

BMJ Open Role of formative assessment in predicting academic success among GP registrars: a retrospective longitudinal study

Paula Heggarty,¹ Peta-Ann Teague,¹ Faith Alele ², Mary Adu,¹ Bunmi S Malau-Aduli ¹

To cite: Heggarty P, Teague P-A, Alele F, *et al.* Role of formative assessment in predicting academic success among GP registrars: a retrospective longitudinal study. *BMJ Open* 2020;**10**:e040290. doi:10.1136/bmjopen-2020-040290

► Prepublication history and supplemental material for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-040290>).

Received 11 May 2020

Revised 25 October 2020

Accepted 29 October 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹College of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia

²College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Queensland, Australia

Correspondence to

A/Professor Bunmi S Malau-Aduli;
bunmi.malauaduli@jcu.edu.au

ABSTRACT

Objectives The James Cook University General Practice Training (JCU GPT) programme's internal formative exams were compared with the Royal Australian College of General Practitioners (RACGP) pre-entry exams to determine ability to predict final performance in the RACGP fellowship exams.

Design A retrospective longitudinal study.

Setting General Practice (GP) trainees enrolled between 2016 and 2019 at a Registered Training Organisation in regional Queensland, Australia.

Participants 376 GP trainees enrolled in the training programme.

Exposure measures The pre-entry exams were Multiple-Mini Interviews (MMI), Situational Judgement Test (SJT) and Candidate Assessment and Applied Knowledge Test. The internal formative exams comprised multiple choice questions (MCQ1 and MCQ2), short answer questions, clinical skills and clinical reasoning.

Primary outcome measure The college exams were Applied Knowledge Test (AKT), Key Feature Problems (KFP) and Objective Structured Clinical Examination (OSCE).

Results Correlations (*r*), coefficients of determination (*R*²) and OR were used as parameters for estimating strength of relationship and precision of predictive accuracy. SJT and MMI were moderately (*r*=0.13 to 0.31) and MCQ1 and MCQ2 highly (*r*=0.37 to 0.53) correlated with all college exams (*p*<0.05 to *p*<0.01), with *R*² ranging from 0.070 to 0.376. MCQ1 was predictive of failure in all college exams (AKT: OR=2.32, KFP: OR=3.99; OSCE: OR=3.46); while MCQ2 predicted failure in AKT (OR=2.83) and KFP (OR=3.15).

Conclusion We conclude that the internal MCQ formative exams predict performance in the RACGP fellowship exams. We propose that our formative assessment tools could be used as academic markers for early identification of potentially struggling trainees.

INTRODUCTION

Speciality training programmes can be very challenging for some junior doctors. Despite meeting the high academic standards required for postgraduate medical training, a proportion of trainees struggle to perform well in their fellowship exams.^{1 2}

Strengths and limitations of this study

- This is the first study investigating the predictive roles of both pre-entry exams and in-training assessment on General Practice (GP) trainees' performance in the college exams at a GP regional training organisation (RTO) that uses a distributed training model.
- This is also the first study to investigate the predictive ability of Candidate Assessment and Applied Knowledge Test in an Australian setting.
- This study was conducted at only one RTO.
- Generalisation of findings to other settings may be limited by the retrospective nature of the study.

A recent public report noted that 22.5% of General Practice (GP) trainees were likely to fail their first attempt of fellowship exams.³ Failure to achieve fellowship is a concern for the candidate who may suffer considerable personal distress with possible financial hardships emanating from the need to pay for resit exams, and perhaps stigma or shame after failing one or more components of the exams.⁴ Therefore, timely identification of struggling trainees is an important step for early remediation plans to prevent failure.^{5 6}

Studies on early identification of students' academic underperformance have been commonly reported in undergraduate medical education.^{7–11} However, there is a paucity of literature on the subject in postgraduate training. Similar to undergraduate medical education, a significant proportion of trainees encounter academic failure during their postgraduate training, hence the need for early support. Studies suggest that remediation should be an explicit part of the medical education programme rather than an afterthought activity put in place after trainees fail.^{12 13} Without proactive engagement in academic support programmes that

foster early identification of GP trainees' educational difficulties, regional training organisation (RTO) remediation programmes will remain a 'band-aid fixing tool' for failed registrars. Such an approach subsequently results in reduced effectiveness of the training programme and comes at an increased cost to individual trainees and the training organisation.^{14 15}

