

Validity Evidence of a Multiple Mini Interview for Selection of Medical Students: Universiti Sains Malaysia Experience

Ahmad Fuad Abdul Rahim, Muhamad Saiful Bahri Yusoff

Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, Malaysia

To cite this article: Ahmad Fuad Abdul Rahim, Muhamad Saiful Bahri Yusoff. Validity evidence of a multiple mini interview for selection of medical students: Universiti Sains Malaysia experience. *Education in Medicine Journal*. 2016;8(2):49-63. DOI: 10.5959/eimj.v8i2.437

To link to this article: <http://dx.doi.org/10.5959/eimj.v8i2.437>

ABSTRACT

Introduction: Student selection is important for selecting the best candidates into medical courses. The Multiple Mini Interview (MMI) is a valid selection tool for the task. The School of Medical Sciences (SMS), Universiti Sains Malaysia (USM) has employed MMI as a medical admission tool to select potential candidates into its medical program. **Objective:** This paper described the implementation of the MMI and reported the preliminary evaluation data on its validity evidence. **Method:** A 9-station MMI (with 5 manned and 4 rest stations) was employed for the 2015 student selection exercise. Interview data were analysed and questionnaire surveys were administered to both interviewers and candidates. Validity, reliability, feasibility and acceptability were determined. **Results:** Unidimensional construct on confirmatory factor analysis (CFA) and interviewer comments provided evidence of construct validity. Overall reliability was 0.94, good enough for high-stakes decisions. Interviewers and candidates' comments on feasibility and acceptability were also generally positive. The consistency of the difficulty and discrimination indices of similar stations between sessions was identified as among the areas for improvement. **Conclusions:** We are cautiously optimistic regarding the utility of the MMI in the SMS; future improvements are planned but the present implementation seems sufficient for the stated purposes.

Keywords: *Student selection, Multiple mini interview, MMI, Validity study, Validity evidence, Universiti Sains Malaysia, USM*

CORRESPONDING AUTHOR

Dr Ahmad Fuad Abdul Rahim, Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan | Email: fuad@usm.my

Introduction

The Medical Degree Program in the School of Medical Sciences (SMS), Universiti Sains Malaysia (USM) is a 5-year program. After USM received the Accelerated Program for Excellence (APEX) status from the Malaysian Higher Education Ministry in 2008, the SMS was allowed to conduct its own student-selection program beginning with the 2009/2010 session intake (1, 2).

Realising the importance of non-academic attributes for future professional practice (3, 4) and the fact that candidates called for the selection were already screened in terms of academic competency, the SMS decided to look at non-academic attributes. Several years' worth of evaluation data (2) helped us to focus on four domains: reasons for applying, emotional-quotient-focused observations, personality attributes and

language proficiency and communication skills.

The initial selection method was the personal interview using a two-person panel of interviewers for each candidate; each session lasting for 15 to 20 minutes (2). Although the personal interview is a popular method of student selection in medical education institutions within and outside the country, there are still inherent weaknesses in the format (4, 5) including low generalisability (i.e. the limited degree to which a candidate's performance in the interview can be generalised to his or her general competency), and interviewer bias. This was also evident in our evaluation data (2). The SMS, after scrutinising the literature for the latest trends in student selection, therefore decided to use the MMI beginning with the student selection for academic year 2014/2015. Based on evaluation reports (e.g. (5), (6)) the SMS decided to forgo pilot testing and proceeded immediately with the MMI.

Since the inception of MMI in 2004 (7), many studies have provided convincing evidence to support its validity in student selection. The first evidence was related to its content that measures the non-cognitive performance of potential candidates, for example critical thinking, ethical decision making, communication skills and knowledge of healthcare system (7). The second evidence was related to its response process that was perceived as high acceptability by both candidates and examiners (8), its scores were not affected by level of exposure to MMI training or coaching, and thus least biases due to irrelevant constructs (9). The third evidence was related to its internal structure, in which researches consistently showed a high level of internal consistency indicating high reliability (8). The fourth evidence was related to its relations to other variables, for examples MMI had the strongest correlation with clerkship performance compared to personal interview, grade point average, autobiographical submission and simulated tutorial (10). The last evidence was related with its consequences on professional

behaviours, for example candidates who were selected by MMI showed the least percentage of individual flagged for unprofessional behaviours compared to personal interview, grade point average, autobiographical submission and simulated tutorial (10, 11). Despite the favourable validity evidence, it is very resource intensive (7, 9–11).

This paper firstly described the implementation of the MMI in the SMS and secondly reported the preliminary evaluation data.

