statistic was calculated to ascertain the effect size and interpreted using the guidelines proposed by Cohen (2013). Our results showed a highly statistically significant increase in the number of *correct* responses from PT1 (M=40.15, SD=12.11) to PT2 (M=64.39, SD=13.46), $t(102)=-16.67$, $p \leq .001$ (two tailed). The mean increase in number of *correct* responses per student was -24.24 (95% CI=-27.13 to -21.36). The eta squared statistic (.73) indicated a large effect size. There was a highly statistically significant decrease in the number of incorrect responses between PT1 (M=28.17, SD=11.68) and PT2 (M=19.15, SD=9.12), $t(102)=6.92$, $p \leq .001$ (two tailed). The mean decrease in *incorrect* responses per student was 9.03 (95% CI=6.44 to 11.62). The eta squared statistic (.32) indicated a large effect size. There was a highly statistically significant decrease in the number of *don't know* responses between PT1 (M=31.69, SD=15.16) and PT2 (M=16.47, SD=11.88), $t(102)=8.56$, $p \leq .001$ (two tailed). The mean decrease in *don't know* responses per student was 15.22 (95% CI=11.69 to 18.75). The eta squared statistic (.42) indicated a large effect size. The mean scores of students' *correct*, *incorrect* and *don't know* responses between the two test intervals are presented in Figure 3.

Final Viva Results

103 students participated in oral viva assessments in week 15 of the semester. Based on the second progress test results, the mean number of potential viva themes was 35, each with 4 sub-theme items (based on incorrect or don't know responses). The cohort's performance within this final assessment item produces a mean student score of 71% with the range: 23-100%. More than 10% of the class achieved 100%.

Student Perceptions

88 students (91%) voluntarily completed the survey directly following the final viva. Table 3 illustrates the level of broad agreement obtained from the survey. Survey questions were designed to capture student perceptions relating to their experiences with the test, its effects upon their learning, and value of the approaches.

Table 3. Survey Question Response Ratings

Questions	% Broad Agreement
Test content effectively reflect the breadth of the undergraduate curriculum	89.7%
Questions challenged my understanding	95.5%
Test 1 was effective identifying gaps in my knowledge & understanding	95%
Re-sitting the identical test was an effective way to measure personal development	86.4%
I was satisfied with the amount I learned between the 2 tests	76.1%
Negative marking discouraged me from guessing answers	85.2%
I normally guess answers in exams	55.7%
The viva encouraged me to effectively target personal knowledge development	93.2%
Explaining my answers verbally enabled me to demonstrate my understanding	83%
It was beneficial to include this type of industry assessment approach to university teaching	89.7%

Free text responses proved additionally informative. We used thematic analysis to analyse the student comment across the written responses. The recognised themes emerging from the student feedback could mostly be organised within a small number of different classifications. Themes most commonly reflected were: 1. *Challenging Experience Good*, 2. *Challenging Experience Bad*, or 3. *Personal Development/ Achievement*. When these were considered in parallel with the quantitative responses it appeared consistent that students found the test to reflect the breadth of curriculum and effective at identifying knowledge gaps and challenging them to learn. However, students appeared divided over how well they received and responded to this.

Positive comments included: *“forced me out of my comfort zone”*, *“it was terrifying but very helpful in the end”*, *“Stressful but effective”*, *“challenging....definitely learned a lot”* *“more confident”*, but there were also critical reviews: *“felt discouraged from choosing (when didn’t know answers)”*, *“A lot of content to cover in a short time which made me feel pressured & stressed”*, *“difficult if you are not comfortable being scrutinised”*, *“stressful to get my abilities to expected standards”*.

Student comments also reflected on the learning process: *“Made to learn in a comprehensive manner”*, *“good preparation for the future”* *“very useful -Broad study was required ...exactly what we need...”* *“learning to self-learn is more valuable than being spoon-fed information”*

The two test scores, quantitative ratings and qualitative reports appear to support a similar conclusion: the subject was challenging but highly effective at generating learning and engaging students.

