

# Early predictors of need for remediation in the Australian general practice training program: a retrospective cohort study

Parker Magin<sup>1,2</sup>  · Rebecca Stewart<sup>3</sup> · Allison Turnock<sup>3,4</sup> · Amanda Tapley<sup>2</sup> · Elizabeth Holliday<sup>1,5</sup> · Nick Cooling<sup>4</sup>

Received: 8 February 2016 / Accepted: 13 October 2016 / Published online: 21 October 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** Underperforming trainees requiring remediation may threaten patient safety and are challenging for vocational training programs. Decisions to institute remediation are high-stakes—remediation being resource-intensive and emotionally demanding on trainees. Detection of underperformance requiring remediation is particularly problematic in general (family) practice. We sought to establish early-training assessment instruments predictive of general practice (GP) trainees' subsequently requiring formal remediation. We conducted a retrospective cohort study of trainees from a large Australian regionally-based GP training organization. The outcome factor was requirement for formal remediation. Independent variables were demographic factors and a range of formative assessments conducted immediately prior to or during early-stage training. Analyses employed univariate and multivariate logistic regression of each predictor assessment modality with the outcome, adjusting for potential confounders. Of 248 trainees, 26 (10.5 %) required formal remediation. Performance on the Colleague Feedback Evaluation Tool (entailing feedback from a trainee's clinical colleagues on clinical performance, communication and probity) and External Clinical Teaching Visits (half-day sessions of the trainee's clinical consultations observed directly by an experienced GP), along with non-Australian primary medical qualification, were significantly associated with requiring remediation. There was a non-significant trend for association with performance on the Doctors Interpersonal Skills Questionnaire (patient feedback on interpersonal elements of the consultation). There were no significant associations with entry-selection scores or formative exam or assessment

---

✉ Parker Magin  
parker.magn@newcastle.edu.au

<sup>1</sup> School of Medicine and Public Health, University of Newcastle, Newbolds Building, University Drive, Callaghan, NSW 2308, Australia

<sup>2</sup> General Practice Training Valley to Coast, Mayfield, NSW, Australia

<sup>3</sup> Tropical Medical Training, Townsville, QLD, Australia

<sup>4</sup> School of Medicine, University of Tasmania, Hobart, TAS, Australia

<sup>5</sup> Clinical Research Design, IT and Statistical Support Unit, Hunter Medical Research Institute, Newcastle, NSW, Australia

scores. Our finding that ‘in vivo’ assessments of complex behaviour, but not ‘in vitro’ knowledge-based assessments, predict need for remediation is consistent with theoretical understanding of the nature of remediation decision-making and should inform remediation practice in GP vocational training.

**Keywords** Family practice · General practice · Education, medical, graduate · Professional competence · Remedial teaching · Educational measurement

## Introduction

Under performing trainees in vocational training programs in hospital or community practice are common (Yao and Wright 2000, 2001; Reamy and Harman 2006), and demands for greater patient safety have raised the stakes for their early detection, support and remediation (Byrne 2012). Underperforming trainees requiring remediation are challenging to vocational training programs and to fellow trainees (Yao and Wright 2001).

Remediation has been defined as a purposeful intervention including “any form of additional training, supervision, or assistance above that required for a typical resident.” (Riebschleger and Haftel 2013). The term can also be used to define the status of a training doctor “falling from good standing, to remediation, to probation, and finally to termination” (Weizberg et al. 2015). Remediation can be viewed as the ‘elevation of feedback and evaluation’ to a higher level (Kimatian and Lloyd 2008).

Hauer et al. (2009), described remediation as occurring in 3 steps: (1) the identification of deficiencies in an individual’s performance; (2) an attempt to provide remedial education to that individual; and (3) reassessment to determine the impact of the education. In the first of these three steps it is rare for trainees to self-report their problems. Most problems are detected by direct observation or critical incidents (Yao and Wright 2000) rather than purposive measures to detect underperformance.

Detection of under-performance may be more difficult in the vocational training setting compared with education for a primary medical degree due to the relatively high clinical responsibilities of trainees, less direct observation, and less systematic formative assessment (Hauer et al. 2009). Within postgraduate settings there also can be further differences in detection of underperforming trainees. In general practice (family practice) vocational training early detection can be more challenging than hospital vocational training settings (Gladman 2011). General practice training is often conducted in geographically-dispersed (often rural) small independent practices with doctors working in relative isolation, even within a practice. In Australia, general practitioner (GP) trainees have considerable autonomy compared to hospital-based specialty training. This autonomy has implications for detection of under-performance per se, and also for detection of possible trainee factors involved in under-performance, such as stress. While more than half of all Australian GP trainees confidentially reported significant levels of stress they are often not detected by their trainers (Larkins et al. 2003).

Thus, in the general practice context, detection of underperforming trainees may be more difficult, and resources for intervention less available, than in training settings with more staff, closer observation of trainees, and systematic monitoring. The impact on patient care of non-detection of under-performance can also be more problematic in general practice given the greater autonomy of trainees in general practice.