Method

Description of the MMI in the SMS

The MMI was first used by the School of Medicine in the University of McMaster, Canada in 2004 (5). It borrows the concepts underlying the Objective Structured Clinical Examination (OSCE), where multiple samplings of students' clinical competencies are taken to overcome content specificity and provide more reliable measurement of their overall competency. In the MMI, students are interviewed for a shorter time compared to the traditional personal interview (hence the name "mini-interview") but in multiple stations (hence "multiple"). Since its inception in 2004, evaluation studies have shown good psychometric qualities for this method (11–17) and it has found a place in many medical educational institutions abroad as their method of choice for student selection.

The SMS, however, had to make several adjustments and modifications in the implementation of the MMI, as follows:

Number of Stations and Timing

The main considerations for the SMS regarding the number of stations and timing are the number of candidates expected and the number of interviewers available. Approximately 500 candidates are called for interview each year, based on the consideration of interviewing up to

four times the expected intake. It is worth noting that the actual number of eligible applicants is far higher and the final number of potential candidates called for interview is only obtained after screening using strict academic criteria.

Apart from the main interview centre in Kubang Kerian, Kelantan, the SMS also has two interview centres in Sabah and Sarawak; East Malaysian states more than 1500 km away. Although the candidate numbers in these centres are much less, transporting more interviewers (as is inherent in the MMI method) to the centres incurs more cost.

Although OSCE studies have indicated that to achieve acceptable reliability, 8–12 stations are required (18), the SMS, having considered the factors above, decided to begin with a 9-station MMI with 5 manned stations and 4 rest stations. Each station lasts for 7 minutes (5 minutes + 2 minutes writing for interviewers/preparation for candidates). It is comforting to note that the number of stations and time per station, even as an initial project, was within the range of 4–12 stations and 5–15 minutes as reported in a systematic review of MMI practices (19). To cater the large candidate

number, six identical circuits were run concurrently over two days; each session running approximately for an hour with about six sessions per day.

Station Content

The initial decision of the SMS to look at non-cognitive attributes is continued in our MMI, with each station focusing on different aspects or domains. As this was the first time the SMS implemented the MMI, it was pragmatic to include the content domains used by McMaster university, the pioneers in this approach (5): critical thinking, communication skills, ethical awareness (modified from the original “ethical decision-making”) and knowledge of the health care system.

After identifying the predictive components from previous interview data, three further domains were included: standard interview questions, language proficiency and general conduct, making altogether seven domains of interest.

The SMS also decided to assess both verbal and written communication skills. This, then, allowed for blueprinting of the desired domain and methods, as exemplified in Table 1.

Table 1: Example blueprint for MMI in the School of Medical Sciences, USM

	Language Proficiency	General Conduct	Critical Thinking	Ethical Awareness	Communication Skills	Knowledge of the health care system	Standard Interview questions
Verbal							
Station 1	✓	✓	✓	✓			
Station 2	✓	✓	✓	✓			
Station 3	✓	✓			✓		
Written							
Station 4	✓					✓	
Station 5	✓				✓		✓

A blueprint offered several advantages, namely: ensuring *content validity* of the MMI and allowing switching of formats (verbal vs. written) for different MMI sessions. This is to counter expected leakage of question content and format between sessions.

Interviewer Requirement

Interviewers were identified using the same process as with the previous interviews. They were required for verbal as well as written stations. To allay boredom, the interviewers were allowed to switch stations after completing one session.

The MMI in the SMS also used non-academic interviewers. We felt that the use of interviewers from stakeholder groups (hospital and nursing staff and selected members of the public) were justified, as they will be the ones receiving the graduates of the SMS.

Rating Instrument and Procedure

As with McMaster University and as recommended in the MMI and OSCE literature, a global rating instrument was developed and used. Each station rated up to four domains as explained previously. As can be seen in Table 1, for example, one verbal station looks at general conduct, language proficiency, critical thinking and ethical awareness, and another verbal station looks at general conduct, language proficiency and communication skills. Meanwhile, another separate written station looks at language proficiency and knowledge of the health care system.

Although interviewers rated each of these domains separately, they were also asked to give a separate global rating based on the overall performance of the candidates for each station. This was based on the finding that ratings for separate domains actually correlated highly with each other, therefore making using them separately for decision-making redundant (5). However, the separate ratings were still recorded as it

provides a guide for the interviewer for the global rating. It also allows a more detailed feedback regarding a particular candidate if needed.

For the written stations, interviewers rated the answer scripts using the same instrument. The overall performance of the candidate is the average of his or her performance in all five stations.

Implementation for Interviewers

Before the MMI sessions, interviewers were asked to attend an interviewer-training workshop or briefing. The workshop introduced the concept of the MMI; explained the justifications as well as detailed implementation issues. A calibration session, using a videotaped mock interview station, was also conducted. The workshop session was held a few weeks before the interview.

