Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school

Barbara Griffin, David W Harding, Ian G Wilson and Neville D Yeomans

As part of the student selection process for entry into medicine, Australian universities are increasingly using interviews and specialist tests of cognitive and non-cognitive ability, such as the Undergraduate Medicine and Health Sciences Admission Test (UMAT) and, more recently, the Multiple Mini Interview (MMI).[1] The use of such tests is an attempt to overcome the socioeconomic bias thought to be associated with high-school matriculation results.[2,3]

However, selection of medical students is a high-stakes context and the competition for places has created a market for independent businesses that offer expensive coaching programs claiming to improve both ability test scores and interview performance. For example, for a fee of about $1700, one company in Australia is currently advertising a weekend workshop that includes additional material for "100s of hours of skill development". The extent to which applicants attend such programs and the effect of this type of coaching on actual scores has rarely been evaluated.[4]

In addition to the availability of coaching programs, there is evidence that a substantial number of students repeat the cognitive ability tests and interviews in an attempt to improve on their initial results.[5] About 9% of the candidates who sat for the UMAT exam in 2007 were resitting the exam for at least the second time (UMAT Test Management Committee, personal communication).

Any gain from coaching or retesting is known as a "practice effect". There is a large body of research on practice effects in relation to cognitive ability tests in general, but little of this relates specifically to medical entry tests. Much less has been reported on the effect of coaching or repeat attempts at selection interviews.

Although there are no studies on the effect of coaching on the UMAT, a qualitative review of the impact of coaching on the Medical College Admission Test (used mainly in the United States and Canada)[4] showed that it had only a minimal effect in increasing scores.

In a study examining test security breaches, Reiter et al[6] found that MMI performance was not affected by prior access to specific interview questions. This suggests that candidates may not benefit from time to rehearse and seek advice on response content. Nevertheless, having access to questions may be less useful than participating in an actual MMI.

Our study aimed to assess the practice effects from coaching on the UMAT, and the effect of both coaching and repeat testing on the MMI,[7] which is being adopted by an increasing number of medical schools.

METHODS

Of approximately 2300 applicants for entry in 2008 to the new School of Medicine at the University of Western Sydney, 340 were selected for interview (ie, the MMI), based on a threshold Universities Admission Index (UAI) of 93 out of 100 or university Grade Point Average of 5.5 out of 7 and a ranking on the basis of the total UMAT score.

The UMAT

The UMAT is a multiple choice test consisting of three sections, each measuring different aspects of ability: logical reasoning (Part 1), understanding people (Part 2) and non-verbal reasoning (Part 3). The results on all three sections are averaged to obtain the total UMAT score. To fulfil the School's mandate to increase the number of medical practitioners in the local area (which has lower average socioeconomic status than some other parts of Sydney), there was a

ABSTRACT

Objective: To assess the practice effects from coaching on the Undergraduate Medicine and Health Sciences Admission Test (UMAT), and the effect of both coaching and repeat testing on the Multiple Mini Interview (MMI).

Design, setting and participants: Observational study based on a self-report survey of a cohort of 287 applicants for entry in 2008 to the new School of Medicine at the University of Western Sydney. Participants were asked about whether they had attended UMAT coaching or previous medical school interviews, and about their perceptions of the relative value of UMAT coaching, attending other interviews or having a "practice run" with an MMI question. UMAT and MMI results for participants were compared with respect to earlier attempts at the test, the degree of similarity between questions from one year to the next, and prior coaching.

Main outcome measures: Effect of coaching on UMAT and MMI scores; effect of repeat testing on MMI scores; candidates' perceptions of the usefulness of coaching, previous interview experience and a practice run on the MMI.

Results: 51.4% of interviewees had attended coaching. Coached candidates had slightly higher UMAT scores on one of three sections of the test (non-verbal reasoning), but this difference was not significant after controlling for Universities Admission Index, sex and age. Coaching was ineffective in improving MMI scores, with coached candidates actually having a significantly lower score on one of the nine interview tasks ("stations"). Candidates who repeated the MMI in 2007 (having been unsuccessful at their 2006 entry attempt) did not improve their score on stations that had new content, but showed a small increase in scores on stations that were either the same as or similar to previous stations.

Conclusion: A substantial number of Australian medical school applicants attend coaching before undertaking entry selection tests, but our study shows that coaching does not assist and may even hinder their performance on an MMI. Nevertheless, as practice on similar MMI tasks does improve scores, tasks should be rotated each year. Further research is required on the predictive validity of the UMAT, given that coaching appeared to have a small positive effect on the non-verbal reasoning component of the test.

