

Review Article

Multiple Mini Interview as an admission tool in higher education: Insights from a systematic review

Muhamad S. Bahri Yusoff, PhD

Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, Kota Bharu, Malaysia

Received 16 February 2019; revised 24 March 2019; accepted 27 March 2019; Available online 10 May 2019



المخلص

أهداف البحث: تم استخدام المقابلات المختصرة المتعددة حول العالم فيما يتعلق باختبار الطلاب، وخاصة في التعليم المهني الصحي. قدمت هذه الورقة تقريراً عن دليل صلاحية المقابلات المختصرة المتعددة في بيئات تعليمية مختلفة.

طرق البحث: تم البحث في الأدبيات من خلال قواعد بيانات "سكوبس" و"ساينس دايركت" و"جوجل سكولر" و"بب مد" و"إيسكو هوست" بناءً على مصطلحات بحث محددة. وتم تقييم كل مقالة بناءً على العنوان والمخلص والنص الكامل. قُيِّمت المقالات المختارة تقييماً نقدياً وتمت صياغة المعلومات ذات العلاقة بدعم صحة المقابلات المختصرة المتعددة في البيئات التعليمية المختلفة. اتبعت هذه الورقة إرشادات "بريزما" لضمان الالتزام المنهجي في الإبلاغ عن نتائج المراجعات المنهجية.

النتائج: كانت غالبية الدراسات من كندا بنسبة 41.54%، تليها المملكة المتحدة (25.39%) والولايات المتحدة (13.85%) وأستراليا (9.23%)، وكان الباقي (9.24%) من ألمانيا وإيرلندا والإمارات العربية المتحدة واليابان وباكستان وتايوان وماليزيا. علاوة على ذلك، تراوح عدد المقابلات المختصرة المتعددة في الغالب من 7 إلى 12 مع مدة 10 دقائق لكل محطة (بما في ذلك فجوة دقيقتين بين المحطات).

الاستنتاجات: أوضحت النتائج إلى أن المحتوى وعملية الاستجابة والبنية الداخلية للمقابلات المختصرة المتعددة كانت مدعومة بصورة جيدة بالأدلة، إلا أن علاقات وعواقب المقابلات المختصرة المتعددة مع متغيرات مهمة للمخرجات كانت مدعومة بشكل غير منتظم. وقد أظهرت الأدلة أن المقابلات المختصرة المتعددة كانت أداة قبول غير متحيزة وعملية ويمكن تطبيقها وموثوقة ومضمونة المحتوى. إلا أن هناك حاجة إلى مزيد من البحث عن أثارها على النتائج غير الإدراكية.

الكلمات المفتاحية: المقابلات المختصرة المتعددة؛ طريقة قبول؛ تعليم عالي؛ صلاحية؛ التعليم المهني الصحي

Abstract

Objectives: Multiple Mini Interviews (MMI) have been conducted across the globe in the student selection process, particularly in health profession education. This paper reported the validity evidence of MMI in various educational settings.

Methods: A literature search was carried out through Scopus, Science Direct, Google Scholar, PubMed, and EBSCOhost databases based on specific search terms. Each article was appraised based on title, abstract, and full text. The selected articles were critically appraised, and relevant information to support the validity of MMI in various educational settings was synthesized. This paper followed the PRISMA guideline to ensure consistency in reporting systematic review results.

Results: A majority of the studies were from Canada, with 41.54%, followed by the United Kingdom (25.39%), the United States (13.85%), and Australia (9.23%). The rest (9.24%) were from Germany, Ireland, the United Arab Emirates, Japan, Pakistan, Taiwan, and Malaysia. Moreover, most MMI stations ranged from seven to 12 with a duration of 10 min per station (including a 2-min gap between stations).

Conclusion: The results suggest that the content, response process, and internal structure of MMI were well supported by evidence; however, the relation and consequences of MMI to important outcome variables were inconsistently supported. The evidence shows that MMI is a non-biased, practical, feasible, reliable, and content-valid admission tool. However, further research on its impact on non-cognitive outcomes is required.

Corresponding author: Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kota Bharu, Malaysia.

E-mail: msaiful_bahri@usm.my

Peer review under responsibility of Taibah University.



Production and hosting by Elsevier

Keywords: Admission method; Health profession education; Higher education; Multiple mini interview; Validity

© 2019 The Author.

Production and hosting by Elsevier Ltd on behalf of Taibah University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Interviews for the selection of students nowadays have become more important as higher education institutions seek capable candidates to enrol in their courses, especially courses related to health and medical sciences. It is widely known that the study of medicine is highly regarded by society and is often considered a difficult and demanding course, as enrolment places are limited.¹ There is a new interview format known as Multiple Mini Interview (MMI).² MMI was developed to dilute the impact of individual examiners and allow them to perform more valid rating of candidate performance.^{3,4}

MMI is an OSCE-style exercise that consists of multiple and focused encounters to assess various cognitive and non-cognitive skills of the candidates.² Basically, the MMI consists of a series of 6–10 situational interviews, each of which poses a non-medical question designed to assess specific non-academic qualities of applicants.⁵ In terms of the arrangement, each circuit has 6–10 stations, and each station involves a situational interview. One or two interviewers or a panel are placed at each station to mark the candidates. The number of interviewers sometimes depends on the situation given. The flexibility of the MMI allows programs to select applicants whose behaviour best aligns with the expected competency.⁶

A recent systematic review revealed that MMIs used for the selection of undergraduate health programs appear to have reasonable feasibility, acceptability, validity, and reliability.⁷ Furthermore, the systematic review concluded that MMI represented a non-biased selection tool for applicants on the basis of age, gender, or socio-economic status, but applicants of certain ethnic and social backgrounds demonstrated low performance in a very small number of published studies.⁷ The latest article included in the systemic review was in 2014, and it only focused on the utility of MMI in health profession education. This paper reports the latest validity evidence of MMI as an admission tool, either within or outside the health profession education context.

