

The efficacy of medical student selection tools in Australia and New Zealand

Boaz Shulruf^{1,2}, Warwick Bagg³, Mathew Begun¹, Margaret Hay⁴, Irene Lichtwark⁴, Allison Turnock⁵, Emma Warnecke⁵, Timothy J Wilkinson⁶, Phillipa J Poole³

The known The predictive power of composite scores based on the weighted means of the results from multiple selection tools is low, explaining less than 15% of variance in student outcomes, regardless of the statistical models, selection tools, and outcomes used.

The new Discriminant function analysis yields optimal and meaningful cut-scores for each of the selection tools based on binary outcomes. A non-compensatory selection model or a “sufficient evidence” approach to selection would be useful for some medical schools.

The implications These alternative approaches may enhance the efficacy and defensibility of the selection process. They may also minimise the student non-completion rate in medical programs.

Entry to a medical program in Australia is highly competitive. Medical schools must allocate fairly the limited number of places available by identifying applicants with the potential to become successful doctors. To do so, they employ selection tools that include aptitude tests (eg, the Undergraduate Medicine and Health Sciences Admission Test [UMAT]), assessment of academic achievements (eg, final secondary school score or rank; university grade point average [GPA]), selection interviews (panel interviews, multiple mini-interviews), simulations and situational judgement tests, psychological tests, and random selection.¹⁻⁶ Most schools employ a combination of weighted or unweighted scores to prepare a list from which the most highly ranked applicants are selected.^{2,7,8}

The predictive power (efficacy) of selection tools can be estimated in multiple linear regression models. In general, the ability of selection tools to predict outcomes such as course completion (binary) or academic achievement throughout the medical program (including final grade; either continuous or categorical) is limited; the exception is prior academic achievement, which accounts for up to 23% of variance in academic outcomes.⁹ The predictive value of a battery of selection tools is generally calculated with a single equation, and the impact of each tool is measured while controlling variability in the other tools.^{1,10} Determining the efficacy of a combination of tools requires multiple regression models or multivariate analyses.^{2,11,12} For continuous outcomes, the impact of the selection tool score is measured against a unit of change in the outcome score without special treatment of the critical range around the pass–fail threshold. For example, if the pass mark is 50%, the difference between achieving 51% and 98% is significant, but the consequences for the student (and perhaps the university and health care system) are more serious if the student achieves 48% rather than 51%. As the pass rate in medicine usually exceeds 80%, the overall predictive power of a selection tool that poorly distinguishes between a low and a high

Abstract

Objectives: To estimate the efficacy of selection tools employed by medical schools for predicting the binary outcomes of completing or not completing medical training and passing or failing a key examination; to investigate the potential usefulness of selection algorithms that do not allow low scores on one tool to be compensated by higher scores on other tools.

Design, setting and participants: Data from four consecutive cohorts of students (3378 students, enrolled 2007–2010) in five undergraduate medical schools in Australia and New Zealand were analysed. Predictor variables were student scores on selection tools: prior academic achievement, Undergraduate Medicine and Health Sciences Admission Test (UMAT), and selection interview. Outcome variables were graduation from the program in a timely fashion, or passing the final clinical skills assessment at the first attempt.

Main outcome measures: Optimal selection cut-scores determined by discriminant function analysis for each selection tool at each school; efficacy of different selection algorithms for predicting student outcomes.

Results: For both outcomes, the cut-scores for prior academic achievement had the greatest predictive value, with medium to very large effect sizes (0.44–1.22) at all five schools. UMAT scores and selection interviews had smaller effect sizes (0.00–0.60). Meeting one or more cut-scores was associated with a significantly greater likelihood of timely graduation in some schools but not in others.

Conclusions: An optimal cut-score can be estimated for a selection tool used for predicting an important program outcome. A “sufficient evidence” selection algorithm, founded on a non-compensatory model, is feasible, and may be useful for some schools.

pass (despite accurately distinguishing between passing and failing) would appear to be low.¹³ Consequently, the conclusions about how best to apply selection tools that can be drawn from regression models and multivariate analyses are limited.

Another reason for the low predictive value of currently available tools is the difference between research models of student selection and how selection tools are actually applied.⁸ Medical schools rarely publish how the scores of selection tools are combined, but in research studies they are usually analysed as averages or cumulative scores. Further, few studies of combinations of selection tools have controlled for confounders, such as ethnic background or sex.