They were also asked to attend a pre-interview briefing on the morning of the interview. Due to the need to create six identical circuits running concurrently, all interviewers sitting in the same station needed to have the same understanding regarding the station, its implementation and standards expected of candidates.

In the stations, interviewers were provided with an interview manual and the question of the station (see Figure 1 for a sample question). Note that the question states the language that needs to be used (either bahasa Melayu or English, or for some stations they were given the option to choose the language) and the medium (verbal or written). Interviewers were also provided with an interviewer guide for their particular stations (this guide informed them the domains assessed with some content suggestions. At times interviewers were also informed about any roles they were supposed to assume, especially in communication stations). A summarised guide of anchor statements for rating the various domains were also provided.

Question:

Your friend's mother is terminally ill with cancer and is sent home from the hospital. Your friend learns that there is a Chinese sinseh who has herbs to cure his mother's cancer but he cannot come up with the RM50,000 to buy the herbs.

Discuss verbally with the interviewer what you would advise your friend.

This station can be answered in ENGLISH or BAHASA MELAYU. DO NOT mix the languages.

Soalan:

Emak rakan anda sakit tenat dengan kanser dan di hantar pulang dari hospital. Rakan anda mendapat maklumat tentang seorang sinseh Cina yang mempunyai herba untuk merawat kanser ibunya tetapi tidak mampu membayar RM50,000 untuk herba itu.

Bincangkan secara lisan dengan penemuduga nasihat yang anda akan berikan kepada rakan anda.

Soalan ini boleh dijawab dalam BAHASA MELAYU atau BAHASA INGGERIS. JANGAN mencampur-adukkan kedua bahasa tersebut.

Figure 1: Sample MMI Question used in the SMS

Implementation for Candidates

Candidates were given written documentation about the general concept and implementation of the MMI together with their interview letters. In addition, they were given another briefing after registering for the interview.

Evaluation of the MMI in SMS

Objectives for the evaluation were to determine the validity, reliability, acceptability and feasibility of the MMI. Before starting the interview, written consent for filling evaluation questionnaires was obtained from candidates.

For testing acceptability of MMI, self-administered evaluation questionnaires were administered both to candidates and interviewers at the end of each session. The questionnaires included evaluation

items answered using 7-point Likert scales. For comparison purposes, the items were modified from the evaluation questionnaires used by McMaster University in their MMI (5).

For testing feasibility of MMI, a comparative analysis was performed to previous interview exercises in terms of number of interviewers involved per hour, number of candidates seen per hour, and total time seen per candidate (minutes).

For testing the quality of MMI stations, difficulty and discrimination indices of the questions were measured. These indices were then utilised using the Item Classification Guide (18) to classify items (questions) into Level I, Level II, Level III and Level IV. Table 2 further explains the significance of these levels.

Table 2: Item classification guide by difficulty and discrimination

Item class	Item difficulty	Item discrimination	Description
Level I	0.45 to 0.75	+0.20 or higher	Best item statistics; use most items in this range if possible
Level II	0.76 to 0.91	+0.15 or higher	Easy; use sparingly
Level III	+0.10 or higher	+0.10 or higher	Difficult; use very sparingly and only if content is essential – rewrite if possible
Level IV	<0.24 or >0.91	Any discrimination	Extremely difficult or easy; do not use unless content is essential

For testing construct validity of our MMI, confirmatory factor analysis (CFA) was performed to test the measurement model of the latent construct. CFA was performed using Analysis of Moment Structure

(AMOS) software Version 22. The latent construct is considered fit if all goodness-of-fit indices achieve the minimal requirement as stated in Table 1, adopted from Yusoff & Arifin study (20).

Table 3: Goodness-of-fit indices levels of acceptance to signify a model fit

Name of category	Name of index	Level of acceptance
Absolute fit ¹	Root Mean Square of Error Approximation (RMSEA)	less than 0.08 (21)
	Goodness of Fit Index (GFI)	more than 0.9 (22)
Incremental fit ²	Comparative Fit Index (CFI)	more than 0.9 (23)
	Tucker-Lewis Index (TLI)	more than 0.9 (24)
	Normed Fit Index (NFI)	more than 0.9 (25)
Parsimonious fit ³	Chi Square/Degree of Freedom (Chisq/df)	less than 5 (26)

Note: ¹Absolute fit: Measures overall goodness-of-fit for both the structural and measurement models collectively. This type of measure does not make any comparison to a specified null model (incremental fit measure) or adjust for the number of parameters in the estimated model (parsimonious fit measure). ²Incremental fit: Measures goodness-of-fit that compares the current model to a specified “null” (independence) model to determine the degree of improvement over the null model. ³Parsimonious fit: Measures goodness-of-fit representing the degree of model fit per estimated coefficient. This measure attempts to correct for any “overfitting” of the model and evaluates the parsimony of the model compared to the goodness-of-fit.