In response to the need for early diagnosis of trainees' educational difficulties and associated limiting factors, it is important to use the concept of 'assessment for learning' to address learners' learning needs and strategically change the ways in which they learn (Isaacs, 2001).¹⁶ Assessment has the potential to predict performance in practice and aid identification of learning deficiencies, especially with the use of valid and reliable tools.^{2 17} Pre-entry assessment tools used by the Royal Australian College of General Practitioners (RACGP) are the Situational Judgement Test (SJTs), Multiple-Mini Interviews (MMI) and the Candidate Assessment Applied and Knowledge Test (CAAKT) which was introduced recently to replace the SJT. The SJT and the CAAKT aim to assess clinical reasoning and problem-solving attributes in trainees, while fundamental clinical competencies across the Domains of General Practice are assessed by behavioural style clinical questions using the MMI.^{18 19} Although the predictive ability of SJT and MMI on performance in end-of-training assessment had been previously established,²⁰ the predictive ability of CAAKT is currently unknown. In addition, there is need for comparative studies with internally developed assessment tools, particularly in the context of their use in RTOs. Shulruf *et al*²¹ reported the inability of interview/admission test scores to predict subsequent student failure or drop-out and concluded that it might be more useful to focus on post-enrolment factors. Evaluating the effect of internally developed formative assessment schedules could provide important data for adopting a programmatic approach to assessment.²²

Therefore, the following research questions were considered:

1. What is the predictive ability of the RACGP SJT, MMI and CAAKT exams on James Cook University (JCU) GP registrars' performance in RACGP exams (end-of training assessment)?
2. What is the predictive ability of internally developed JCU GP training formative exams on registrars' performance in RACGP exams (end-of training assessment)?

The findings of this study will enhance the development of a systematic assessment programme that fosters early identification of registrars potentially at risk of academic difficulty, thus enabling effective and timely remediation strategies.

MATERIALS AND METHODS

Study setting

James Cook University's (JCU) postgraduate general practice training (GPT) programme was established in

2016 to deliver the Australian General Practice Training (AGPT) programme in 'North West Queensland', an area that encompasses all of Queensland apart from the South East corner of the state. JCU provides training pathways to the award of Fellowship of the Royal Australian College of General Practitioner (FRACGP), Fellowship of the Australian College of Rural and Remote Medicine and Fellowship in Advanced Rural General Practice. JCU, as a RTO, has a mandate to recruit, train and retain a fit for purpose general practice workforce in remote, rural and regional Queensland. As such, JCU GPT aims to lessen attrition of trainees (both voluntary and involuntary), lessen trainees' underperformance (decreasing reliance on remediation) and increase successful completion of training rates. This mandate resulted in the design of internal formative assessment tools that allow for assessment of trainees' academic ability right from the first day of orientation to the programme. The internal formative assessment schedule entails educational diagnosis for identification of trainees' learning needs, hence providing the opportunity for early remediation plans to foster successful completion of the training.

Study procedure and population

Using a retrospective longitudinal study design, we collated academic pre-entry data (SJT, CAAKT and MMI) for all GP trainees who were enrolled with JCU GPT at the time of the study (2016 to 2019). Demographic variables which were obtained from the registrars at the point of entry into JCU GP programme and stored in the University Record System database were included in this study. These variables include gender, training pathway, origin and fellowship status. Data on the registrars' internal formative assessment scores comprising of the orientation assessment (multiple choice questions (MCQ1), short answer questions (SAQ), clinical skills (CS) and clinical reasoning (CR)) and subsequent assessment (MCQ2) of trainees 8 months after the MCQ1, were also collated. Data on college exam scores (Applied Knowledge Test (AKT), Key Feature Problems (KFP) and Objective Structured Clinical Examination (OSCE) exams) for trainees who had completed at least two of the three college exams were obtained. Registrars need to pass AKT and KFP before they can attempt the OSCE exam. All registrars in the study had attempted at least both AKT and KFP. Only the first attempt scores in the college exams were included in the analysis.

MEASURES

Exposure measure

Pre-entry assessment

Prior to 2018, applicants to the RACGP training programme were required to undertake SJT and MMI. The SJT items assess non-academic (professional) abilities such as empathy, integrity and coping with pressure.²³ The questions test awareness of effective lines of action in each situation as well as procedural knowledge.²⁰ If successful

in this assessment, candidates are invited to undertake a MMI which is a structured interview style assessment of between five and eight stations. The MMI assesses clinical competencies across the RACGP Domains of General Practice including interpersonal communication through behavioural style scenarios and questions.²⁰ From 2018, applicants are required to undertake a CAAKT which is a written test that comprises both applied knowledge and SJT questions. The CAAKT replaced the SJT. Based on their successful performance in these assessments, applicants are offered a training position on the AGPT programme training towards a FRACGP.