In Australia, the national general practice vocational training program requires trainees to complete a three to four year full-time equivalent (FTE) program with a general practice component of a minimum of 18 months FTE placement in community general practice for the Royal Australian College of GPs (RACGP) or, alternatively, six months general practice for the Australian College of Rural and Remote Medicine (ACRRM). The remaining terms are completed in hospital contexts or pursuing areas of special interest. The general practice component is conducted within an apprenticeship-like model but with considerable independence of practice. Though having recourse to clinical advice and assistance from experienced GP supervisors, in many respects trainees (in Australia, registrars) have autonomy and responsibilities similar to senior GPs—in areas such as prescribing rights and the ordering of investigations, and in Medicare (universal government-funded health insurance) billing. Separate not-for-profit accredited organizations called Regional Training Providers (RTPs, now known as Regional Training Organizations), using RACGP and ACRRM vocational training standards, oversee the coordination and delivery of GP training in the community, hospitals and dedicated classroom block teaching. It is this regional training structure that is responsible for the remediation of underperforming registrars.

Australian GP trainees must pass the summative Fellowship examinations of the RACGP or ACRRM, to be recognized as specialist medical practitioners by the Australian Health Practitioner Regulation Agency. Prior to these summative Fellowship examinations, Australian GP trainees' performance is subject to a continuous program of formative assessment activities by their RTP (Hays and Wellard 1998; Ingham et al. 2015). These formative assessment activities have the potential to detect underperformance and the need for remediation or career re-alignment prior to sitting the summative examinations.

Underperformance of vocational trainees can be due to poor clinical judgment, knowledge and/or skills, but underlying causes are usually multifactorial (Firth-Cozens and King 2006; Williams et al. 2009; Borkett-Jones and Morris 2010) and there are recognized early behavioral markers of difficulty (Paice and Orton 2005; Beaman and King 2009; Black and Welch 2009).

Remediation typically occurs relatively late in the course of the trainee's clinical training. It is very resource intensive for training organisations (Mugford 2003; Durning et al. 2011; Roberts et al. 2012) and is often emotionally demanding on the trainee. The trainee under remediation often feels stigmatized and reluctant to participate and the trainers may be unwilling or unable to deliver remediation (Bahrami 1997; Leap and Fromson 2006; Hauer et al. 2009). Thus the decision to institute remediation is a high-stakes one and outcomes, especially using traditional analytical approaches, are often poor and protracted (Williams, et al. 2009; Durning et al. 2011).

There are, however, compelling reasons why RTPs need to provide early and effective remediation to underperforming trainees to: (a) reduce the risk of significant harm to the public (Papadakis et al. 2008; Dwyer et al. 2011), (b) avert many of the financial and personal costs inherent in remediation, (c) improve summative assessment outcomes and (d) provide the option for early career re-alignment.

Given the stakes involved for the trainee, training provider and public it is important that reliable early predictors of poor performance be established. They may lead to earlier intervention and resolution of performance difficulties. Some have stated that early detection of difficulties should be regarded as the 'gold standard' for educational supervision (Evans et al. 2010).

Accurate decision making around the need for remediation and associated training progression is crucial for fairness to the trainee, defensibility for the training organization

and the promotion of trust in patients. Decision making theory (Edwards 1954) may be useful for making such high stakes choices. Contemporary decision support models, such as Naturalistic Decision Making (NDM) from the field of cognitive psychology, appear pertinent to complex and error-prone assessment decisions in vocational training. NDM investigates “how experienced people, working as individuals or groups in dynamic, uncertain and often fast paced environments, identify and assess their situation, make decisions and take actions whose consequences are meaningful to them and to the larger organization in which they operate” (Zsombok and Klein 2014, p. 19).

While NDM is principally used to inform clinical decision making it can also be helpful in assessment decisions regarding training progression as it integrates and weighs up risk factors and other relevant data before crucial decisions are made.

The trustworthiness of the decision making around the recommendation to initiate remediation in vocational trainees can also be improved with purposeful design of decision making strategies including triangulation (van der Vleuten et al. 2012), ‘critical dialogues’ between assessors (Van der Vleuten et al. 2010; Gladman 2011) standardization of assessment (Epstein 2007) and careful choice of assessment tools with an emphasis on those that measure overall quality of care (Bingham and Crampton 2011).

Inherent in implementation of early-intervention strategies incorporating these elements within an NDM-informed approach is an appreciation of the validity of individual early assessment components in predicting subsequent under-performance requiring formal remediation. There have been a few studies describing the value of specific formative activities that screen for competence in the vocational setting, including: (a) Work based Assessments (WBA) e.g. the mini-Peer Assessment Tool (Black and Welch 2009), the mini-CEX (Mitchell et al. 2011), Case-based Discussion (Mitchell et al. 2011) and external clinical teaching visits (Yao and Wright 2000; Gladman 2011); (b) periodic knowledge tests e.g. progress testing (Dijksterhuis et al. 2009) and other in-training written examinations (Garibaldi et al. 1994); (c) standardised periodic supervisor and trainee questionnaires (Bingham and Crampton 2011); and (d) informal supervisor feedback, often via telephone (Gladman 2011). While being indicators of competence, many of these ‘screening tools’ come too late in training, under-report underperformance (Bingham and Crampton 2011) and are considered to be an unreliable way of screening for future performance (Mitchell et al. 2011). Importantly, specific tools such WBA may not offer any additional advantage over informal techniques (Barrett et al. 2015).