The convergent validity of our MMI was checked through the size of factor loading and average variance extracted (AVE). Convergent validity refers to the degree to which a measure correlates with other measures that theoretically gauge similar attributes (27). Factor loading more than 0.5 is considered as reasonably high and AVE values more than 0.5 signify convergent validity (28–30). AVE was calculated manually following the formula recommended by Fornell & Larcker (30) and Hair et al. (28).

The reliability of our MMI construct was assessed by internal consistency. Composite reliability (CR) was used to measure the

internal consistency and a CR value more than 0.6 signifies satisfactory level of internal consistency (30).

Results

Validity and Reliability

Using the Item Classification Guide by difficulty and discrimination (18), if each station is considered separately then 40% of the stations are Level I, 43.3% are Level II, 3.3% are Level III and 13.4% are Level IV.

A summary of the difficulty and discrimination indices and the Item Level of each MMI station is given in Table 4.

Table 4: Difficulty and discrimination indices and item level of MMI stations by interview sessions

Parameter	MMI 01	MMI 02	MMI 03	MMI 04	MMI 05
Session 1: N = 160					
Difficulty index	0.57	0.69	0.78	0.84	0.73
Discrimination index	0.63	0.51	0.37	0.30	0.56
Item level	I	I	II	II	I
Session 2: N = 100					
Difficulty index	0.47	0.90	0.85	0.92	0.74
Discrimination index	0.52	0.30	0.37	0.07	0.56
Item level	I	II	II	IV	I
Session 3: N = 85					
Difficulty index	0.64	0.74	0.88	0.81	0.80
Discrimination index	0.48	0.57	0.35	0.30	0.30
Item level	I	I	II	II	II
Session 4: N = 59					
Difficulty index	0.68	0.88	0.86	0.85	0.75
Discrimination index	0.69	0.31	0.38	0.19	0.38
Item level	I	II	II	II	I
Session 5: N = 27					
Difficulty index	0.52	0.59	0.89	0.44	0.96
Discrimination index	0.44	0.56	0.33	0.78	0.11
Item level	I	I	II	III	IV
Session 6: N = 16					
Difficulty index	0.81	0.88	0.94	0.75	1.00
Discrimination index	0.50	0.50	0.25	0.50	0.00
Item level	II	II	IV	I	IV

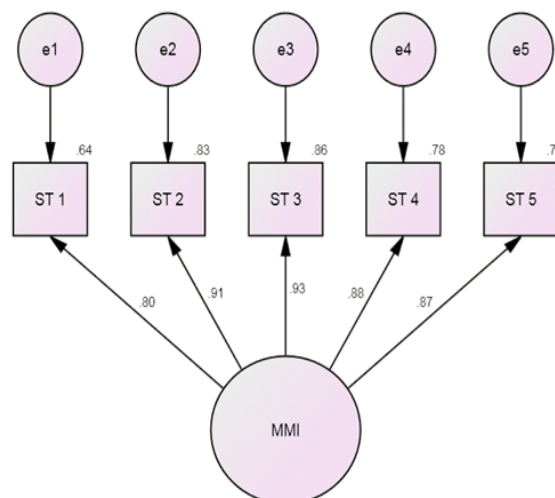
**Figure 2:** Construct validity of MMI

Table 5: The results of confirmatory factor analysis

Variable	χ^2 - statistic (df)	p-value	Goodness of fit indices						
			Cmin/df	RMSEA	GFI	CFI	AGFI	NFI	TLI
One-factor model	9.37 (5)	0.095	1.87	0.040	0.993	0.998	0.979	0.0.996	0.997

Confirmatory factor analysis revealed a unidimensional construct of MMI as all the goodness of fit indices attained an acceptable level of model fit (Figure 2 and Table 5). The results support the construct validity of MMI. The calculated AVE value was 0.77 that signifies convergent validity. In addition, the reliability analysis revealed the CR value was 0.94, suggesting a high level of internal consistency.

Feasibility

A total of 447 (81.7%) candidates out of the expected 547 attended the interview; the majorities attending the first and second sessions. Thirty interviewers were involved per session.

Table 6 summarises the feasibility of our MMI compared with previous interview exercises.

Table 6: Analysis of the feasibility of MMI compared to previous interview exercises.