Internal formative exams

JCU GPT offers all trainees commencing in general practice training posts a number of formative assessments. The registrars are required to take an introductory formative exam at the start of their training which includes four different assessments: a 65-question MCQ paper, a short answer paper (SAQ), observed simulated consultations (CS) and a written clinical reasoning paper (CR). Another 65-question MCQ paper is administered to the trainees 8 months after commencing the training programme. These are in-training formative assessments conducted by JCU to evaluate the registrars' knowledge and learning needs and are different from the summative assessments (college examinations) which are organised independently by the RACGP.

Outcome measures

RACGP summative exams

RACGP pathway trainees are required to successfully pass an AKT which comprises 150 questions that assess applied clinical knowledge, a written key feature paper (KFP) which assesses applied clinical reasoning and clinical decision-making skills and an OSCE which assesses applied knowledge, clinical reasoning, communication skills and professional behaviours.

Data analysis

Given that the national pass scores for the college examinations vary per year/sitting, we standardised the AKT, KFP and OSCE scores before conducting the analyses. Missing data were detected and deleted by pairwise comparative scanning to maximise effective sample size. After assumptions of linearity, homoscedasticity, independence of errors, non-multicollinearity, unusual points and normality of residuals were met, descriptive statistics comprising frequencies and mean scores were computed in a preliminary analysis. IBM SPSS (V.23) was used in running correlations, and multiple linear and logistic regressions.

Correlation analysis: Pearson's pairwise correlation coefficients between all pre-entry, internal formative and college exams were used to estimate the strength of association between these variables and significance thresholds were set at $p < 0.05$ and $p < 0.01$.

Multiple linear regression analysis: Multiple linear regression analyses to estimate the ability of pre-entry (SJT, CAAKT and MMI) and in-training formative (MCQ1, CR, CS, SAQ and MCQ2) exams to predict performance in the final college exams were conducted. For the purpose of these analyses, CAAKT was not included in the multiple linear regression model with SJT and MMI, because SJT and CAAKT are similar, hence such an inclusion could escalate the Variance Inflation Factor and trigger multicollinearity. Therefore, a separate multiple linear regression model that included only CAAKT and MMI to assess their ability to predict RACGP Fellowship exams was run.

Logistic regression analysis: The probability of failing each college examination as a result of failure in the internal formative exams was estimated using separate logistic regression analyses. Only significant internal formative assessment variables that were predictive of all college exams in the linear regression were included in the logistic regression model. The formative assessment scores were converted to categorical pass or fail variables. The pass mark for the internal assessment was 55% and scores below 55% were considered a fail. Gender and origin were adjusted for and fitted as covariates in the multiple linear and logistic regression models for continuous and categorical variables, respectively.

Patient and public involvement

Patients or the public were not involved in this study.

RESULTS

Assessment data of 376 trainees were included in the study. Demographic profiles showed that 58.2% were women, 1.9% of Aboriginal and Torres Strait Islander descent, 64.1% Australian medical graduates (AMG), 96.5% enrolled in the FRACGP Only training pathway and 61.7% had become fellows (online supplemental table 1).

All score categories in the study showed normal distribution, with the exception of SJT and CAAKT which were slightly negatively skewed (-1.58 and -1.5 , respectively), as commonly observed in a typical SJT score distribution.^{20 24} Nevertheless, parametric analysis was conducted for all score categories, due to the large data set. Previous research had demonstrated the robustness of using parametric analyses for large data sets, except in instances of extreme skewness.²⁵

Descriptive statistics revealed that in the pre-entry exams, mean scores for SJT, MMI and CAAKT were 84.09%, 31.78% and 68.05%, respectively. Mean scores for the internal formative assessment ranged from 49.24% in CR to 62.4% in MCQ2. The mean scores for AKT, KFP and OSCE were 72.09%, 60.85% and 71.79%, respectively (online supplemental table 2). The first attempt pass rates for the RACGP exams were 86.9% for AKT, 76.6% for KFP and 92.6% for OSCE (data not shown). In addition, univariate analysis revealed that performance was not influenced by gender in the AKT ($\chi^2 = 1.940$,

Table 1 Pearson correlation coefficients between pre-entry and college exams

Categories	SJT	MMI	CAAKT	AKT	KFP	OSCE
Pre-entry exams						
SJT	—					
MMI	0.18**	—				
CAAKT	†	0.28	—	0.04	-0.02	0.47**
College exams						
AKT	0.31*	0.22**	0.04	—	0.66**	0.40**
KFP	0.25**	0.13*	-0.02	0.66**	—	0.43**
OSCE	0.26*	0.24**	0.47**	0.40**	0.43**	—

*p<0.05, **p<0.01.