Centralized electronic analysis and reporting systems run by training organisations, also called ‘at-risk trainee dashboards’, are capable of summing data from multiple sources and may be useful adjuncts to detecting underperformance (Black and Welch 2009) Such reporting systems provide a substrate for the identification, assessment, decision-making and action within dynamic, uncertain environments demanded by NDM. Some centres use such systems to categorise underperforming trainees according to ‘level of concern’ e.g. Level 1 (Minor), Level 2 (Serious), and Level 3 (Concerns that threaten progression of training) (Beaman and King 2009).

Overall, there appears to be a lack in confidence (from these limited number of studies) as to whether trainees requiring remediation can be accurately predicted from existing competence or demographic data. While the gold standard of more extensive screening, from multiple sources, from trained experts, using aggregated data e.g. programmatic assessment model (van der Vleuten et al. 2012) would provide more reliable detection of underperformance, perhaps there are some key indicators that could provide the substrate for an efficient path to early detection and thence to less resource-intensive and stigmatizing remediation.

We aimed to address this issue by conducting an in-depth analysis of the association of early training factors, including formative assessments, with GP vocational trainees' subsequent requirement for remediation. This was performed in one of Australia's 17 RTPs. This RTP routinely records a broad range of trainee demographic and performance indicators.

## Methods

We conducted a retrospective cohort study using routinely collected data.

### Setting

The study was conducted in one of Australia's 17 RTPs. RTPs are government-funded, not-for-profit, geographically-defined organisations charged with delivery of general practice vocational training. The study RTP covers a large area of northern Queensland including urban to very remote regions. In 2015 264 trainees were in active training with the RTP and they were supervised in 250 practices with supervisory input by 404 accredited supervisors.

### Participants

The inclusion criterion was being a trainee enrolled with the RTP and commencing training during 2010–2014.

Exclusion criteria were having transferred into or out of the RTP during training.

### Outcome factor

The dichotomous outcome was a trainee being required to undergo formal remediation during their vocational training program with the RTP. Such formal remediation is usually externally-funded but may be internally funded by the RTP. A decision that remediation is required occurs in Term 2 or 3 of training (each term is 6-months FTE). The requirement for remediation is a decision of a committee of senior executive, educational and administrative RTP staff. It is arrived at after consideration of multiple factors. These include quantifiable in-training assessments, but encompass a far wider range of often subjective and anecdotal considerations which are not quantified (nor can be practicably quantified).

### Predictor variables

- Performance on the annual national training program selection assessment (a continuous variable, scored as deciles with 10 being highest performance).
- Review and Readiness Multiple Choice Questionnaire (R and R MCQ). R and R MCQ is an internal RTP MCQ-based assessment of trainee clinical knowledge undertaken prior to the commencement of Term 1. R and R MCQ was a continuous variable, scored as a percentage.
- Review and Readiness Consultation (R and R Cons). R and R Cons is an internal RTP role-play-based assessment of trainee clinical consultation communication skills

undertaken prior to the commencement of Term 1. R and R Cons was scored as a dichotomous outcome: Issues identified/No issues identified.

- Colleague Feedback Evaluation Tool (CFET) collects feedback from 12-15 of a trainee's clinical colleagues on their clinical performance, communication and probity (Campbell, Narayanan, Burford, and Greco 2010). CFET was a continuous variable, scored as a percentage.
- Performance on the Doctors Interpersonal Skills Questionnaire (DISQ). DISQ collects feedback from 30 patients. This feedback focuses specifically on the interpersonal elements of the interaction (Campbell et al. 2010). Patients complete the assessment immediately post-consultation with the trainee. DISQ was a continuous variable, scored as a percentage.
- Performance at External Clinical Teaching Visits (ECTVs) in which a half-day session of the trainee's clinical consultations is observed directly by an experienced GP (designated by the regional training provider, and external to the trainee's practice). The ECTV was scored as a dichotomous outcome: issues identified/no issues identified (in either of the two ECTVs performed in Term 1).

## Covariates

Demographic variables were trainee age, trainee gender, College in which enrolled (RACGP or ACRRM), and country of trainee primary medical qualification (Australian or non-Australian). Other trainee-level independent variables were whether training was undertaken part-time in Term 1, the number of days leave taken in Term 1 (dichotomized to  $\leq 10$  or  $>10$ ), and the number of education hours (organized by the RTP for trainees during education 'release' from the practice) completed in Term 1 (a continuous variable).

The practice-level variable was the Australian Standard Geographical Classification-Remoteness Areas (ASGC-RA) remoteness category of the trainee's Term 1 practice location (categorized for our analyses as RA-1&2 (major cities and inner regional), RA-3 (outer regional) and RA-4&5 (remote and very remote).

All predictors and covariates (excepting pre-training assessments) related to the first six months of GP training.

## Data extraction

Predictor variables and covariates from 2012 onwards were already held in an electronic database. Other data were extracted from routinely collected and recorded training data and entered in the existing database. This pre-2012 data was extracted from both electronic and hard-copy records.

## Statistical analyses

*Descriptive statistics* Continuous variables were described using means and standard deviations. Group differences in means were tested using ANOVA. Categorical variables were described using counts and percentages. Group differences were tested using Chi square or Fisher's Exact test, as appropriate.

*Modelling approach* Due to a relatively small sample size and some missing data for a number of predictors and covariates, multivariate models were fitted separately for each of the six predictors (providing individual association of each predictor with the outcome

after adjusting for potential confounders). Analyses of both complete case and multiply imputed data were performed.