Interview method	Applicants	Number of interviewers involved per hour	Number of candidates seen per hour	Total time seen per candidate (minutes)
Personal interview (2009–2014)	515 (average applicants from 2009–2014)	38	57	20
MMI (2015)	447 (63 minutes per 9 candidates)	30	54	25

Acceptability

Table 7 summarises the responses of the candidates regarding their experience. Candidates were very positive as regards the fairness and implementation of the MMI (i.e. their ability to accurately portray themselves, anxiety caused by the format, effect of format on applying to USM, and adequacy and clarity of instructions) where the mean of ratings were all above 5.

Candidates' agreements were more spread when asked about special knowledge required to answer any of the stations and their difficulty. However, the majority of their mean ratings were all above 3.5 and near to 4, except for one station (Station 3). Similar results were found as regards the time available for the stations. The mean ratings were all above 3.5 except for one station (Station 1).

Table 8 summarises the responses of interviewers regarding their experience. The majority of the interviewers were positive about their ability to accurately assess the candidates, the administration of the MMI, the adequacy of preparatory materials

provided and the clarity of instructions given to candidates.

A qualitative study of the candidates' subjective responses will be presented in a separate report.

Table 7: The feedback of candidates towards MMI

Item	Minimum	Maximum	Mean	Std. Deviation
Do you believe that you were able to present an accurate portrayal of your ability? 1: Definitely not, 3: Not really, 5: Somewhat, 7: Definitely	2	7	5.74	1.090
Compared to the personal interview do you think the MMI would cause candidates more or less anxiety? 1: A lot more, 3: A little more, 5: A little less, 7: A lot less	1	7	5.23	1.598
Would the use of the MMI encourage you from applying to USM? 1: Definitely not, 3: Not really, 5: Somewhat, 7: Definitely	1	7	5.62	1.819
Were the instructions given before the MMI adequate to prepare you for the experience? 1: Definitely not, 3: Not really, 5: Somewhat, 7: Definitely	1	7	6.22	1.130
Were the instructions given before each station clear enough? 1: Definitely not, 3: Not really, 5: Somewhat, 7: Definitely	1	7	6.53	.856
Do you think any of the interviews required specialised knowledge? 1: None, 4: Somewhat, 7: A lot				
MMI 01	1	7	4.63	2.199
MMI 02	1	7	3.90	2.022
MMI 03	1	7	3.70	2.029
MMI 04	1	7	4.09	2.134
MMI 05	1	7	4.13	2.082

(continued on next page)

Table 7: (continued)

Item	Minimum	Maximum	Mean	Std. Deviation
How difficult was each interview?				
1: Easy, 3: Somewhat Easy, 5: Difficult, 7: Very Difficult				
MMI 01	1	7	3.78	1.682
MMI 02	1	7	3.59	1.536
MMI 03	1	7	3.38	1.527
MMI 04	1	7	3.57	1.643
MMI 05	1	7	3.53	1.635
Was the time available for each station appropriate?				
1: Too little, 4: Well timed, 7: Too much				
MMI 01	1	7	3.12	1.394
MMI 02	1	7	3.63	1.167
MMI 03	1	7	3.68	1.047
MMI 04	1	7	3.59	1.210
MMI 05	1	7	3.67	1.184

Table 8: The feedback of examiners towards MMI

Item	Minimum	Maximum	Mean	Std. Deviation
Do you believe that you were able to develop an accurate portrayal of the candidates?				
1: Definitely not, 3: Not really, 5: Somewhat, 7: Definitely				
	3	7	5.38	1.106
Compared to a more traditional interview, do you think the MMI would be more or less difficult to administer (from the point of view of an examiner)?				
1: A lot more, 3: A little more, 5: A little less, 7: A lot less				
	3	7	5.74	1.138
Were the materials provided before the MMI adequate to prepare you for the experience?				
1: Definitely not, 3: Not really, 5: Somewhat, 7: Definitely				
	3	7	5.84	0.817
Were the instructions given to candidates before your station clear enough?				
1: Definitely not, 3: Not really, 5: Somewhat, 7: Definitely				
	2	7	6.08	0.900

Discussion

Validity and Reliability

The confirmatory factor analysis revealed a unidimensional construct, which supports our decision to define a candidate's overall performance in the MMI as the mean of the scores in the five stations. Validity evidence of this construct include content validity evidence, as intended by the blueprinting process (31). Interviewers were also positive regarding their ability to assess the candidates fairly. Further evidence includes interviewer and candidate preparation before the MMI, intended as response process evidence. Further evidence of validity is needed to look at predictive validity and follow-up studies for this cohort are planned.