†Correlation could not be computed because at least one of the variables is constant.

AKT, Applied Knowledge Test; CAAKT, Candidate Assessment and Applied Knowledge Test; KFP, Key Feature Problems; MMI, Multiple-Mini Interviews; OSCE, Objective Structured Clinical Examination; SJT, Situational Judgement Test.

df=1, p=0.167) and OSCE (0.391, df=1, p=0.532) exams. However, females obtained significantly higher scores than their male counterparts in the KFP exam ($\chi^2=5.225$, df=1, p=0.026). A similar trend was observed with origin (AMG/international medical graduates (IMG)), whereby performance in the AKT ($\chi^2=0.391$, df=1, p=0.530) and OSCE ($\chi^2=0.071$; df=1, p=0.817) exam were not influenced, but AMGs obtained significantly higher scores than IMGs in the KFP exams ($\chi^2=11.990$, df=1, p=0.001).

Table 1 presents the correlation coefficients between the pre-entry exams (SJT, MMI and CAAKT) and the college exams. Results showed that performance on SJT and MMI are significantly and positively correlated with performance on all college assessments (r values ranging from 0.13 to 0.31; p<0.05 to p<0.01). However, these correlations are weak, with MMI having the lowest correlation with KFP (r=0.13, p<0.05). On the other hand,

CAAKT was only significantly correlated with performance on OSCE (r=0.47, p<0.01). Overall, SJT showed a stronger correlation with the college exams compared with the MMI. It is important to note that CAAKT and SJT did not yield any correlation value because they are similar.

Table 2 presents the correlation coefficients between the internal formative examinations and the college exams. Overall, all other internal exams except clinical skills were significantly correlated with the college exams. Of all assessments, MCQ1 and MCQ2 showed stronger correlations with all the college exams (r values ranging from 0.37 to 0.53; p<0.01) compared with SAQ and CR (r values ranging from 0.15 to 0.37; p<0.05 to p<0.01).

Multiple regression analysis was conducted to ascertain the extent to which pre-entry exams and internal formative exams significantly predict scores in each college exam. Gender and origin were included in the multiple linear regression analyses and were not significant predictors of performance in the college examination. Table 3 shows the predictive relationship between the pre-entry examination scores and the college examination. Pre-entry SJT was predictive of college AKT examination (p<0.001), KFP (p<0.001) and OSCE (p<0.001). MMI was predictive of college AKT (p=0.002) and OSCE (p=0.001) but was not predictive of KFP. It was found that 0.7% to 12.5% of the variation in the college exams could be explained by the pre-entry exams (SJT and MMI). Pre-entry CAAKT was predictive of OSCE (p=0.001) but not of AKT and KFP. When the predictive role of both CAAKT and MMI (in one regression model) were assessed, MMI was predictive of AKT (p=0.027) and KFP (p<0.001) but lost its predictive power with OSCE (data not shown) which differs from the model with SJT.

The results presented in table 4 show the predictive relationship between the internal formative examination scores and the college examination. All independent variables (internal formative exam scores) were entered into

Table 2 Pearson correlation coefficients between internal formative and college exams

Categories	MCQ1	SAQ	CS	CR	MCQ2	AKT	KFP	OSCE
Internal formative exams								
MCQ1	—	0.34**	-0.007	0.32**	0.52**	0.48**	0.48**	0.40**
SAQ	0.34**	—	0.29**	0.15*	0.33**	0.25**	0.15**	0.17**
CS	-0.007	0.29*	—	0.93	0.05	0.11	-0.08	-0.09
CR	0.32**	0.15*	0.93	—	0.34**	0.33**	0.37**	0.17*
MCQ2	0.52**	0.33**	0.05	0.34**	—	0.53**	0.52**	0.37**
College Exams								
AKT	0.48**	0.25**	0.11	0.33**	0.53**	—	0.66**	0.40**
KFP	0.48**	0.15**	-0.08	0.37**	0.52**	0.66**	—	0.43**
OSCE	0.40**	0.17**	-0.09	0.17*	0.37**	0.40**	0.43**	—

*p<0.05, **p<0.01.

AKT, Applied Knowledge Test; CR, clinical reasoning; CS, clinical skills; KFP, Key Feature Problems; MCQ, multiple choice questions; OSCE, Objective Structured Clinical Examination; SAQ, short answer questions.