For the complete case analyses (the primary analysis), univariate logistic regression was used to screen predictors and covariates for univariate association with the outcome. Covariates showing univariate association with the outcome at  $p < 0.2$  were included in initial multivariate models for each predictor. Model reduction was then performed, removing covariates not associated with the outcome at  $p < 0.2$  and whose removal did not produce substantive changes in other model coefficients. The rationale for model reduction was to avoid model overfitting. For the final model for each predictor, goodness-of-fit was checked using the Hosmer and Lemeshow test, and by assessing residuals and influence statistics. Predictive utility of the model was assessed using the c-statistic (Area under the Receiver Operator Characteristic Curve: AUC) and its 95 % confidence interval.

The multiple imputation analyses assumed data were missing at random. A total of 25 imputed datasets were generated using the MI procedure within SAS, implementing a fully conditional specification (FCS). Continuous variables were imputed using linear regression, binary variables were based on a logistic model (logit link) and categorical variables were imputed using a discriminant function method. Models were then fitted to the individual imputed datasets and parameter estimates were combined using the SAS MIANALYZE procedure.

The modelling approach for multiply imputed data was similar to complete case analyses. Variable selection was based on complete case results, and multivariate models were fitted to the imputed datasets. The one exception is that one additional model was fitted for the outcome (remediation required/not required). This model included all predictors showing evidence of association (at  $p < 0.2$ ) with the outcome in the multivariate model for that predictor. Also included were all covariates showing univariate association (at  $p < 0.2$ ) with the outcome. Model reduction was performed as described above.

All statistical analyses were programmed using SAS v9.4.

## Ethics approval

Approval was from the University of Newcastle Human Research and Ethics Committee (H-2014-0402).

## Results

There were 248 trainees who met the inclusion criteria. The demographics of these trainees are presented in Table 1.

Of these 248 trainees, 26 required formal remediation (10.5 %). The univariate associations of predictors and covariates with requiring remediation are presented in Table 2. The predictor variables CFET (lower score) and ECTV ('issue identified') and the covariate non-Australian primary medical qualification were all significantly associated with requiring remediation.

The six multivariate models with outcome factor 'needed remediation', one for each predictor variable, are presented in Table 3. The c-statistic for each model is also presented. These are the complete case analysis models. The predictor variables CFET (lower score, OR 0.84 for each percentage point of score) and ECTV ('issue identified', OR 10.1) were significantly associated with requiring remediation. There was a non-significant trend

**Table 1** Characteristics of participating trainees and their formative evaluations

Variable	Class/statistic	Total (N = 248)*
Age	Mean (SD)	33.39 (6.671)
DISQ	Mean (SD)	87.6 (5.844)
CFET	Mean (SD)	86.97 (5.594)
R&R MCQ	Mean (SD)	77.13 (10.4)
ED hours: 1st 6 month	Mean (SD)	71.48 (25.44)
Gender	Male	98 (40 %)
	Female	150 (60 %)
R&R consultation	No action	168 (86 %)
	Assistance/action	28 (14 %)
ECT	No action	152 (85 %)
	Assistance/action	27 (15 %)
Selection decile	1	24 (9.9 %)
	2	21 (8.7 %)
	3	29 (12 %)
	4	24 (9.9 %)
	5	21 (8.7 %)
	6	27 (11 %)
	7	24 (9.9 %)
	8	23 (9.5 %)
	9	21 (8.7 %)
	10	28 (12 %)
College	ACRRM	69 (28 %)
	RACGP	176 (72 %)
Country of training	Australia	196 (79 %)
	International	52 (21 %)
>10 Leave days: 1st 6 month	No	183 (86 %)
	Yes	30 (14 %)
Part-time: 1st 6 month	No	176 (83 %)
	Yes	37 (17 %)
ASGC remoteness	MajorCity/InnerRegional	17 (8.0 %)
	OuterRegional	170 (80 %)
	Remote/VeryRemote	25 (12 %)

*DISQ* Doctors Interpersonal Skills Questionnaire, *CFET* Colleague Feedback Evaluation Tool, *R and R MCQ* Review and Readiness Multiple Choice Questionnaire, *R and R Cons* Review and Readiness Consultation, *ECTVs* External Clinical Teaching Visits, *ACRRM* Australian College of Rural and Remote Medicine, *RACGP* Royal Australian College of General Practitioners, *ASGC* Australian Standard Geographic Classification

\* Numbers may not total 248 due to missing data

( $p = 0.067$ ) for an association with DISQ (lower scores). Being a non-Australian graduate was a significant association in all but the DISQ model (where  $p = 0.055$ ), with ORs of 2.94 to 8.10.

The six imputed data multivariate models (detailed results not shown) provided similar results to the complete case analyses, with CFET and ECTV being the only predictor variables significantly associated with requiring remediation.