MJA 2008; 189: 270–273

<table>
Comparison of UMAT and MMI results for coached and non-coached groups					
	Mean score (SE)				
Group	UMAT 1*	UMAT 2*	UMAT 3*	Total MMI	MMI Station 2*
Coached	63.61 (0.63)	57.25 (0.63)	63.65 (0.61)	3.54 (0.04)	3.81 (0.07)
Non-coached	61.93 (0.66)	58.55 (0.63)	61.33 (0.65)	3.56 (0.04)	4.01 (0.07)
P	0.065	0.147	0.010	0.724	0.044
</table>

SE = standard error. MMI = Multiple Mini Interview. UMAT = Undergraduate Medicine and Health Sciences Admission Test. *UMAT Part 1, logical reasoning; UMAT Part 2, understanding people; UMAT Part 3, non-verbal reasoning; MMI Station 2, understanding communication. ◆

lower cut-off UMAT score for applicants residing in the region so that about half the interview places could be allocated to local applicants.

The MMI

The MMI consists of a series of short, structured interviews ("stations") used to assess personal qualities. The 2007 MMI (for 2008 entry to the course) consisted of nine stations, each measuring the same underlying construct as the 2006 MMI. However, only two stations had questions identical to those in the 2006 MMI; four were alternative forms, and three were completely new. Two related ratings were given by the interviewers at each station using 5-point scales, with anchors that described examples of poor (score 1), satisfactory (score 3) and outstanding (score 5) performance. For example, at the station assessing communication skills, candidates were rated on their ability to explain information clearly (1st rating) and in a sensitive manner (2nd rating). An average of the two scores for each station and an average score for all nine stations were calculated.

Interviewers included medical practitioners, university staff and community representatives, all of whom undertook an intensive 3-hour training session on interviewing techniques and using the rating scales, with a further half-hour briefing immediately before conducting the MMI.

We investigated the effect of retesting on MMI scores, comparing performances in 2007 on stations that were identical (in terms of questions asked) to those of the 2006 MMI with performance on stations that were alternative forms of the 2006 version and with those that were completely new.

Research survey

After attending the MMI, candidates were invited to complete a self-report survey. They were asked whether they had attended UMAT coaching, how many other medical school interviews they had attended so far, and how well they thought UMAT training, attending interviews at other universities and having a "practice run" with different MMI questions would help them do their best at an MMI. Statements in bold print on the consent form, the information form and the actual survey form assured candidates that their responses would have no influence at all on the selection process.

Statistical analysis

Independent *t* tests using SPSS, version 15.0 (SPSS Inc, Chicago, Ill, USA) were used to assess differences between coached and non-coached groups. For candidates who did the MMI in 2006 and repeated it in 2007, *z* scores (showing the position of each student in relation to the mean and standard deviation of the distribution of all scores) for the total MMI and for each station for both 2006 and 2007 were calculated. Paired *t* tests were used to compare their relative position in the 2006 cohort with their position in the 2007 cohort.

Ethical approval

The Human Research Ethics Committee of the University of Western Sydney approved our study.

RESULTS

Results are based on 287 candidates (84% of all students selected for interview) who agreed to participate in our study. Their mean age was 18.34 years (SD, 2.64 years) and 51.3% were males.

Prevalence of coaching

Just over half of the 287 respondents (51.4%) indicated they had attended coaching. In comparison with those who had not attended coaching, the coached group had a significantly higher UAI (98.13 v 97.47; $t = 2.18$; $P = 0.03$), were slightly younger (mean age, 17.92 v 18.76 years; $t = 2.58$; $P = 0.01$), and were more likely to be male (57.6% v 44.4%; $\chi^2 = 4.88$; $P = 0.03$). Although a smaller proportion of local-area candidates (49.4%) than non-local candidates (53.7%) had attended coaching, the difference was not statistically significant.

Interview experience

The majority of respondents (203 [70.7%]) had attended at least one other medical school interview before attending the MMI, but these would have been almost exclusively panel interviews, as our university is the only one in Australia that conducts MMIs for undergraduate admission, and only eight of the graduate interviewees reported having attended interviews at universities offering postgraduate courses. Seventeen candidates who completed our MMI in 2006 and were unsuccessful had reapplied and were again selected for interview in 2007.

Candidate perceptions

All respondents thought that having a "practice run" with an MMI question would be the most effective way of helping them do their best (compared with coaching and other interview experience), but those who had attended the 2006 MMI rated practice even more favourably than those who had not (mean rating on five-point scale, 3.60 v 3.00; $t = 2.41$; $P = 0.017$). Coaching was considered the least helpful tool by all respondents, although those who had attended coaching rated it more helpful than those who had not (mean rating on five-point scale, 2.32 v 1.98; $t = 3.34$; $P = 0.001$). Those who had attended interviews at other universities did not rate their experience more or less useful than those who had not attended prior interviews.