Table 3 Comparison of predictive abilities of pre-entry exam scores on college exam scores

Variables							
Pre-entry exams	College exams	R ²	β	SE _B	B	t	P value
SJT	AKT	0.125	0.096	0.02	0.278	4.897	<0.001
MMI			0.015	0.005	0.175	3.081	0.002
CAAKT*		0.002	0.003	0.008	0.043	0.400	0.69
SJT	KFP	0.070	0.081	0.02	0.233	3.993	<0.001
MMI			0.008	0.005	0.090	1.547	0.123
CAAKT*		0.000	-0.002	0.008	-0.02	-0.184	0.854
SJT	OSCE	0.110	0.079	0.021	0.227	3.729	<0.001
MMI			0.017	0.005	0.204	3.338	0.001
CAAKT*		0.218	0.035	0.010	0.467	3.421	0.001

*The inclusion of CAAKT in the model with SJT and MMI yielded no results (an error was detected). Therefore, the predictive relationship between CAAKT and the college exams was conducted separately. Gender and origin were adjusted in the analyses and were not significant. β, unstandardised coefficient; AKT, Applied Knowledge Test; B, standardised coefficient; CAAKT, Candidate Assessment and Applied Knowledge Test; KFP, Key Feature Problems; MMI, Multiple-Mini Interviews; OSCE, Objective Structured Clinical Examination; P, significance level; SE_B, standard error of the coefficient; SJT, Situational Judgement Test.

the model at once to indicate how well they are able to predict each dependent variable (RACGP exam scores) and how much variance each of these independent variables explains in the dependent variable over and above the other independent variables adjusting for gender and origin. MCQ1 was predictive of the college AKT examination ($p<0.001$), KFP ($p<0.001$) and OSCE ($p=0.003$). Similarly, MCQ2 was also predictive of the college AKT

($p<0.001$), KFP ($p<0.001$) and OSCE ($p=0.014$). CR was predictive of KFP ($p=0.003$), but not AKT and OSCE. There was no significant relationship between SAQ and CS and the RACGP exams. Overall, the internal formative assessments accounted for 20.5% to 37.6% of the variance in performance in the RACGP exams.

Table 5 shows that failing MCQ1 was predictive of failing all RACGP exams, while failing MCQ 2 was

Table 4 Comparison of predictive abilities of internal formative exam scores on college exam scores

Variables							
Internal formative exams	College exams	R ²	β	SE _B	B	t	P value
MCQ1	AKT	0.358	0.027	0.008	0.252	3.527	<0.001
SAQ			0.000	0.005	0.004	0.060	0.952
CS			0.003	0.004	0.083	1.351	0.178
CR			0.008	0.002	0.120	1.906	0.058
MCQ2			0.036	0.007	0.355	4.981	<0.001
MCQ1	KFP	0.376	0.028	0.008	0.259	3.677	<0.001
SAQ			-0.004	0.005	-0.051	-0.0780	0.437
CS			-0.003	0.002	-0.099	-1.630	0.105
CR			0.012	0.004	0.184	2.963	0.003
MCQ2			0.036	0.007	0.349	4.964	<0.001
MCQ1	OSCE	0.205	0.029	0.010	0.267	3.020	0.003
SAQ			0.002	0.007	0.029	0.350	0.727
CS			-0.004	0.003	-0.106	-1.393	0.166
CR			0.001	0.005	0.020	0.261	0.795
MCQ2			0.022	0.009	0.219	2.483	0.014

Gender and origin were adjusted in the analyses and were not significant.

β, unstandardised coefficient; AKT, Applied Knowledge Test; B, standardised coefficient; CR, clinical reasoning; CS, clinical skills; KFP, Key Feature Problems; MCQ, multiple choice questions; OSCE, Objective Structured Clinical Examination; P, significance level; SAQ, short answer questions; SE_B, standard error of the coefficient.

Table 5 Failing internal formative assessment as predictors of failing each college examination

Variables							
Internal formative exams	College exams	β	SE _{β}	df	B	P value	95% CI
MCQ1							
Pass	AKT	Ref					
Fail		0.840	0.393	1	2.316	0.033	1.071 to 5.006
MCQ2							
Pass		Ref					
Fail		1.041	0.371	1	2.831	0.005	1.369 to 5.856
MCQ1							
Pass	KFP	Ref					
Fail		1.383	0.331	1	3.987	<0.001	2.084 to 7.626
MCQ2							
Pass		Ref					
Fail		1.149	0.317	1	3.153	<0.001	1.693 to 5.875
MCQ1							
Pass	OSCE	Ref					
Fail		1.243	0.517	1	3.46	0.016	1.259 to 9.545
MCQ2							
Pass		Ref					
Fail		-1.151	0.782	1	0.316	0.141	0.068 to 1.464

Gender and origin were adjusted in the analyses and were not significant.