**Table 2** Univariable associations of a trainee requiring remediation (associations with predictor variables and co-variables)

Variable	Class/statistic	No (n = 222)	Yes (n = 26)	OR (95 % CI)	p value
Age	Mean (SD)	33.16 (6.378)	35.4 (8.77)	1.04 (0.99, 1.10)	0.1175
DISQ	Mean (SD)	88.02 (5.34)	85.22 (7.907)	0.93 (0.85, 1.01)	0.0662
CFET	Mean (SD)	87.65 (5.317)	82.69 (5.588)	0.86 (0.78, 0.96)	0.0053
R&R MCQ	Mean (SD)	76.98 (9.787)	78.36 (14.78)	1.01 (0.97, 1.06)	0.5564
ED hours: 1st 6 month	Mean (SD)	71.16 (25.66)	73.77 (24.23)	1.00 (0.99, 1.02)	0.6240
Gender	Male	88 (40 %)	10 (38 %)		
	Female	134 (60 %)	16 (62 %)	1.05 (0.46, 2.42)	0.9075
R&R consultation	No action	152 (87 %)	16 (76 %)		
	Assistance/action	23 (13 %)	5 (24 %)	2.07 (0.69, 6.18)	0.1946
ECT	No action	139 (90 %)	13 (52 %)		
	Assistance/action	15 (9.7 %)	12 (48 %)	8.55 (3.31, 22.1)	<0.0001
Selection decile	1	21 (9.7 %)	3 (12 %)		
	2	17 (7.9 %)	4 (15 %)		.
	3	26 (12 %)	3 (12 %)		.
	4	23 (11 %)	1 (3.8 %)		.
	5	18 (8.3 %)	3 (12 %)		.
	6	22 (10 %)	5 (19 %)		.
	7	23 (11 %)	1 (3.8 %)		.
	8	21 (9.7 %)	2 (7.7 %)		.
	9	19 (8.8 %)	2 (7.7 %)		.
	10	26 (12 %)	2 (7.7 %)	0.93 (0.81, 1.07)	0.3278
College	ACRRM	64 (29 %)	5 (19 %)		
	RACGP	155 (71 %)	21 (81 %)	1.73 (0.63, 4.80)	0.2891
Country of training	Australia	182 (82 %)	14 (54 %)		
	International	40 (18 %)	12 (46 %)	3.90 (1.68, 9.07)	0.0016
>10 Leave days: 1st 6 months	No	163 (87 %)	20 (77 %)		
	Yes	24 (13 %)	6 (23 %)	2.04 (0.74, 5.58)	0.1664
Part-time: 1st 6 months	No	155 (83 %)	21 (81 %)		
	Yes	32 (17 %)	5 (19 %)	1.15 (0.40, 3.29)	0.7895
ASGC remoteness	MajorCity/ InnerRegional	13 (7.0 %)	4 (15 %)		
	OuterRegional	149 (80 %)	21 (81 %)	0.46 (0.14, 1.54)	0.2061
	Remote/ VeryRemote	24 (13 %)	1 (3.8 %)	0.14 (0.01, 1.34)	0.0874

*DISQ* Doctors Interpersonal Skills Questionnaire, *CFET* Colleague Feedback Evaluation Tool, *R and R MCQ* Review and Readiness Multiple Choice Questionnaire, *R and R Cons* Review and Readiness Consultation, *ECTVs* External Clinical Teaching Visits, *ACRRM* Australian College of Rural and Remote Medicine, *RACGP* Royal Australian College of General Practitioners, *ASGC* Australian Standard Geographic Classification

**Table 3** Multivariate models with outcome ‘trainee requiring mediation’ for each the six predictor variables (i.e. formative assessment modalities)

Variable	Class	OR (95 % CI)	<i>p</i>	Area under the curve (95 % CI)
DISQ		0.93 (0.85, 1.01)	0.0686	0.667 (0.503, 0.830)
Country of training	International	2.94 (0.98, 8.82)	0.0551	
CFET		0.84 (0.74, 0.95)	0.0056	0.853 (0.746, 0.961)
Country of training	International	8.10 (1.85, 35.4)	0.0054	
>10 Leave days: 1st 6 months	Yes	3.61 (0.74, 17.7)	0.1134	
R&R consultation	Assistance/action	1.53 (0.44, 5.27)	0.5028	0.693 (0.562, 0.825)
Country of training	International	5.43 (1.99, 14.8)	0.0009	
>10 Leave days: 1st 6 months	Yes	2.53 (0.75, 8.54)	0.1344	
R&R MCQ		0.99 (0.95, 1.04)	0.6630	0.675 (0.531, 0.818)
Country of training	International	5.43 (2.01, 14.7)	0.0009	
>10 Leave days: 1st 6 months	Yes	2.33 (0.71, 7.63)	0.1614	
ECT: in 1st 6 month	Assistance/action	10.1 (3.69, 27.8)	<0.0001	0.764 (0.658, 0.870)
Country of training	International	2.95 (1.11, 7.87)	0.0308	
Selection decile		0.96 (0.83, 1.12)	0.6226	0.678 (0.559, 0.798)
Country of training	International	4.08 (1.67, 10.0)	0.0021	
>10 Leave days: 1st 6 months	Yes	2.13 (0.74, 6.12)	0.1590	

*DISQ* Doctors Interpersonal Skills Questionnaire, *CFET* Colleague Feedback Evaluation Tool, *R and R MCQ* Review and Readiness Multiple Choice Questionnaire, *R and R Cons* Review and Readiness Consultation, *ECTVs* External Clinical Teaching Visits

**Table 4** Multivariate model with outcome ‘trainee requiring mediation’ (including all six predictor variables plus co-variables, and using multiply imputed data)

Variable	Class	OR (95 % CI)	<i>p</i> value	Area under the curve (95 % CI)
ECT	Assistance/action	8.20 (3.03, 22.2)	<0.0001	0.769 (0.666, 0.871)
Country of training	International	3.50 (1.38, 8.83)	0.0081	

In the full multivariate model (all predictor variables plus covariates) using multiply imputed data (see Table 4) only ECTV (‘issue identified’) and non-Australian primary medical qualification were significantly associated with requiring remediation.