Effect of coaching on UMAT scores

There was no significant difference between the coached and non-coached groups on their UMAT scores for the tests of logical reasoning or understanding people, but those who were coached had higher scores on the test of non-verbal reasoning ($P = 0.01$) (Box).

However, after hierarchical regression analyses controlling for UAI, sex and age, the difference between coached and non-coached students on Part 3 of the UMAT was no longer significant.

Effect of coaching on MMI scores

Coaching made no difference to the total MMI score (Box), even after controlling for UAI, sex and age. However, on one of the nine stations (Station 2, which assessed communication skills), the coached group had significantly lower scores than the non-coached group ($P = 0.044$).

Effect of interview experience on MMI scores

There were no significant differences on the total MMI score or on any of the individual station scores between those who had attended interviews at other universities and those for whom the MMI was their first interview.

Effect of repeating the MMI

Seventeen candidates who did the MMI in 2006 repeated it in 2007. There was a significant increase in the total interview z score from the first to second interview attempt (from -0.72 to 0.00; $t = 4.14$; $P = 0.001$), despite three of the 17 candidates performing worse on the second attempt. There was a small but significant increase in four of the nine individual station scores. A comparison of raw scores showed similar but mostly non-significant improvements.

A comparison of the average improvement in z scores between 2006 and 2007 showed that there was a significant improvement on stations whose content was exactly the same as the 2006 stations (mean increase, 0.53 SD; $P = 0.021$) and on stations that were alternative forms of the 2006 stations (mean increase, 0.57 SD; $P = 0.005$). However, there was no improvement in performance on stations that were new in 2007.

DISCUSSION

Entry into medical school is extremely competitive, and it appears that coaching for the selection tests is widespread in Australia, as in other parts of the world.[4] Most coaching programs include training for both the UMAT and interviews, but our results suggest that such training is ineffective for improving MMI scores, and in fact may even be associated with reduced scores on some stations. However, at least one part of the UMAT, the test of non-verbal reasoning, may be susceptible to improvement with coaching.

Practice effects appear to differ among dimensions of cognitive ability, with quantitative and analytical tests more easily solved by learning specific problem-solving skills, whereas verbal tests that tap general information and acquisition of new knowledge are less amenable to improvement with retesting or coaching.[5] Because the non-verbal reasoning section of the UMAT requires candidates to solve pattern series using quantitative and specific skills, it is not surprising that this section was more affected by coaching than either the logical reasoning section (which requires acquisition of new knowledge to solve the items) or the understanding people section (which requires general understanding of interpersonal relationships and functioning).

The presence of a practice effect is important, as it may change the construct and predictive validity of a test, affecting the fairness to all applicants if differential outcomes occur.[7] A recent meta-analysis of practice effects on cognitive ability tests used in general selection contexts[5] found that test scores increased by about 0.25 SD from the first to second administration. The effect was larger when coaching was delivered between tests (although repetition was a potential confounder). In the context of medical student selection, Lievens et al[8] showed that retesting on the Flemish Medical and Dental Studies admission exam actually altered the construct validity of the test, in that it became less "g-loaded". In other words, the test measured what it was designed to measure (general cognitive ability [g]) on the first administration, but after retesting, results reflected proportionally more variance due to non-g factors such as narrow test-specific skills and test-wiseness. This is a particularly important issue, as the generalisability of test results resides primarily in g.[8] On the basis of their findings, Lievens et al questioned the practice of allowing candidates to repeat such tests.

A search of the literature failed to identify any other studies on the effects of retesting on interview scores. In terms of coaching, two of three published studies found a positive effect on interview performance,[9,10] whereas the third[11] showed no effect.

Strengths of our study were the high response rate and the reasonably large sample size. Another strength was the opportunity to examine the effect of coaching on the results of testing multiple personal characteristics rather than just one global rating. On the other hand, a possible problem with comparing coached and non-coached groups is that the groups are unlikely to be equivalent, as the use of coaching is voluntary and may be linked with factors such as personality, ability and socioeconomic status.[7]

Furthermore, although the effect of coaching on the non-verbal reasoning section of the UMAT became insignificant after controlling for age, sex and UAI, it should be remembered that the participants in our survey group were among the top UMAT performers and there is no way of knowing whether a higher proportion of this group underwent coaching than of the group of lower performers who did not reach our threshold for invitation to interview.