It is a relief to note that the overall reliability of the MMI exercise was 0.94, an acceptable level to make high-stake decisions (32). The overall reliability reported by Eva and colleagues in their pilot was 0.65–0.81 (5). Even though in their pilot MMI studies more stations were utilised, only one circuit was used, involving the rating of 117 candidates. The reliability of 0.94 in our exercise is understandable as in our MMI, 447 candidates were rated using six identical circuits over several sessions. It is reassuring to note that a similar 5-station MMI which rated 237 candidates also reported a similar reliability of > 0.9 (33). Despite the reassuring data, we will still consider adding more stations to future MMIs as long as feasibility permits, as common assessment sense would suggest that it will increase the validity and reliability further.

What is more concerning, however, is the variation of difficulty and discrimination of the stations between sessions. This is most likely attributable to interviewer training or question design. If the item level does not vary very much between sessions, this might indicate the quality of the question, assuming that the standards of the interviewers are similar. On the

other hand, a highly fluctuating item level between sessions might indicate varying interviewer standards between sessions. However, investigation of this issue is impeded by the fact that questions for the morning and afternoon sessions were different, as were the questions from one day to another. Apart from that, at any station interviewers were different from session to session, and that for every session, there are six identical circuits running the same stations. This makes making any conclusions based on the differences between sessions difficult. It is likely that in the future interviewers will be asked to stay in the same station for the duration of the morning or afternoon sessions. It is interesting to note that the initial intention to allow rotation of examiners after every session was to allay boredom, as suggested by an interviewer in the pilot as reported by Eva and colleagues (5). In retrospect, the caution mentioned in the same article that against this is "...the cost of lessening the examiner-identified benefit of an increased 'ability to set a standard for expected responses' and the improved 'consistency of comparing responses' that develops from seeing a large number of candidates work through the same station" probably holds true. At any rate, this is extremely valuable information for future question design and MMI implementation and highlights the importance of interviewer training. One of the aims for future MMI and training sessions in the SMS is surely to reduce this variation as much as possible.

Feasibility

It was decidedly an advantage that since its inception the SMS uses the OSCE as one of its assessment formats. Minimal introduction was needed for the interviewers. However, familiarising new support staff was very important as close coordination and vigilance were needed to handle the large numbers of candidates and interviewers and running multiple MMI circuits simultaneously. Staff with previous experience handling OSCEs were a bonus

but we noticed that it took just a short time for new staff to understand the flow of the MMI.

The SMS is also blessed with an examination ward, complete with multiple mini-wards and discussion rooms. As such, having a suitable venue for the MMI was not a problem. However, this luxury was not available in the two interview centers in Sabah and Sarawak. In both centres, large conference rooms were utilised. We managed to construct one circuit consisting of the nine stations in each of the two centres. A single circuit was adequate to cater for the reduced number of candidates in both. Each station was acceptably far apart from one another but of course some compromises had to be made on noise and privacy. This is certainly a factor to be considered in the future, for example if stations using standardised patients are to be used.

In terms of human resource, there are some differences noted. To finish up to 500 candidates over two days, the previous personal interviews utilised up to 19 panels of 2 interviewers at a time, meaning up 38 interviewers were used at any time. They are expected to take up to 20 minutes per candidate, including time for filling up the rating form. In terms of efficiency, this means that the interview can handle up to 57 candidates per hour. However, experience told us that invariably there were differences in the time taken for each panel to finish each candidate; conscientious interviewers usually extended the time taken to 30 minutes or more, while some took less time than the standard 20 minutes. The faster panels usually ended up taking more candidates, as support staff usually redirects waiting candidates to fill up empty interview rooms. Panels finishing late in the afternoon were also not uncommon.

In contrast, the efficiency of our MMI is approximately 54 candidates per hour (each session taking 63 minutes with 6 circuits and 9 candidates per circuit). Each candidate was seen for a total of 25 minutes (by five different interviewers) as compared to

being seen for 20 minutes (which was still highly variable) by a pair of interviewers in one sitting. Because it ran on a strict time schedule, interviewers expressed their satisfaction that everybody saw the same number of candidates for the same durations. No interviewer went home late.

Financially, however, the MMI does incur some extra costs in the form of honorariums for non-SMS interviewers and travel and accommodation expenses for five interviewers who were sent to the peripheral interview centres. Previously only two interviewers (one panel) were sent.

Acceptability

As with other institutions reporting the implementation of MMI (5, 34–40), the MMI was overall positively accepted by both candidates and interviewers in our school.