## Discussion

### Summary of major findings

In this retrospective study from a single regional GP vocational training provider we found that 10.5 % of trainees required remediation. We found that, of six formative in-training Term 1 assessment activities, performance on only two (the CFET and ECTV) were significantly associated with requiring formal remediation later in training. There was a non-significant trend for an association of DISQ scores with requiring remediation.

A striking finding is that these three assessments involve ‘in vivo’ evaluations of trainees’ performance in integrated clinical tasks involving patient care in the clinical setting. The assessments are made by parties across the spectrum of those interacting with or observing the trainee during their clinical work—patients (DISQ), work colleagues (CFET), and external observing GPs (ECTV). ‘In vitro’ assessments, including those testing more ‘academic’ constructs including knowledge, were not associated with requiring remediation. These findings can be seen to be consistent with results of assessments that target higher levels of Miller’s pyramid (Miller 1990) being predictive of later need for remediation, while those assessing lower levels are not predictive.

### Comparison with previous literature

Despite the widespread use and acceptance of workplace based assessment (WBA) in teaching hospitals in the UK and Australia, there has been surprisingly limited research into the impact of WBA on trainee performance (Miller and Archer 2010). Specifically, the utility of WBA tools in detecting underperforming trainees and in measuring a subsequent improvement in their performance remains unclear (Barrett et al. 2015). Of all the WBA assessment tools, multisource feedback (MSF) has shown the most potential to improve performance (Violato et al. 2008; Black and Welch 2009). Our finding of the utility of CFET, a form of MSF, is consistent with this limited published literature but also provides evidence for its potential use as a screening tool in identifying poor early-training performance.

We found the ECTV to be predictive of later need for remediation. The ECTV is a well-accepted, but under researched, WBA tool in the Australian GP training context for over 20 yrs (Hays and Peterson 1996). It provides direct observation of consultations, and timely, structured feedback between an experienced visiting medical educator and trainee. It also has the likely hallmarks of a useful screening tool for underperformance (Barrett et al. 2015) and is used in all training practices and all RTPs in Australia. These useful features need further elucidation.

The most similar tool to the ECTV, in other jurisdictions, is the mini-CEX. The mini-CEX also involves direct observation of a trainee in a focused clinical encounter, usually by an experienced senior colleague using a rating form for feedback (Nair et al. 2008). As our study did not include mini-CEX data we cannot make direct comparisons, but it is possible that the ECTV performs better than the mini-CEX at detecting underperforming trainees as the mini-CEX appears to only measure a single global dimension of competence (Cook et al. 2010) while the ECTV is more holistic, assesses a greater variety of skills and domains and educators are experienced with the underperformance benchmark (Kefala et al. 2005; Gladman 2011).

In our study, covariate analysis demonstrated the most significant demographic and academic risk factors for requirement for remediation were: overseas country of the trainees’ basic medical degree, and number of leave days in the first 6 months of training. We also found selection scores were *not* accurate predictors of need for remediation. These findings are consistent with limited previous research. Across all levels of medical education there are very few studies that have examined demographic and academic risk factors likely to lead to the need for remediation (Shah et al. 1997; Yao and Wright 2000; Winston et al. 2014). One subjective United States study in vocational training, reported that program directors felt that risk factors for underperformance included being overseas trained, being from an under-represented minority, and being aged over 35. Undergraduate performance, entrance/selection scores, and prior records (e.g. Board scores,

recommendation letters), however, were perceived as being only marginally helpful in screening for potential problem residents (Yao and Wright 2000). Selection processes in Australian general practice training program have not been found to be useful predictors of poor performance later in the program.

### Strengths and limitations

A strength of the study lies in our access to data concerning a range of potential predictors and covariates of all trainees of a large regional training program and in the extensive analyses which this data made possible.

There are, though, a number of limitations of our study. Being conducted in a single regional training organization means the results must be generalized with caution. The RTP in this study, however, encompasses a broad range of trainee and practice demographics and the assessment instruments evaluated in our study are identical or similar to those used by other the other 16 Australian regional training organizations. Thus the external validity of the study is likely to be acceptable.

The retrospective design using routinely collected data resulted in appreciable missing data for some variables. We were able, however, to address this issue by a sensitivity analyses using multiple imputation which showed the findings of our complete case analyses to be robust. The relatively small absolute number of trainees requiring remediation meant that we had statistical power in the primary complete case analysis only to model each predictor variable separately against our outcome (though we analyzed a full model using imputed data).