In contrast, a full range of typical interviewees was examined, so the results in relation to the lack of coaching effects on the MMI are likely to be more reliable. Although coaching could prepare candidates with examples of "good" responses, interviewers may actually have given lower scores to those whose responses appeared "rehearsed" or lacking a genuine quality. The MMI investigated in our study emphasised the use of behavioural interviewing (questions asked candidates to describe examples of their past behaviour[11]) as a further guard against possible score elevation due to coaching. In addition, training in "impression management" technique is unlikely to be of benefit in the MMI style of interview, as candidates attend each station for only a brief period.

The MMI was nevertheless susceptible to retesting effects, and interview candidates themselves believed that a practice run on an MMI station would be the most effective way of helping them do their best. The MMI score is a significant component of the final selection ranking, and seven of the 17 repeat candidates performed sufficiently well (relative to the 2007 interviewee cohort) to gain a place in the 2008 student intake. Given that there was no practice effect when station content was new, medical schools may need to consider revising at least some of their MMI content from year to year.

It is possible that the retest effects on the MMI that we observed were the result of regression towards the mean. As there were 100 places available for the 340 candidates interviewed, the chance of selection was almost one in three. However, given that a number of those interviewed eventually accepted positions at other universities, offers were made beyond the top 100 ranked applicants, which further increased the chance of selection. In such situations, those returning for retesting are all among the lowest ranked at the first test, and improvement suggests regression towards the mean.[5]

In contrast, only 15% of the more than 2000 eligible applicants were offered an interview on the basis of their UMAT score. Any changes in UMAT scores as a result of retesting are therefore unlikely to be due to regression towards the mean, because the retest candidate pool will probably have an average initial score almost equivalent to that of the entire group who completed the first test.[5]

The important issue yet to be resolved is whether practice effects on either the UMAT or the MMI change their predictive ability. In relation to coaching, a candidate's first score may be the better predictor of success in medical school or beyond because gains in performance on the test "reflect construct-relevant improvements that do not extend into the criterion domain".[5]

Further research is required on the effect of coaching and repetition on the predictive validity of the UMAT and MMI so that medical schools can critically evaluate the practical implications of their use. The reassuring message from our study, however, is that practice effects appear to be small or sometimes even negative.

AUTHOR DETAILS

Barbara Griffin, BPsych(Hons), PhD, MAPS, Senior Research Fellow[1]
David W Harding, BA(Hons), MClinPsych, MAPS, PhD Candidate[2]
Ian G Wilson, MB BS, PhD, FRACGP, Professor of Medical Education[2]
Neville D Yeomans, MD, FRACP, AGAF, Foundation Dean[2]
1 College of Health and Science, University of Western Sydney, Sydney, NSW.
2 School of Medicine, University of Western Sydney, Sydney, NSW.
Correspondence: b.griffin@uws.edu.au

REFERENCES

1 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004; 38: 314–326.
2 Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002; 324: 952-957.
3 Story M, Mercer A. Selection of medical students: an Australian perspective. *Intern Med J* 2005; 35: 647-649.
4 McGaghie WC, Downing SM, Kubilius R. What is the impact of commercial test preparation courses on medical examination performance? *Teach Learn Med* 2004; 16: 202-211.
5 Hausknecht JP, Halpert JA, Di Paolo NT, Moriarty Gerrard MO. Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *J Appl Psychol* 2007; 92: 373-385.
6 Reiter HI, Salvatori P, Rosenfeld J, et al. The effect of defined violations of test security on admissions outcomes using multiple mini-interviews. *Med Educ* 2006; 40: 36-42.
7 Ryan AM, Ployhart RE, Greguras GJ, Schmit MJ. Test preparation programs in selection contexts: self-selection and program effectiveness. *Personnel Psychol* 1998; 51: 599-621.
8 Lievens F, Reeve CL, Heggestad ED. An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *J Appl Psychol* 2007; 92: 1672-1682.
9 Maurer T, Solamon J, Troxtel D. Relationship of coaching with performance in situational employment interviews. *J Appl Psychol* 1998; 83: 128-136.
10 Maurer TJ, Solamon JM, Andrews KD, Troxtel DD. Interviewee coaching, preparation strategies, and response strategies in relation to performance in situational employment interviews: an extension of Maurer, Solamon, and Troxtel (1998). *J Appl Psychol* 2001; 86: 709-717.
11 Campion MA, Campion JE, Hudson JP Jr. Structured interviewing: a note on incremental validity and alternative question types. *J Appl Psychol* 1994; 79: 998-1002.