Candidates thought that they were able to portray themselves accurately and generally agreed that the MMI format caused less anxiety. This is important as further evidence to the validity of the exercise. It is interesting to note that the third item in the candidate questionnaire was worded differently from the original version, in that they were asked whether the MMI would encourage them to apply to USM, as opposed to discourage. Following the trend, this was also answered positively. We are heartened to observe that items about adequacy of the instructions were also answered positively. As regards whether stations required specialised knowledge, the mean of the responses for the stations hovered around 4: somewhat. This is good feedback to us, as our aim is to have the stations not requiring any specialised knowledge at all. This may require more fine-tuning of the questions at the committee level. The difficulty of the stations also hovered around 4, between “Somewhat easy” and “Difficult”. This is also desirable as stations, which are too easy or too difficult, would not be discriminative. The responses to adequacy of time was particularly relevant, as for the

sake of feasibility we took the decision to run the stations in five minutes, compared to the eight minutes as reported by Eva and colleagues. To our relief, most of the stations were rated near 4 (well-timed); only one station obtained a mean rating of 3.12, which is still acceptable.

The interviewers were also of the opinion that they were somewhat able to develop an accurate portrayal of candidates, another evidence of the validity of the MMI. They were also positive that the MMI was less difficult to administer compared to the traditional interview. It is interesting to note that the mean rating for this item in USM was more than 10% more positive than in McMaster. Although a qualitative report of the evaluation will be offered later, anecdotal feedback included praise for the MMI, which started and ended on time, compared to the traditional personal interview, which was often prolonged out of the interviewers' concern about not making an adequate decision based on the time available. Related to this, interviewers expressed their relief that the decision regarding a candidate is now shared between five examiners, and not resting squarely on their shoulders as was the previous method.

Conclusion

In conclusion, we are cautiously optimistic of the positive feedback obtained from the evaluation data, despite the relatively small number of stations and shorter time per station. We have identified areas for further improvement, including standardisation of station quality and interviewer training. Continuous monitoring and evaluation is planned for improvement and obtaining predictive validity evidence.

Acknowledgements

We would like to thank the Dean and Academic Deputy Dean of SMS for their full support of the MMI process, and we would like to express our gratitude to Prof.

Kevin Eva for coming over to our Kelantan Campus to give an introductory workshop on MMI.

References

1. Yusoff MSB, Abdul Rahim AF, Jaa'far R. Medical education department roles and initiatives towards achieving APEX agenda. *Education in Medicine Journal*. 2014;6(2):e1–e7. doi: 10.5959/eimj.v6i2.261.
2. Yusoff MSB, Rahim AFA, Baba AA, Esa AR. Medical student selection process and its pre-admission scores association with the new students' academic performance in Universiti Sains Malaysia. *Int Med J*. 2011;18(4):329–33.
3. Albanese MA, Snow MH, Skochelak SE, Huggett KN, Farrell PM. Assessing personal qualities in medical school admissions. *Academic Medicine*. 2003;78(3):313–21. doi: 10.1097/00001888-200303000-00016.
4. Eva K, Reiter H. Where judgement fails: pitfalls in the selection process for medical personnel. *Adv Health Sci Educ*. 2004;9(2):161–74.
5. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Medical Education*. 2004;38(3):314–26. doi: 10.1046/j.1365-2923.2004.01776.x.
6. Dowell J, Lynch B, Till H, Kumwenda B, Husbands A. The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Medical Teacher*. 2012; 34(4): 297–304. doi: 10.3109/0142159X.2012. 652 706.
7. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ*. 2004;38(3):314–26. doi: 10.1046/j.1365-2923.2004.01776.x.
8. Eva KW, Macala C. Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Medical Education*. 2014;48(6):604–13. doi: 10.1111/medu.12402.