Discussion

Development of our learning list in collaboration with members from industry was critical to the identification, interpretation and validation of specific content detail. This reflects the general importance of the industry stakeholder relationship within curriculum design (Orrell, 2011; Jackson, 2015). Previously, university curriculum and industry-based practice guidelines had been considered and developed by each group in isolation, or with *ad hoc* opinions sought. Our collaborative test building approach advanced a mutual appreciation and addressed assumptions from each group. The decision to include several recent graduates to the review committee provided invaluable insight to student reactions and test strategies during design and aligns with the benefits espoused for the engagement of students as partners (Healey, 2014; Matthews, 2017).

Capstone subjects and progress tests (PTs) may appear incompatible at first glance. PTs offer longitudinal student performance data, encouraging paced learning across a whole program and discouraging intensive bursts of isolated study, where capstones represent a final learning push (Houston & Thompson, Kinzie, 2013). PT avoidance of cramming and binge learning (Schuwirth & van der Vleuten, 2012) is challenged in intensive single semester delivery. They do however share some important common ground. Both aim to facilitate learning through immersing students into a full experience of the discipline, its practices, knowledge and expectations. We accept that the confines of a single semester subject mean we forgo the beneficial longitudinal performance data. However, data from three tests (two MCQ and one viva) is a marked improvement on student data achieved from the former single summative test. A conventional PT philosophy discourages student focus on test preparation as a strategy to avoid superficial and less sustainable rote learning (Dijksterhuis et al., 2009, Van Berkel et al., 1994). In contrast we repeatedly promoted our learning list, openly advertising the 400 items (relating to 100 questions). Essentially these represented an extensive set of mini learning outcomes against which students were to be measured on three

occasions during the single semester. Where PTs direct students' focus to the wider curriculum instead of a test, we potentially met this ideal part way with our design.

Comparing PT1 and PT2 results, the 64% increase in total *correct* student responses and reductions in *incorrect* (33.5%) and *don't know* responses (47%), alongside the students' reported experiences and the observed paramedic assessor feedback, suggest considerable learning growth. Improvements to student test scores in an examination they had previously attempted, following 10 weeks of focussed teaching and learning design may seem unremarkable and likely a predictable result, but this does not represent the complete picture. This was by far the most comprehensive test the students had encountered in the history of our degree and represented knowledge critical to their future work as paramedics. Mastery of 400 learning items deemed essential to on-road practice, places greater stakes beyond a simple test score result, with foreseeable consequences linked to knowledge gaps or poor decisions. The value of using such a comprehensive test is also reflected in the literature as a means to address the practice of student strategic revision ahead of deeper sustained learning, with the importance of 'whole discipline knowledge' emphasised (Van Berkel et al. 1994; Norman et al. 2010). Until now paramedic students had not been measured on their 'whole knowledge' and the broader expectations of the paramedic role. Nor had they been previously exposed to a correct minus incorrect scoring approach. This is considered important to the reliability of making test-based decisions about students, with the justifications posed in medicine that it is unacceptable practice for doctors to be forced to guess responses when they are unsure (Schuwirth & van der Vleuten, 2012). This same argument is appropriate for the paramedic.

To the cohort of previously high achieving students embarking on their final academic phase, many already with conditional offers of employment, an adjusted class mean score of 14% on PT1 close to the end was extremely confronting. We were very interested to explore the effect our first use of negative marking had upon student test behaviour and posed a survey question about what amount of negative weighting it would take to deter students from guessing an answer in a test. While the responses varied, -1 was the most common response with 35% supporting this. Remarkably 8.3% indicated that no weighting amount would stop them from guessing to potentially optimise their scores. More than half of the respondents (55%) indicated it was normal for them to guess answers in exams. With the PT reflecting curriculum students had previously satisfied, these responses coupled with a PT1 correct score of only 40%, compelled us to question the role chance had played with inflating previous student grades.

While the numbers of *correct*, *incorrect* and *don't know* responses all showed pleasing shifts between the tests, student attitude towards the PT1 result proved pivotal to their success. Students more willing to accept the critical PT1 results proved far quicker to engage with the learning structure of the subject and respond to knowledge gaps. The free text feedback results echoed this with the identified 'challenging experience good' and 'challenging experience bad' themes, suggesting that while many students felt they were challenged by the testing process, individual 'like' or 'dislike' of being challenged influenced student decisions about the value of the approach. Many embraced the testing format and opportunities to target knowledge gaps, while others struggled with receiving such extensive critical feedback and vehemently defending a right to chance test results. The perceived impacts on GPA close to their course completion outweighed any learning benefit of the innovation for these few.