Building on earlier research,^{1,2,10,11,14-17} we proposed a different approach to evaluating the predictive value of selection tools. Specifically, our study focused on the efficacy of selection tools for predicting medical program success when the outcomes are binary (completing or not completing medical school; passing or

¹University of New South Wales, Sydney, NSW. ²Centre for Medical and Health Sciences Education, University of Auckland, Auckland, New Zealand. ³University of Auckland, Auckland, New Zealand. ⁴Monash University, Melbourne, VIC. ⁵University of Tasmania, Hobart, TAS. ⁶University of Otago, Christchurch, New Zealand.
 ✉ b.shulruf@unsw.edu.au • doi: 10.5694/mja17.00400 • See Editorial, p. 202 • Published online 12/03/2018

1 The study population, by medical school and sex

	Total	Women	Men
University of Auckland	703	367 (52%)	336 (48%)
Monash University	820	460 (56%)	360 (44%)
University of Otago	575	310 (54%)	265 (46%)
University of Tasmania	456	250 (55%)	206 (45%)
University of New South Wales	824	457 (56%)	367 (44%)
<i>Total</i>	<i>3378</i>	<i>1844 (55%)</i>	<i>1534 (45%)</i>

failing a key examination) rather than continuous (eg, test scores, final GPA) or ordinal (eg, fail, pass, pass with distinction). We proposed that, although whether a student is good or very good at medical school is important, it is more critical to distinguish between their passing and failing. Our investigation thus differed from earlier studies by focusing on outcomes from a consequential rather than a metric perspective.

Our second aim was to investigate whether or not our approach would support alternative methods for assessing the quality of selection algorithms that do not allow low scores on one tool to be compensated by high scores on other tools, as do, for example, averages or cumulative scores.

Methods

Context and data

Five undergraduate medical schools in Australia and New Zealand provided both selection and outcome data for 3378 students from four consecutive cohorts of students (enrolled 2007–2010) (Box 1). The predictor (independent) variables were student scores on the selection tools employed: a measure of prior academic achievement (GPA or Australian Tertiary Admission Rank [ATAR]), UMAT, and selection interviews (as applicable). The outcome (dependent) variables were graduation from the program in a timely fashion, and passing the final clinical skills assessment at the first attempt (Box 2). Timely graduation was defined as graduating no later than one year beyond the minimum study period. Some students complete their degree later for academic or personal reasons, but this is relatively rare; access to relevant data in this regard was not approved by the institutions' ethics committees for reasons of privacy. Methods for the final clinical skills assessment included objective structured clinical examinations and workplace-based assessments.

Statistical analysis

As we could not measure the probability of success of applicants not admitted to medical school, we employed a feasible model in which the probability of success in the main outcomes among the admitted applicants at each school was assessed optimistically; that is, an algorithm that maximises the estimated accuracy of the selection decision. We used discriminant function analysis (DFA)^{18,19} to identify the cut-score for each selection tool that best discriminated between students who achieved or failed to achieve an outcome. DFA predicts the classification of subjects into pre-defined categories of a dependent variable (outcome) according to one or more continuous or binary independent variables (predictors). DFA and logistic regression yield results that are almost identical when there are no nominal predictors and the

2 Variables included in the analyses reported in this article

Independent variables	Description
UMAT (mean score)	Mean score of all three UMAT components
UMAT1*	Logical reasoning and problem solving
UMAT2*	Understanding people
UMAT3*	Non-verbal reasoning
Grade point average or Australian Tertiary Admission Rank	Measure of prior academic achievement at secondary school (Australian Tertiary Admission Rank; Australian schools) or during first year of university or prior degree (grade point average; New Zealand medical schools)
Dependent variables	
Timely graduation	Graduation no later than a year beyond minimum time
Final clinical skills assessment	Pass/fail the final assessment in program
UMAT = Undergraduate Medicine and Health Sciences Admission Test. * Three components of the UMAT examination. ♦	

distribution of the continuous predictors is approximately bell-shaped;²⁰ these conditions applied to our study. DFA was preferred to logistic regression because its purpose is classification, whereas that of logistic regression is identifying the likelihood of observing a particular outcome following a one unit change in a continuous predictor.