β , unstandardised coefficient; AKT, Applied Knowledge Test; B, standardised coefficient; KFP, Key Feature Problems; MCQ, multiple choice questions; OSCE, Objective Structured Clinical Examination; P, significance level; SE _{β} , standard error of the coefficient.

predictive of failing only AKT and KFP. Candidates who failed MCQ1 exam were 2.3 times more likely to fail the AKT exam ($p=0.033$, OR=2.316, 95% CI 1.071 to 5.006). Participants who failed MCQ2 were 2.8 times more likely to fail the AKT exam ($p=0.005$, OR=2.831, 95% CI 1.369 to 5.856). Similarly candidates who failed the MCQ1 and MCQ2 were 4.0 times and 3.2 times more likely to fail the KFP exam ($p<0.001$, respectively). Participants who failed MCQ1 were 3.46 (95% CI 1.259 to 9.545) times more likely to fail the OSCE component ($p=0.016$), however, failure in MCQ2 was not predictive of failing the OSCE component.

DISCUSSION

This study aimed to assess the predictive role of the pre-entry SJT, MMI, CAAKT and the internally developed JCU GPT formative exams on trainees' performance in the RACGP examinations (end of training assessment). The findings of the study showed that SJT and MMI were significantly correlated with exam performance in the end of training exams. However, only SJT was predictive of performance in all end of training exams while MMI was predictive of AKT and OSCE but not KFP. CAAKT was correlated with and predictive of only the OSCE. Evaluation of the predictive abilities of the internally developed formative assessments showed that only the scores from

the orientation MCQ1 and the end of semester MCQ2 were predictive of performance in all the RACGP exams. Obtaining a high score on the MCQ1 and MCQ2 predicts a high score and an increased probability of passing the RACGP exams. Obtaining a fail as an outcome in the orientation MCQ1 was predictive of failing all RACGP examinations, while a failure in the MCQ2 was only predictive of failing the AKT and KFP components of the examinations.

Although SJT and MMI pre-entry assessments were correlated with the end of training examination, the correlations were comparatively lower than the internal formative assessment.²⁰ In addition, the current pre-entry assessment (CAAKT) that replaced SJT was found to be less correlated with the AKT and KFP components of the RACGP college exams compared with the previous assessment (SJT). The pre-entry exams were also less predictive of College exam performance in comparison to the internal formative MCQ and CR exams. The significant correlation between the pre-entry exams and college exams may in part be related to similarities in construct.²⁰ Patterson *et al* (2016) reported a significant correlation between clinical skills assessment and performance in OSCE.²⁰ However, interviews have been shown to be inconsistent predictors of future performance.²⁰ The poor predictive ability of the newly introduced CAAKT

buttresses the fact that pre-entry assessments may not be the best predictors of performance in subsequent exams during training.¹¹ Previous studies in undergraduate medical education have reported conflicting findings on the ability of interview/admission test scores to predict student academic performance during the undergraduate preclinical years and clinical years.^{26–28} While some authors reported that the admission tests were predictive of academic performance during the preclinical years,²⁶ others stated that the test was not predictive of performance in preclinical and clinical years.^{27–28} There are suggestions that it might be more useful to focus on post-enrolment factors.²¹

The evidence from this study indicates that the in-training formative assessments may be useful programmatic ‘assessment for learning’ tools² and also better predictors of academic performance in the summative assessment of the RACGP exams. While there are limited studies on the predictive role of in-training assessment in GP training, numerous studies in other specialities have identified the predictive role of in-training examinations in fellowship exams.^{29–31} The internal MCQs were designed as formative feedback tools to assist the trainees and trainers to assess mastery of content and competence while serving as a guidepost for the fellowship examination.³² Evidence from our study has enabled the JCU GPT programme to intervene early in providing additional support for trainees at risk of failing to progress through the training programme. Performance in the internal formative assessment has provided the most useful information and guidance for identifying trainees who may experience academic difficulty early in the programme.²⁹ With early identification of an individual at risk of failing the fellowship exam, the more time the training programme has to intervene and provide additional support.²⁹ Although the details are not included in this current study, those who failed the orientation MCQ1 had additional educational support programmes constructed for them which enabled them to identify and remediate knowledge gaps before sitting the end of semester MCQ2 and subsequent fellowship examinations.