PT claims to being linked to a reduction in test-related anxiety (Heeneman et al. 2017) was certainly different to our own experience when applied to a single semester subject and for the first time for our students. Regardless of the purely formative nature of PT1, the results were clearly inconsistent with the expectations for many students. Student awareness that the next time they would face a summative test on the same instrument which had left much knowledge very exposed proved a

source of some nervousness for much of the semester. While it was not feasible in this study to compare results and experiences across previous cohorts given the changes to content and approaches, anecdotally, exam related stress was not reported to be higher than in previous years. By retaining and re-using the same test questions for PT1 and PT2 we attempted to address concerns of students memorising questions ahead of prioritising substantive learning through frequent requirements of students to demonstrate their knowledge in PBL and practical exercises. The inclusion of an oral viva further encouraged deeper student understanding. We have no way of establishing if students did or did not memorise any of the questions, however during a subject exit interview, students shared how unfamiliar they felt with the specific questions wording after having been so focussed on the learning list, with several conveying their genuine surprise that the 2 tests were identical despite being made aware this was to be the case from the start of the subject.

Our decision to include a viva to the PT offered a variation on the versatility of PTs. Although there are examples in the literature of alternatives to MCQ PTs, such as OSCEs (Pugh et al., 2014), we were unable to find reports of the use of PT content across several linked assessment formats. We had introduced the viva assessment in an earlier iteration of the subject and have found it continues to be well received by students.

Regardless of whether they liked or disliked the test-driven design, there was clear consensus the method had been highly effective at contributing to relatively rapid learning growth.

Conclusion

Consensus was reached between 100 core themes and 400 related essential concepts which can offer an illustration of the core knowledge requirements for the Australian paramedic, and a comprehensive guide for the expectations of graduates. We have been able to construct a test which we feel to be a valid instrument for measuring knowledge of this content. The approach we have used to embed the testing within a single subject offers students a transparent guide of the expectations of the discipline, and support to respond to these. We are confident that the integrated nature of content within the test design and the rigour used to ensure student results more accurately reflect student understanding.

We found clear student admissions of chance score contributions in the past and the need for a significant penalty requirement in order to influence any change to student test behaviour. While the use of negative marking may remain up for debate, discouraging student paramedics from making practice decisions when they are unsure, continues to be a position supported by our expert group. An effect on other behaviour, for example cramming, is difficult for us to measure directly, however we have been able to demonstrate regular student engagement with the test material across a semester. We recognise the limitation of the data being constrained to student performance measures and student perceptions, however these remain two fundamental measures of the success of any teaching innovations.

Our design and evaluation offer a model for others considering introducing a programmatic assessment approach to a teaching program, who may first require a test of the viability of the process prior to a wholesale course wide commitment and undertaking.

We have ambitions to transition the full paramedic program to programmatic assessment, which will require far broader acceptance and approval from internal and external interest groups.