For each outcome at each school, the cut-score was defined as the middle value between the two centroid points for each group yielded by DFA. To estimate statistical errors, expressed as a 95% confidence interval (CI), each DFA was calculated 100 times for random subsamples (50%) of the data.²¹ Confidence intervals for each cut-score could not be generated by bootstrapping (which employs 1000 random subsamples of different sizes) because bootstrapping provides 95% CIs for DFA coefficients but not for the cut-scores. Coe's guideline²² was used to estimate effect sizes from the percentage of classifications that were correct. An effect size under 0.4 is considered small, and one of 0.6 or more as large. The association between the number of cut-scores met and timely graduation was measured by Goodman and Kruskal's gamma.¹

All analyses were conducted in SPSS 22 (IBM).

Alternative selection algorithms

After establishing cut-scores, we examined whether alternative algorithms might be useful for selection purposes. In contrast to a compensatory algorithm — in which a high score on one tool can compensate for a low score on another — a non-compensatory algorithm requires that students reach thresholds on each of several tools. A third option would be to apply a rule of "sufficient evidence": that is, if the result for one selection tool indicates that an applicant is likely to succeed, they can be selected regardless of their scores on other tools.

Ethics approval

The research ethics committees of each institution endorsed the approval of this study by the Human Research Ethics Committee of the University of New South Wales (reference, HC15421, HREAP G: Health, Medical, Community and Social). To maintain the privacy of students and schools, we have not reported student

outcomes, nor have we described the interview or assessment methods; further, schools may have since altered their selection models. In this article, results are reported anonymously for schools A to E; the order is unrelated to that of the schools in Box 2.

Results

The efficacy of each tool in predicting each of the two outcomes is summarised in Box 3 and Box 4. The efficacy of each selection tool was independent of that of the other tools. For both timely graduation and final clinical skills assessment, prior academic achievement (GPA or ATAR) was the most effective selection tool, with medium to very large effect sizes for all schools. The predictive power of UMAT scores was limited, with two exceptions: a medium effect size of mean UMAT score in predicting timely graduation at School A, and of individual UMAT components for the same outcome at School E. The effect sizes for selection interviews predicting timely graduation or passing the final clinical skills assessment were small; at School A, the estimated effect size for predicting timely graduation was negative, meaning that the cut-score predicted failure more often than it predicted success.

Having determined the optimal cut-scores, we explored the utility of other selection algorithms. The feasibility of a “sufficient evidence” algorithm can be tested only if it includes two or more selection tools with medium or large effect sizes (0.4 or more²²), a condition that applied only to the outcome of timely graduation. The first tool selected for testing the “sufficient evidence” algorithm was the GPA/ATAR, as it consistently had the largest effect size; the second was the UMAT, because it was used by all schools.

The association between the number of cut-scores met (0, 1 or 2) and the likelihood of timely graduation is shown in Box 5. For School A, a second cut-score added little to the predictive power of the selection decision (increased from 95% to 97%), while only 81% of those who reached neither cut-score graduated in timely fashion. School A might thus benefit from using a “sufficient evidence” approach. Schools C and D might benefit from assessing only the GPA/ATAR cut-scores, as reaching the cut-score significantly increased the proportion who graduated on time compared with applicants who did not (from 78% to 88% and from 85% to 92% respectively). On the other hand, for Schools B and E, a cut-score-based selection algorithm would achieve little, as meeting even one cut-score did not significantly improve the timely graduation rate (Box 5).

Discussion

We have described a new approach to assessing the efficacy of medical student selection processes, after examining how well current selection tools predict two pragmatic binary outcomes: timely graduation and passing the final clinical skills assessment.

It has been reported that the predictive power of composite scores based on the weighted means of the results from multiple selection tools is low, explaining less than 15% of variance in student outcomes, regardless of the statistical models, selection tools, and outcomes used.^{1-4,9-11,13,14,16,23} We found similar effect sizes to previous studies, but the comprehensiveness and novel types of analyses are the major strengths of our study. We analysed large datasets from five medical schools that applied the student selection tools in different ways. As a result, our results may be more generalisable than those of studies that have focused on a single medical school, or on a particular tool in