Finally, the independence of the outcome (remediation) from the predictor variables should be considered. Those charged with making the decision concerning need for remediation will have access to records of performance on the assessments that constitute our predictor variables. But these are only six predictors amongst myriad factors that influence this complex decision. Input from practice or RTP staff in the form of submission of 'Notification of Concern' forms is encouraged and the substance of these Notifications of Concern cover a wide range of issues. Much weight is placed on subjective or informal evidence—informal concerns reported by practice staff, especially supervisors, or by training organization staff (including office as well as education staff) not identified in other feedback modalities. These notifications carry weight and can trigger a consideration by the Remediation Committee. Included in 'Notification of Concern' from practice staff are those concerning adverse patient outcomes, near-miss critical events or other instances of poor clinical judgement. Critical events (Yao and Wright 2000) or a sequence of less serious events (Yao and Wright 2001) are common triggers for remediation in vocational training. But other concerns including issues of professionalism, communication and clinical knowledge can also prompt concern. Personal reports by trainees of difficulties in coping, of external stressors, or of psychological illness may also be factors in decision-making. The committee making decisions on remediation is governed by explicit policies. Judgements take into account triangulation of multiple data sources, as above, together with specific elicited input from the registrar, their supervisor, and their Medical Educator as part of the assessment process. Registrars requiring remediation typically struggle in multiple domains. We are confident that our outcome was largely independent of the predictor variables.

Informal assistance provided by a registrar's supervisor or Medical Educator as a result of a poor performance on one of our predictor assessments will be of far lesser intensity than formal remediation and is also unlikely to address shortcomings in other domains.

Thus, such informal action is unlikely to have attenuated associations in our analyses between need for remediation and those of our predictors which were found to be not significant.

Pragmatically, remediation is a very resource-intensive, expensive, and potentially fraught process. Identifying reliable predictors (even if not theoretically independent predictors) early in training will be of great utility (see below).

### Implications for educational policy and future research

We found that a tenth of trainees at a large Australian vocational training organization required formal remediation—a resource-intensive, expensive and often distressing process. We also found that three ‘in vivo’ assessment instruments (CFET, ECTV and DISQ) delivered early in training have reasonable capacity to predict later requirement for formal remediation (especially when combined with consideration of the country of the trainee’s first medical degree). These findings suggest that such ‘in vivo’ assessments should be a vital component of (necessarily complex) remediation prediction strategies. It is not surprising that this is so. We sought to frame remediation prediction within an NDM model where experienced people make decisions in a dynamic, uncertain clinical-educational environment. Instruments that capture this complexity in vivo would be expected to be integral to this assessment model. Additional exploration of the NDM model for high stakes clinical education decision making would be worthwhile.

Replication of our findings in other GP vocational training settings will be of importance and it would be of interest to see if our findings apply to other specialty training programs. Through early detection of under-performance it is plausible that earlier, less intense, interventions may avert the need for formal remediation. Further research should assess whether the institution of such earlier, simpler interventions targeted by consideration of performance on CFET, ECTV and DISQ (or equivalent instruments) does, indeed, avert the need for formal remediation. The cost-effectiveness of this process of targeted early intervention should also be evaluated.

**Acknowledgments** This project was funded by an Education Research Grant, 2015, Australian Government Department of Health.

### References

- Bahrami, J. (1997). Remediation. *British Medical Journal*, *315*, S2–S7114.
- Barrett, A., Galvin, R., et al. (2015). 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. *Systematic Reviews*, *4*, 65.
- Beaman, M., & King, J. (2009). Supporting professional performance: Managing trainees in difficulty South West Peninsula Deanery NHS Report.
- Bingham, C., & Crampton, R. (2011). A review of prevocational medical trainee assessment in New South Wales. *Medical Journal of Australia*, *195*, 410–412.
- Black, D., & Welch, J. (2009). The under-performing trainee-concerns and challenges for medical educators. *The Clinical Teacher*, *6*, 79–82.
- Borkett-Jones, H., & Morris, C. (2010). Managing the trainee in difficulty. *British Journal of Hospital Medicine*, *71*, 286–289.
- Byrne, P. (2012). Are they safe in there? Patient safety and trainees in the practice. *Australian Family Physician*, *41*, 26–29.
- Campbell, J., Narayanan, A., Burford, B., & Greco, M. (2010). Validation of a multi-source feedback tool for use in general practice. *Education for Primary Care*, *21*, 165–179.