Furthermore, timing of in-training assessments is very important as they play a significant role in early identification of poorly performing trainees who are at risk of failing RACGP assessment. It is generally stated that the earlier the issue is identified and clarified, the earlier a process of assistance/academic support can be provided.²⁹ Current evidence suggests that in training assessments that are properly structured foster active learning.³³ For these assessments to be effective, it is important for them to be ongoing activities rather than an end of training process. These in-training assessments can be used as academic markers to identify and provide support for struggling trainees, thereby, providing the opportunity for identified gaps to be addressed through early support programmes.³⁴

This study reflects the benefit of ongoing in-training formative assessments that assess both clinical and

theoretical knowledge³⁵ at one GP RTO, the findings may be applicable to other GP training organisations in Australia. In addition, other speciality training organisations may want to consider investigating the role of in-training formative assessment using MCQs which have proven to be predictive of performance in fellowship examinations. As healthcare systems increasingly depend on service provided by junior or trainee doctors, early detection (and remediation) of poor performance is vital for both patient care and career progression of the individual doctor.³⁶

Strengths and limitations

This study is the first study to investigate the predictive role of both the pre-entry exams, and in training assessment on GP trainees’ performance in the end of training assessments at a GP RTO that uses a distributed training model. However, our study has some limitations that need to be taken into consideration. First, the study was conducted at only one RTO. Second, generalisation of the study findings to other settings may be limited by the retrospective nature of the study. Future studies could consider investigating the predictive value of the in-training exams over a longer period of time. Additionally, the effectiveness of the academic support provided to underperforming registrars could be further explored.

CONCLUSION

The findings of this study suggest that the internally developed formative MCQs are valuable assessment instruments which could assist in the early identification of GP trainees who are at risk of underperformance in their RACGP fellowship exams and subsequently foster early planning for appropriate supervision and additional support for this group of trainees.

Twitter Faith Alele @D_Faith and Bunmi S Malau-Aduli @bmaluaduli

Acknowledgements The authors acknowledge all the James Cook University General Practice registrars who participated in the study. Ms Allison Bonner, Ms Kimberley Martinsen and Ms Kath Paton are appreciated for facilitating the data collection process.

Contributors BSMA, PH and PAT conceived and designed the study. BSMA, PH and PAT supervised the data collection. BSMA and FA conducted the data analysis and PH and PAT contributed to the interpretation of the data. BSMA, MA and FA wrote the first draft of the manuscript. All authors contributed to the writing and review of the manuscript and approved the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethics approval (H6771) for this research was granted by the James Cook University Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content

includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Faith Alele <http://orcid.org/0000-0001-5570-3641>