References

- Australian Health Practitioner Regulation Agency 2019, *Professional capabilities for registered paramedics*, viewed 30/10/2019, <https://www.paramedicineboard.gov.au/Professional-standards/Professional-capabilities-for-registered-paramedics.aspx>
- Bennett, J, Freeman, A, Coombes, L, Kay, L & Ricketts, C 2010, 'Adaptation of medical progress testing to a dental setting', *Medical Teacher*, vol.32, no.6, pp.500-502.
- Boud, D 2000, 'Sustainable assessment: rethinking assessment for the learning society', *Studies in continuing education*, vol.22, no.2, pp.151-167.
- Boud, D & Soler, R 2016, 'Sustainable assessment revisited'. *Assessment & Evaluation in Higher Education*, vol.41, no.3, pp.400-413.
- Burton, RF 2001, 'Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers', *Assessment & Evaluation in Higher Education*, vol.26, no.1, pp. 41-50.
- Burton, RF 2004, 'Multiple choice and true/false tests: reliability measures and some implications of negative marking', *Assessment & Evaluation in Higher Education*, vol.29, no.5, pp. 585-595.
- Burton, RF 2005, 'Multiple-choice and true/false tests: myths and misapprehensions', *Assessment & Evaluation in Higher Education*, vol.30, no.1, pp. 65-72.
- Council of Ambulance Authorities 2013, Paramedic Professional Competency Standards Version 2.2., viewed 30/10/2019, https://www.caa.net.au/images/documents/resources_for_universities/Paramedic_Professional_Competency_Standards_V2.2_February_2013_PEPAS.pdf
- Carter, H & Thompson, J 2015, 'Defining the paramedic process', *Australian journal of primary health*, vol.21, no.1, pp.22-26.
- Cohen, J 2013, *Statistical power analysis for the behavioral sciences*, Routledge.
- Cole, M 2009, 'Using Wiki technology to support student engagement: Lessons from the trenches', *Computers & Education*, vol.52, no.1, pp. 141-146.
- Coombes, L, Ricketts, C, Freeman, A & Stratford, J 2010, 'Beyond assessment: feedback for individuals and institutions based on the progress test', *Medical teacher*, vol.32, no.6, pp. 486-490.
- Dijksterhuis, M, Scheele, F, Schuwirth, L, Essed, G, Nijhuis, J & Braat, D 2009, 'Progress testing in postgraduate medical education', *Medical teacher*, vol.31, no.10, pp. e464-e468.
- Freeman, A, van der Vleuten, C, Nouns, Z & Ricketts, C 2010, 'Progress testing internationally', *Medical Teacher*, vol.32, no.6, pp. 451-455.
- Gordon, TJ 1994, 'The delphi method', *Futures Research Methodology*, vol.2, no.4, pp.1-30.
- Guss, DA & Posluszny, M 1984, 'Paramedic orotracheal intubation: a feasibility study' *The American Journal of Emergency Medicine*, vol.2, no.5, pp. 399-401.
- Haladyna, TM, Downing, SM & Rodriguez, MC 2002, 'A review of multiple-choice item-writing guidelines for classroom assessment', *Applied Measurement in Education*, vol.15, no.3, pp.309-333.
- Hauff, SR, Hopson, LR, Losman, E, Perry, MA, Lybson, ML, Fischer, J & Santen, SA 2014, 'Programmatic assessment of level 1 milestones in incoming interns', *Academic Emergency Medicine*, vol.21, no.6, pp. 694-698.
- Healey, M 2014, 'Students as partners in learning and teaching in higher education', Workshop Presented at University College Cork, vol.12, pp.15.
- Heeneman, S, Schut, S, Donkers, J, van der Vleuten, C & Muijtjens, A 2017, 'Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment', *Medical Teacher*, vol.39, no.1, pp. 44-52.