3 Efficacy of selection tools in predicting timely graduation

Selection tool	School	Cut-score (95% CI)	Classification correct	Estimated effect size
GPA or ATAR	A	81.8 (81.1–82.2)	72.8%	1.20
	B	97.7 (97.6–97.7)	64.0%	0.67
	C	96.5 (96.5–96.5)	70.0%	1.02
	D	98.5 (98.4–98.5)	71.4%	1.11
	E	87.3 (87.1–87.4)	65.7%	0.76
Interview	A	75.0 (75.0–75.0)	48.4%	< 0
	B	86.4 (86.2–86.5)	58.3%	0.39
	C	NA	NA	NA
	D	82.8 (82.7–82.9)	54.6%	0.23
	E	NA	NA	NA
UMAT (mean)	A	54.7 (54.5–54.9)	59.9%	0.46
	B	62.0 (61.9–62.0)	45.0%	< 0
	C	57.3 (57.3–57.4)	52.5%	0.15
	D	60.7 (60.6–60.7)	52.3%	0.14
	E	73.4 (73.1–73.8)	57.4%	0.35
UMAT1	A	55.9 (55.7–56.2)	54.8%	0.24
	B	61.4 (61.3–61.5)	46.8%	< 0
	C	57.5 (57.4–57.6)	50.1%	0.07
	D	60.7 (60.5–60.8)	48.7%	< 0
	E	80.5 (80.0–81.1)	59.7%	0.45
UMAT2	A	53.2 (52.9–53.4)	52.7%	0.16
	B	59.6 (59.5–59.7)	47.0%	< 0
	C	56.1 (55.9–56.2)	48.4%	< 0
	D	58.0 (57.9–58.1)	48.3%	< 0
	E	68.9 (68.3–69.5)	62.8%	0.60
UMAT3	A	55.0 (54.8–55.3)	54.3%	0.22
	B	65.2 (65.0–65.4)	40.1%	< 0
	C	58.4 (58.3–58.5)	46.4%	< 0
	D	63.2 (63.0–63.4)	52.4%	0.15
	E	71.2 (70.6–71.8)	59.1%	0.42

ATAR = Australian Tertiary Admission Rank; CI = confidence interval; GPA = grade point average; NA = not applicable (data were not available); UMAT = Undergraduate Medicine and Health Sciences Admission Test. ♦

a particular jurisdiction.^{1,8,23} It is the broadest study of its type to date.

We employed techniques rarely used to investigate the efficacy of selection by medical schools. DFA is an effective tool for

4 Efficacy of selection tools in predicting a pass in the final clinical skills assessment

Selection tool	School	Cut-score (95% CI)	Classification correct	Estimated effect size
GPA or ATAR	A	NA	NA	NA
	B	97.3 (97.3–97.4)	72.7%	1.20
	C	96.5 (96.5–96.5)	70.0%	1.02
	D	98.6 (98.5–98.6)	73.2%	1.23
	E	88.1 (88.1–88.1)	59.4%	0.44
Interview	A	NA	NA	NA
	B	86.8 (86.7–87.0)	55.7%	0.27
	C	NA	NA	NA
	D	83.9 (83.2–84.6)	53.1%	0.17
	E	NA	NA	NA
UMAT (mean)	A	NA	NA	NA
	B	61.6 (61.5–61.7)	48.4%	< 0
	C	57.3 (57.2–57.3)	53.0%	0.17
	D	60.7 (60.5–60.9)	52.2%	0.14
	E	78.5 (78.5–78.5)	44.5%	< 0
UMAT1	A	NA	NA	NA
	B	61.5 (61.4–61.6)	45.2%	< 0
	C	57.6 (57.5–57.7)	49.7%	< 0
	D	60.6 (60.4–60.8)	48.2%	< 0
	E	85.0 (85.0–85.1)	52.5%	0.15
UMAT2	A	NA	NA	NA
	B	58.8 (58.7–59.0)	50.5%	0.08
	C	56.3 (56.2–56.4)	47.5%	< 0
	D	57.8 (57.7–57.9)	50.8%	0.09
	E	73.6 (73.4–73.7)	55.3%	0.26
UMAT3	A	NA	NA	NA
	B	64.3 (64.1–64.5)	46.9%	< 0
	C	58.1 (57.9–58.2)	48.1%	< 0
	D	63.8 (63.6–64.1)	52.2%	0.14
	E	76.8 (76.3–77.4)	47.5%	< 0

ATAR = Australian Tertiary Admission Rank; CI = confidence interval; GPA = grade point average; NA = not applicable (data were not available); UMAT = Undergraduate Medicine and Health Sciences Admission Test. ♦

facilitating decision making,^{18,24} and our approach of 100 random re-samplings is more conservative (ie, yields broader 95% CIs) than bootstrapping.²¹ We estimated optimal cut-scores for each selection tool and school for predicting binary outcomes. The cut-scores and the proportions of correct classifications may

5 Percentage of timely graduation, by number of cut-scores met

School	Gamma*	P	Number of cut-scores achieved		
			0	1	2
A	0.578	< 0.001	81%	95%	97%
B	0.031	0.79	86%	87%	NA
C	0.353	0.004	78%	88%	NA
D	0.314	0.014	85%	92%	NA
E	0.248	0.14	95%	96%	98%

NA = not applicable: effect size of tool was less than 0.4. * This statistic (range, 0–1) quantifies the association between timely graduation and the number of cut-scores achieved. ♦

be useful for evaluating current tools and policies and informing future selection decisions. This technique could also be applied to generating cut-scores that predict progression from earlier to later program phases.