Bunmi S Malau-Aduli <http://orcid.org/0000-0001-6054-8498>

REFERENCES

- Guerrasio J, Garrity MJ, Aagaard EM. Learner deficits and academic outcomes of medical students, residents, fellows, and attending physicians referred to a remediation program, 2006-2012. *Acad Med* 2014;89:352-8.
- Prideaux D, Roberts C, Eva K, et al. Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 conference. *Med Teach* 2011;33:215-23.
- Royal Australian College of General Practitioners. RACGP education exam report 2018.2 AKT, 2018. Available: <https://www.racgp.org.au/FSDEDEV/media/documents/Education/Registrars/Fellowship%20Pathways/Exams/AKT-20182-exam-report.pdf> [Accessed 01 May 2020].
- Dyrbye LN, Thomas MR, Shanafelt TD. Medical student distress: causes, consequences, and proposed solutions. *Mayo Clin Proc* 2005;80:1613-22.
- Cleland J, Leggett H, Sandars J, et al. The remediation challenge: theoretical and methodological insights from a systematic review. *Med Educ* 2013;47:242-51.
- Winston KA, van der Vleuten CPM, Scherpbier AJJA. Prediction and prevention of failure: an early intervention to assist at-risk medical students. *Med Teach* 2014;36:25-31.
- Curtis E, Wikaire E, Jiang Y, et al. Examining the predictors of academic outcomes for Indigenous Māori, Pacific and rural students admitted into medicine via two equity pathways: a retrospective observational study at the University of Auckland, Aotearoa New Zealand. *BMJ Open* 2017;7:e017276.
- Grainger B, Yelder J, Reid P, et al. Predictors of medical student remediation and their underlying causes: early lessons from a curriculum change in the University of Auckland medical programme. *N Z Med J* 2017;130:73-82.
- Gray C, Toms N. A holistic remediation intervention for struggling undergraduate medical students affords sustained progress test performance recovery. *MedEdPublish* 2018;7.
- Maslov Kruzicevic S, Barisic KJ, Banozic A, et al. Predictors of attrition and academic success of medical students: a 30-year retrospective study. *PLoS One* 2012;7:e39144.
- Sladek RM, Bond MJ, Frost LK, et al. Predicting success in medical school: a longitudinal study of common Australian student selection tools. *BMC Med Educ* 2016;16:187.
- Cleland J, Cilliers F, van Schalkwyk S. The learning environment in remediation: a review. *Clin Teach* 2018;15:13-18.
- Kalet A, Chou CL, Ellaway RH. To fail is human: remediating remediation in medical education. *Perspect Med Educ* 2017;6:418-24.
- Bennion LD, Durning SJ, LaRochelle J, et al. Untying the Gordian knot: remediation problems in medical schools that need remediation. *BMC Med Educ* 2018;18:120.
- Frellsen SL, Baker EA, Papp KK, et al. Medical school policies regarding struggling medical students during the internal medicine clerkships: results of a national survey. *Acad Med* 2008;83:876-81.
- Isaacs G. Assessment for learning. Teaching and educational development Institute, the University of Queensland 2001.
- Patterson F, Ashworth V, Zibarras L, et al. Evaluations of situational judgement tests to assess non-academic attributes in selection. *Med Educ* 2012;46:850-68.
- Eva KW, Macala C, Fleming B. Twelve tips for constructing a multiple mini-interview. *Med Teach* 2019;41:510-6.
- Mercer A. Who gets in, and who doesn't: selecting medical students: an Australian case study. *Education Research and Perspectives* 2009;36.
- Patterson F, Rowett E, Hale R, et al. The predictive validity of a situational judgement test and multiple-mini interview for entry into postgraduate training in Australia. *BMC Med Educ* 2016;16:87.
- Shulruf B, Poole P, Wang GY, et al. How well do selection tools predict performance later in a medical programme? *Adv Health Sci Educ Theory Pract* 2012;17:615-26.
- Schuwirth L, van der Vleuten C, Durning SJ. What programmatic assessment in medical education can learn from healthcare. *Perspect Med Educ* 2017;6:211-5.
- Patterson F, Lievens F, Kerrin M, et al. Designing selection systems for medicine: the importance of balancing predictive and political validity in high-stakes selection contexts. *International Journal of Selection and Assessment* 2012;20:486-96.
- Roberts C, Clark T, Burgess A, et al. The validity of a behavioural multiple-mini-interview within an assessment centre for selection into specialty training. *BMC Med Educ* 2014;14:169.
- Lievens F, Buyse T, Sackett PR. The operational validity of a video-based situational judgment test for medical college admissions: illustrating the importance of matching predictor and criterion construct domains. *J Appl Psychol* 2005;90:442-52.
- Migliaretti G, Bozzaro S, Siliquini R, et al. Is the admission test for a course in medicine a good predictor of academic performance? A case-control experience at the school of medicine of Turin. *BMJ Open* 2017;7:e017417.
- Yates J, James D. The value of the UK clinical aptitude test in predicting pre-clinical performance: a prospective cohort study at Nottingham medical school. *BMC Med Educ* 2010;10:55.
- Yates J, James D. The UK clinical aptitude test and clinical course performance at Nottingham: a prospective cohort study. *BMC Med Educ* 2013;13:32.
- Johnson GA, Bloom JN, Szczotka-Flynn L, et al. A comparative study of resident performance on standardized training examinations and the American Board of ophthalmology written examination. *Ophthalmology* 2010;117:2435-9.
- Juul D, Schneidman BS, Sexson SB, et al. Relationship between Resident-In-Training examination in psychiatry and subsequent certification examination performances. *Acad Psychiatry* 2009;33:404-6.
- Lohr KM, Clauser A, Hess BJ, et al. Performance on the adult rheumatology in-training examination and relationship to outcomes on the rheumatology certification examination. *Arthritis Rheumatol* 2015;67:3082-90.
- Schuwirth LW and van der Vleuten CP. The use of progress testing. *Perspect Med Educ* 2012;1:24-30.
- Ringsted C, Henriksen AH, Skaarup AM, et al. Educational impact of in-training assessment (ITA) in postgraduate medical education: a qualitative study of an ITA programme in actual practice. *Med Educ* 2004;38:767-77.
- Hays R, Wellard R. In-Training assessment in postgraduate training for general practice. *Med Educ* 1998;32:507-13.
- Stewart R, Cooling N, Emblen G, et al. Early predictors of summative assessment performance in general practice post-graduate training: a retrospective cohort study. *Med Teach* 2018;40:1166-74.
- Barrett A, Galvin R, Steinert Y, et al. A BEME (Best Evidence in Medical Education) systematic review of the use of workplace-based assessment in identifying and remediating poor performance among postgraduate medical trainees. *Syst Rev* 2015;4:65.