- Houston, D & Thompson, J 2017, 'A bridge to 'being' a practitioner: the role of pedagogical practice-in-context knowledge in the design, delivery and experience of a capstone subject', *Research and Development in Higher Education: Curriculum Transformation*, vol.40, pp. 175-185.
- Houston, D & Thompson, J N 2017, 'Blending Formative and Summative Assessment in a Capstone Subject: 'It's not your tools, it's how you use them'', *Journal of University Teaching & Learning Practice*, vol.14, no.3
- Howe, A, Campion, P, Searle, J & Smith, H 2004, 'New perspectives—approaches to medical education at four new UK medical schools' *Bmj*, vol.329, no.7461, pp. 327-331.
- Jackson, D 2015, 'Employability skill development in work-integrated learning: Barriers and best practice', *Studies in Higher Education*, vol.40, no.3, pp. 350-367.
- Kinzie, J 2013, 'Taking stock of capstones and integrative learning', *Peer review*, vol.15, no.4, pp. 27.
- Lesage, E, Valcke, M & Sabbe, E 2013, 'Scoring methods for multiple choice assessment in higher education—Is it still a matter of number right scoring or negative marking?', *Studies in Educational Evaluation*, vol.39, no.3, pp. 188-193.
- Matthews, KE 2017, 'Five propositions for genuine students as partners practice', *International Journal for Students as Partners*, vol.1, no.2
- McHarg, J, Bradley, P, Chamberlain, S, Ricketts, C, Searle, J & McLachlan, JC 2005, 'Assessment of progress tests', *Medical education*, vol.39, no.2, pp. 221-227.
- Muijtjens, A, Hoogenboom, R, Verwijnen, G & van der Vleuten, C 1998, 'Relative or absolute standards in assessing medical knowledge using progress tests', *Advances in Health Sciences Education*, vol.3, no.2, pp. 81-87.
- Muijtjens, AM, Schuwirth, LW, Cohen-Schotanus, J & van der Vleuten, CP 2008, 'Differences in knowledge development exposed by multi-curricular progress test data', *Advances in Health Sciences Education*, vol.13, no.5, pp. 593-605.
- Norman, G, Neville, A, Blake, JM & Mueller, B 2010, 'Assessment steers learning down the right road: impact of progress testing on licensing examination performance', *Medical Teacher*, vol.32, no.6, pp. 496-499.
- Notari, M 2006, 'How to use a Wiki in education: Wiki based effective constructive learning', *Proceedings of the 2006 international symposium on Wikis, 2006 ACM*, pp.131-132.
- O'Brien, K, Moore, A, Dawson, D & Hartley, P 2014, 'An Australian story: paramedic education and practice in transition', *Australasian Journal of Paramedicine*, vol.11, no.3
- Orrell, J 2011, 'Good practice report: Work-integrated learning', *ALTC: Strawberry Hills*.
- Parker, K & Chao, J 2007, 'Wiki as a teaching tool' *Interdisciplinary Journal of e-learning and Learning Objects*, vol.3, no.1, pp. 57-72.
- Pugh, D, Touchie, C, Wood, TJ & Humphrey-Murto, S, 2014, 'Progress testing: is there a role for the OSCE?', *Medical education*, vol.48, no.6, pp.623-631.
- Ricketts, C, Freeman, AC & Coombes, LR 2009, 'Standard setting for progress tests: combining external and internal standards' *Medical Education*, vol.43, no.6, pp.589-593.
- Schuwirth, L & Ash, J 2013, 'Assessing tomorrow's learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could forget', *Medical Teacher*, vol.35, no.7, pp. 555-559.
- Schuwirth, LW 2004, 'Assessing medical competence: finding the right answers', *The clinical teacher*, vol.1, no.1, pp. 14-18.
- Schuwirth, LW & van der Vleuten, CP 2012, 'The use of progress testing', *Perspectives on medical education*, vol.1, no.1, pp. 24-30.
- Thompson, J, Grantham, H & Houston, D 2015, 'Paramedic capstone education model: Building work ready graduates', *Australasian Journal of Paramedicine*, vol.12, no.3

- Thompson, J, Houston, D, Dansie, K, Rayner, T, Pointon, T, Pope, S, Cayetano, A, Mitchell, B & Grantham, H 2017, 'Student & tutor consensus: a partnership in assessment for learning', *Assessment & Evaluation in Higher Education*, vol.42, no.6, pp. 942-952.
- Tio, RA, Schutte, B, Meiboom, AA, Greidanus, J, Dubois, EA & Bremer, AJ 2016, 'The progress test of medicine: the Dutch experience', *Perspectives on Medical Education*, vol.5, no.1, pp. 51-55.
- Torke, S, Abraham, RR, Ramnarayan, K & Asha, K 2010, 'The impact of viva-voce examination on students' performance in theory component of the final summative examination in physiology', *Journal of Physiology and Pathophysiology*, vol.1, no.1, pp. 10-12.
- Van Berkel, HJ, Nuy, HJ & Geerlings, T 1994, 'The influence of progress tests and block tests on study behaviour', *Instructional Science*, vol.22, no.4, pp. 317-333.
- Verhoeven, B, Verwijnen, G, Scherpbier, A & van der Vleuten, C 2002, 'Growth of medical knowledge', *Medical Education*, vol.36, no.8, pp. 711-717.
- Vleuten, CVD, Verwijnen, G. & Wijnen, W 1996, 'Fifteen years of experience with progress testing in a problem-based learning curriculum', *Medical Teacher*, vol.18, no.2, pp. 103-109.
- Wade, L, Harrison, C, Hollands, J, Mattick, K, Ricketts, C & Wass, V 2012, 'Student perceptions of the progress test in two settings and the implications for test deployment', *Advances in Health Sciences Education*, vol. 17, no.4, pp. 573-583.
- Wood, DF 2003, 'Problem based learning', *Bmj*, vol.326, pp.328-330.
- Wrigley, W, van der Vleuten, CPM, Freeman, A & Muijtjens, A 2012, 'A systemic framework for the progress test: Strengths, constraints and issues: AMEE Guide No. 71', *Medical Teacher*, vol.34, pp. 683-697.