9. Reiter HI, Salvatori P, Rosenfeld J, Trinh K, Eva KW. The effect of defined violations of test security on admissions outcomes using multiple mini-interviews. *Med Educ.* 2006;40(1):36–42. doi: 10.1111/j.1365-2929.2005.02348.x.
10. Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education.* 2007;41(4):378–84. doi: 10.1111/j.1365-2929.2007.02709.x.
11. Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine.* 2004;79(10):S40–S2. doi: 10.1097/00001888-200410001-00012.
12. Dore KL, Kreuger S, Ladhani M, Rolfson D, Kurtz D, Kulasegaram K, et al. The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Academic Medicine.* 2010;85(10):S60–S3. doi: 0.1097/ACM.0b013e3181ed442b.
13. Eva KW, Reiter HI, Rosenfeld J, Norman GR. The relationship between interviewers' characteristics and ratings assigned during a multiple mini-interview. *Academic Medicine.* 2004;79(6):602–9. doi: 10.1097/00001888-200406000-00021.
14. Eva KW, Reiter HI, Rosenfeld J, Trinh K, Wood TJ, Norman GR. Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *JAMA.* 2012;308(21):2233–40. doi: 10.1001/jama.2012.36914.
15. Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education.* 2009;43(8):767–75. doi: 10.1111/j.1365-2923.2009.03407.x.
16. Reiter HI, Salvatori P, Rosenfeld J, Trinh K, Eva KW. The effect of defined violations of test security on admissions outcomes using multiple mini-interviews. *Medical Education.* 2006;40(1):36–42. doi: 10.1111/j.1365-2929.2005.02348.x.
17. Rosenfeld J, Reiter H, Trinh K, Eva K. A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Adv Health Sci Educ.* 2008;13(1):43–58.
18. Downing SM, Yudkowsky R. *Assessment in health professions education.* New York: Routledge; 2009.
19. Pau A, Jeevaratnam K, Chen YS, Fall AA, Khoo C, Nadarajah VD. The multiple mini-interview (MMI) for student selection in health professions training – a systematic review. *Medical Teacher.* 2013;35(12):1027–41. doi: 10.3109/0142159X.2013.829912.
20. Yusoff MSB, Arifin WN. Educational environment and psychological distress of medical students: the role of deep learning approach. *Journal of Taibah University Medical Sciences.* 2015;10(4):411–8. doi: 10.1016/j.jtumed.2015.08.005.
21. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociological Methods & Research.* 1992;21(2):230–58. doi: 10.1177/0049124192021002005.
22. Jöreskog K, Sörbom D. *LISREL VI users guide.* 3rd ed. Moorsville, IN: Scientific Software; 1984.
23. Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin.* 1990;107(2):238. doi: 10.1037/0033-2909.107.2.238.
24. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin.* 1980;88(3):588. doi: 10.1037/0033-2909.88.3.588.
25. Bollen KA. A new incremental fit index for general structural equation models. *Sociological Methods & Research.* 1989;17(3):303–16. doi: 10.1177/0049124189017003004.
26. Marsh HW, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: first-and higher order factor models and their invariance across groups.

- Psychological Bulletin. 1985;97(3):562. doi: 10.1037/0033-2909.97.3.562.
27. Streiner LD, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. New York: Oxford University Press; 2008. doi: 10.1093/acprof:oso/9780199231881.001.0001
 28. Hair JJ, Black W, Babin B, Anderson R. Multivariate data analysis. Upper Saddle River, NJ: Pearson Prentice-Hall; 2009.
 29. Zainudin A. Structural equation modeling using AMOS graphic. Shah Alam: UiTM Press; 2012.
 30. Fornell C, Larcker D. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*. 1981;18(1):39–50. doi: 10.2307/3151312.
 31. Downing SM. Validity: on meaningful interpretation of assessment data. *Medical Education*. 2003;37(9):830–7. doi: 10.1046/j.1365-2923.2003.01594.x.
 32. Downing SM. Reliability: on the reproducibility of assessment data. *Medical Education*. 2004;38(9):1006–12. doi: 10.1111/j.1365-2929.2004.01932.x.
 33. Fraga JD, Oluwasanjo A, Wasser T, Donato A, Alweis R. Reliability and acceptability of a five-station multiple mini-interview model for residency program recruitment. *J Community Hosp Intern Med Perspect*. 2013;3(3–4).
 34. Brownell K, Lockyer J, Collin T, Lemay JF. Introduction of the multiple mini interview into the admissions process at the University of Calgary: acceptability and feasibility. *Medical Teacher*. 2007;29(4):394–6. doi: 10.1080/01421590701311713.
 35. Hofmeister M, Lockyer J, Crutcher R. The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Medical Education*. 2009;43(6):573–9. doi: 10.1111/j.1365-2923.2009.03380.x.
 36. Dore KL, Kreuger S, Ladhani M, Rolfson D, Kurtz D, Kulasegaram K, et al. The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Acad Med*. 2010;85(10 Suppl):S60–3. doi: 10.1097/ACM.0b013e3181ed442b.
 37. McAndrew R, Ellis J. An evaluation of the multiple mini-interview as a selection tool for dental students. *Br Dent J*. 2012;212(7):331–5. doi: 10.1038/sj.bdj.2012.267.
 38. Cameron AJ, Mackeigan LD. Development and pilot testing of a multiple mini-interview for admission to a pharmacy degree program. *Am J Pharm Educ*. 2012;76(1):10. doi: 10.5688/ajpe76110.
 39. Perkins A, Burton L, Dray B, Elcock K. Evaluation of a multiple-mini-interview protocol used as a selection tool for entry to an undergraduate nursing programme. *Nurse Educ Today*. 2013;33(5):465–9. doi: 10.1016/j.nedt.2012.04.023.
 40. Ahmed A, Qayed KI, Abdulrahman M, Tavares W, Rosenfeld J. The multiple mini-interview for selecting medical residents: first experience in the Middle East region. *Medical Teacher*. 2014;36(8):703–9. doi: 10.3109/0142159X.2014.907875.