The cut-scores for timely graduation and passing the final clinical skills assessment were similar for different selection tools employed by a school. This may not be surprising; the clinical skills assessment, required for graduation, is undertaken late in the medical program. We selected these two outcomes because they are understandable and were available for all schools. The outcome of subsequent workplace performance, while important, is moderated by influences beyond the undergraduate environment, including experience as interns, postgraduate training, and personal circumstances. Timely graduation and passing the final clinical skills assessment might, however, be considered surrogate measures of subsequent clinical practice, as medical schools are guided by graduate outcomes when designing summative clinical skills assessments and making decisions about progression, including about whether a student should graduate.

If all selection tools are regarded as being important for selection, our findings suggest that schools might consider non-compensatory algorithms, as students should not be selected if they perform poorly on a selection tool that reliably predicts outcomes.^{7,17} The “sufficient evidence” approach could be implemented in two ways. If the selection tools each measure a different, independent set of attributes (ie, correlations between scores on different tools are small), an applicant achieving a higher number of cut-scores has provided more evidence that they are suitable for medical training across a broader range of attributes than one who has achieved fewer, and should be ranked higher for selection. Alternatively, if a particular tool reliably predicts an outcome, and the other tools add little to predictive accuracy, achieving the cut-score on the reliable tool would suffice for selection (a “sufficient evidence” approach), while the results from multiple tools could be combined for applicants who did not achieve the cut-score on the reliable tool.

Our investigation confirms the findings of local studies that prior academic performance is the most effective tool for predicting academic outcomes, and that neither UMAT nor interview scores have a consistent impact.^{1,2,10,16,17} Most studies have assessed continuous outcomes, and admission GPA/ATAR explained no more than 10% of variance.^{1,2,10,16,17} Only one study examined the effectiveness of selection tools in predicting success *v* failure at the end of the program as a binary outcome.¹ This binary approach may be increasingly important as schools seek to further diversify

their student intake to better meet health and community needs. Schools may need to focus more on improving the accuracy and validity of selection tools close to the selection threshold mark rather than at the upper end of the scale.²⁵

One limitation of our study is that we could not measure outcomes for unsuccessful applicants. Another is that cut-scores were based on outcomes for earlier student cohorts, and we assumed that these applicants were similar to the current admission cohort. Further, the failure rates for our two outcomes were low, as is typical for a study examining success and failure in medical programs. When one of the dependent variable categories includes only a small proportion of the study population, the probability of a type I error increases. Nonetheless, we analysed data from four consecutive cohorts at each of five medical schools in Australia and New Zealand, and the analysis included 100 random re-samplings, achieving narrow 95% CIs for the cut-scores. This suggests that our assumption of stability across cohorts is valid, and supports the generalisability of our findings. Finally, the low predictive value of a selection tool such as interview may indicate an inability to effectively select for outcome success, but the tool may still have a

high level of effectiveness in “selecting out” unsuitable applicants whose data are not included in the study.

In conclusion, there is no gold standard for selecting medical students. Schools must choose fairly the applicants they believe have the aptitude to meet the community’s health care needs. Completing the medical training program is the first step. Medical schools periodically evaluate their selection tools and algorithms according to accumulated evidence derived from institutional data. Our study suggests that new methods may be useful for guiding selection policy and processes, including using meaningful cut-scores for selection based on important program outcomes, and applying non-compensatory selection algorithms or a “sufficient evidence” algorithm.

Acknowledgements: This study was supported by a grant from the UMAT Consortium and the Australian Council for Educational Research.

Competing interests: No relevant disclosures.

Received 3 May 2017, accepted 27 Nov 2017. ■

© 2018 AMPCo Pty Ltd. Produced with Elsevier B.V. All rights reserved.