- Cook, D. A., Beckman, T. J., Mandrekar, J. N., & Pankratz, V. S. (2010). Internal structure of mini-CEX scores for internal medicine residents: Factor analysis and generalizability. *Advances in Health Sciences Education, 15*, 633–645.
- Dijksterhuis, F., Scheele, F., Schuwirth, L., et al. (2009). Progress testing in postgraduate medical education. *Medical Teacher, 31*, e464–e468.
- Durning, S., et al. (2011). Viewing ‘strugglers’ through a different lens: How a self regulated learning perspective can help medical educators with assessment and remediation. *Academic Medicine, 86*, 1–8.
- Dwyer, A. J., Morely, P., Reid, E., & Angelatos, C. (2011). Distressed doctors: A hospital based support program for poorly performing and “at risk” junior medical staff. *Medical Journal of Australia, 194*, 466–469.
- Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51*, 380–417.
- Epstein, R. (2007). Assessment in medical education. *New England Journal of Medicine, 356*, 387–396.
- Evans, D., Alstead, E., & Brown, J. (2010). Applying your clinical skills to students and trainees in academic difficulty. *The Clinical Teacher, 7*, 230–235.
- Firth-Cozens, J., & King, J. (2006). Are psychological factors linked to performance? In J. Cox, J. King, A. Hutchinson, & P. McAvoy (Eds.), *Understanding doctors’ performance* (pp. 61–77). Oxford: Radcliffe Press.
- Garibaldi, R. A., Trontell, M. C., Waxaman, H., et al. (1994). The in-training examination in internal medicine. *Annals of Internal Medicine, 121*, 117–123.
- Gladman, G. (2011). The Identification of general practice registrar needing assistance. *The Australasian Medical Journal, 4*, 308–314.
- Hauer, K. E., Ciccone, A., Henzel, T. R., Katsufrakis, P., Miller, S. H., Norcross, W. A., et al. (2009). Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Academic Medicine, 84*, 1822–1832.
- Hays, R., & Peterson, L. (1996). An evaluation of visiting clinical teachers in general practice training. *Education for General Practice, 7*, 54–58.
- Hays, R., & Wellard, R. (1998). In-training assessment in postgraduate training for general practice. *Medical Education, 32*, 507–513.
- Ingham, G., Morgan, S., et al. (2015). Are GP supervisors confident they can assess registrar competence and safety, and what methods do they use? *Australian Family Physician, 44*, 154–158.
- Kefalas, B., Usherwood, T., & Knight, A. (2005). External clinical teaching visits in the regionalised environment. *Australian Family Physician, 34*(Suppl 1), 10–12.
- Kimatian, S. J., & Lloyd, S. H. (2008). Remediation and due process for trainees: Setting the stage for future success. *International Anesthesiology Clinics, 46*, 113–125.
- Larkins, S., Spillman, M., Vanlint, J., & Hays, R. (2003). Stress, personal and educational problems in vocational training. *Australian Family Physician, 32*, 473–476.
- Leap, L. L., & Fromson, J. A. (2006). Problem doctors: Is there a system-level solution? *Annals of Internal Medicine, 144*, 107–115.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*, S63–S67.
- Miller, A., & Archer, J. (2010). Impact of workplace based assessment on doctors’ education and performance: A systematic review. *British Medical Journal, 341*, c5064.
- Mitchell, C., Bhat, S., Herbert, A., & Baker, P. (2011). Workplace-based assessments of junior doctors: Do score predict training difficulties. *Medical Education, 45*, 1190–1198.
- Mugford, B. (2003). Performance and remediation in general practice training. *Australian Family Physician, 32*, 544–546.
- Nair, B. R., Alexander, H., McGrath, B. P., et al. (2008). The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Medical Journal of Australia, 189*(3), 159–161.
- Paice, E., & Orton, V. (2005). Managing the trainee doctor in difficulty. In A. Hastie, I. Hastie, & N. Jackson (Eds.), *Postgraduate medical education and training: A guide for primary and secondary care* (pp. 165–175). Oxford: Radcliffe.
- Papadakis, M. A., Arnold, G. K., Blank, L. L., Holmboe, E. S., & Lipner, R. S. (2008). Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Annals of Internal Medicine, 148*, 869–876.
- Reamy, B., & Harman, J. (2006). Residents in trouble: An in depth assessment of the 25 year experience of a single family medicine residency. *Family Medicine, 38*, 252–257.
- Riebschleger, M. P., & Haftel, H. M. (2013). Remediation in the context of the competencies: A survey of pediatrics residency program directors. *Journal of Graduate Medical Education, 5*, 60–63.

- Roberts, N. K., Williams, R. G., Klingensmith, M., et al. (2012). The case of the entitled resident: A composite case study of a resident performance problem syndrome with interdisciplinary commentary. *Medical Teacher, 34*, 1024–1032.
- Shah, S., Knowlden, F. R., Harris, M., & Yeo, G. (1997). General practice trainees in the borderline zone of competence: The experience of remediation at the regional general practice training and education centre. *Education for General Practice, 8*, 102–112.
- van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: Building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology, 24*, 703–719.
- Van der Vleuten, C., Schuwirth, L., et al. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher, 34*, 205–214.
- Violato, C., Lockyer, J., & Fidler, H. (2008). Changes in performance: A 5-year longitudinal study of participants in a multi-source feedback programme. *Medical Education, 42*, 1007–1013.
- Weizberg, M., Smith, J. L., Murano, T., Silverberg, M., & Santen, S. A. (2015). What does remediation and probation status mean? A survey of emergency medicine residency program directors. *Academic Emergency Medicine, 22*, 113–116.
- Williams, R. G., Roberts, N. K., Schwind, C. J., & Dunningham, G. L. (2009). The nature of general surgery resident performance problems. *Surgery, 145*, 651–658.
- Winston, K. A., Van der Vleuten, C. P., & Scherpbier, A. J. (2014). Prediction and prevention of failure: An early intervention to assist at risk medical students. *Medical Teacher, 36*, 25–31.
- Yao, D. C., & Wright, S. M. (2000). National survey of internal medicine residency program directors regarding problem residents. *JAMA, 284*, 1099.
- Yao, D. C., & Wright, S. M. (2001). The challenge of problem residents. *Journal of General Internal Medicine, 16*, 486–492.
- Zsombok, C. E., & Klein, G. (2014). Naturalistic Decision Making: Where are we now? In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (3rd ed., p. 19). New York: Psychology Press.