- 1 Shulruf B, Poole P, Wang YG, et al. How well do selection tools predict performance later in a medical programme? *Adv Health Sci Educ* 2012; 17: 615-626.
- 2 Poole P, Shulruf B, Rudland J, Wilkinson T. Comparison of UMAT and admission GPA on the prediction of performance on medical school assessments: a national, cross-institution Study. *Med Educ* 2012; 46: 163-171.
- 3 Mercer A, Puddey I. Admission selection criteria as predictors of outcomes in an undergraduate medical course: A prospective study. *Med Teach* 2011; 33: 997-1004.
- 4 Urlings-Strop LC. Selection of medical students: a controlled experiment. *Med Educ* 2009; 43: 175-183.
- 5 Sharma S, Gangopadhyay M, Austin E, Mandal MK. Development and validation of a situational judgment test of emotional intelligence. *Int J Sel Assess* 2013; 21: 57-73.
- 6 Adam J, Bore M, McKendree J, et al. Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Med Educ* 2012; 12: 69.
- 7 Ma C, Harris P, Cole A, et al. Selection into medicine using interviews and other measures: much remains to be learned. *Iss Educ Res* 2016; 26: 623-634.
- 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: 215-223.
- 9 Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002; 324: 952-957.
- 10 Edwards D, Friedman T, Pearce J. Same admissions tools, different outcomes: a critical perspective on predictive validity in three undergraduate medical schools. *BMC Med Educ* 2013; 13: 173.
- 11 Puddey I, Mercer A. Socio-economic predictors of performance in the Undergraduate Medicine and Health Sciences Admission Test (UMAT). *BMC Med Educ* 2013; 13: 155.
- 12 Quinlivan J, Lam L, Wan S, Petersen R. Selecting medical students for academic and attitudinal outcomes in a Catholic medical school. *Med J Aust* 2010; 193: 347. <https://www.mja.com.au/journal/2010/193/6/selecting-medical-students-academic-and-attitudinal-outcomes-catholic-medical>
- 13 Arulampalam W, Naylor R, Smith J. A hazard model of the probability of medical school drop-out in the UK. *J R Stat Soc Series A Stat Soc* 2004; 167: 157-178.
- 14 Poole P, Shulruf B. Shaping the future medical workforce: take care with selection Tools. *J Prim Health Care* 2013; 5: 269-275.
- 15 Poole P, Shulruf B, Harley B, et al. Shedding light on the decision to retain an interview for medical student selection. *N Z Med J* 2012; 125: 81-88.
- 16 Mercer A, Abbott P, Puddey I. Relationship of selection criteria to subsequent academic performance in an Australian undergraduate dental school. *Eur J Dent Educ* 2013; 17: 39-45.
- 17 Wilkinson D, Zhang J, Parker M. Predictive validity of the Undergraduate Medicine and Health Sciences Admission Test for medical students' academic performance. *Med J Aust* 2011; 194: 341-344. <https://www.mja.com.au/journal/2011/194/7/predictive-validity-undergraduate-medicine-and-health-sciences-admission-test>
- 18 Garson GD. Discriminant function analysis. Asheboro (NC): Statistical Associates Publishing, 2012. <http://www.statisticalassociates.com/discriminantfunctionanalysis.htm> (viewed Nov 2017).
- 19 Haas RE, Nugent KE, Rule RA. The use of discriminant function analysis to predict student success on the NCLEX-RN. *J Nurs Educ* 2004; 43: 440-446.
- 20 Cupples L, Heeren T, Schatzkin A, Colton T. Multiple testing of hypotheses in comparing two groups. *Ann Intern Med* 1984; 100: 122-129.
- 21 Wood M. Statistical inference using bootstrap confidence intervals. *Signif (Oxf)* 2004; 1: 180-182.
- 22 Coe R. It's the effect size, stupid: what effect size is and why it is important. British Educational Research Association Annual Conference, Exeter, United Kingdom, 12–14 September, 2002. <http://www.leeds.ac.uk/educol/documents/00002182.htm> (viewed Nov 2017).
- 23 Sladek R, Bond M, Frost L, Prior K. Predicting success in medical school: a longitudinal study of common Australian student selection tools. *BMC Med Educ* 2016; 16: 187.
- 24 Kreiter C, Solow C. A statistical technique for the development of an alternate list when using constrained optimization to make admission decisions. *Teach Learn Med* 2002; 14: 29-33.
- 25 Hay M, Mercer A, Lichtwark I, et al. Selecting for a sustainable workforce to meet the future healthcare needs of rural communities in Australia. *Adv Health Sci Educ Theory Pract* 2017; 22: 533-551. ■