Materials and Methods

We conducted this systematic review based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) for a standard reporting of systematic review.⁸ PRISMA helps to provide complete transparency and good reporting for systematic and meta-analysis review. PRISMA includes 27 checklist items to improve title,

abstract, methods, results, discussion, and funding reporting quality.

Study questions

The primary focus was to discover evidence to support the validity of MMI from five sources⁹: 1) **Content:** Do MMI stations represent the construct? The extent of MMI includes a specific set of items (i.e. station characteristics) to reflect the content of the intended attribute to be assessed; 2) **Response process:** Are MMI items completely understood by the subjects? This concerns the relationship between the intended construct and the thought processes of subjects while responding to the items; 3) **Internal structure:** Do MMI items measure the proposed constructs? This deals with the degree of relationship between and among items and constructs as proposed and commonly represented by reliability and factor structure; 4) **Relations to other variables:** Do MMI scores correlate with other variables? This is about the relationship of MMI scores to external variables measured by another instrument assessing similar concepts or specific sets of criteria. It can be represented in the form of convergent, discriminant, predictive, and concurrent validity; and 5) **Consequences of a measurement:** Do MMI scores really make a difference? This addresses evidence regarding the significance of measurement scores on specific intended or unintended outcomes.

Study eligibility

Broad criteria were utilised to present a comprehensive MMI outlook within and beyond health profession education. Original articles published in English that reported the validity evidence of MMI either within or outside the health profession education context were included.

Study identification

A literature search was performed through Scopus, Science Direct, EBSCO Host, Google Scholar, and PubMed database to search articles related to the MMI using search terms such as ‘Multiple Mini Interview’ and ‘MMI’. No time limit was specified in searching, and the last search date was in December 2016.

Study selection

The author performed the initial screening process of articles based on the title and abstract. Criteria such as participants, study design, validity evidence, and outcomes were the key issues for in-depth screening of the full articles. The selected articles underwent an in-depth appraisal based on the priori criteria for inclusion in the systematic review (The study selection is illustrated in [Figure 1.](#))

Results

Study flows

A total of 7470 potential articles were identified during the literature search using the search terms. Throughout the screening process, 69 articles were selected for the in-depth full-text study. After critical evaluation of the full texts, 64 articles were included in the systematic review.

Table 1 shows 49 articles reported evidence to support the content of MMI, while Table 2 shows that 40 articles support the internal structure^{2,4,10-47}; 37 articles support the response process^{2,5,6,11,12,14,18-20,23,25-28,30,31,33,37,45-62}; 21 articles support the relation to other variables^{4,15-17,21,24,25,29,38-40,44,54,57,59,61,63-67} and four articles support consequences^{31,35,36,49,58,66,68,69} of MMI. MMI has mainly been implemented in medical and health sciences: 51 (76%) reports were in medicine (i.e. 58% undergraduate and 18% postgraduate), while 16 (24%) reports were from other health sciences (i.e., 7% dentistry, 6% pharmacy, 3% nursing, 1.5% rehabilitation sciences, 1.5% physician assistance, 1.5% health sciences, 1.5% para-medicine, and 1.5% veterinary). No MMIs were reported outside of the medical and health sciences. Out of 64, a majority of the studies were from Canada (41.54%), followed by the United

Kingdom (25.39%), the United States (13.85%), and Australia (9.23%). The rest (9.24%) were from Germany, Ireland, the United Arab Emirates, Japan, Pakistan, Taiwan, and Malaysia.

Content. Table 1 summarizes the content of MMI in terms of station number, number of days required to conduct MMI, time required per station, and competencies assessed. Most MMIs were implemented in two to three days, per circuit consisting of seven to 12 stations, with each MMI station requiring seven to 10 min. The 10 most frequent competencies assessed by MMI were professionalism (n = 48); communication skills (n = 33); teamwork (n = 32); ethics and morals (n = 26); critical thinking and problem solving (n = 25); motivation to study (n = 19); empathy (n = 16); management skills (n = 14); resilience (n = 13); and interpersonal skills (n = 10).

Internal structure. Six studies reported an MMI internal consistency level (Cronbach's alpha) of less than 0.7,^{19,22,23,26,28,38} whereas 18 articles reported at least 0.70.^{11-13,16,18,19,21-24,26,31,32,34,37,43,44,47} The lowest and highest levels of internal consistency were 0.54³⁸ and 0.98,¹³ respectively (Table 2).

Response process. Both applicants and examiners were positive about the experience and potential of MMI as a student selection method^{2,6,14,26-28,33,37,46-49,51-53,55,56,60,61};

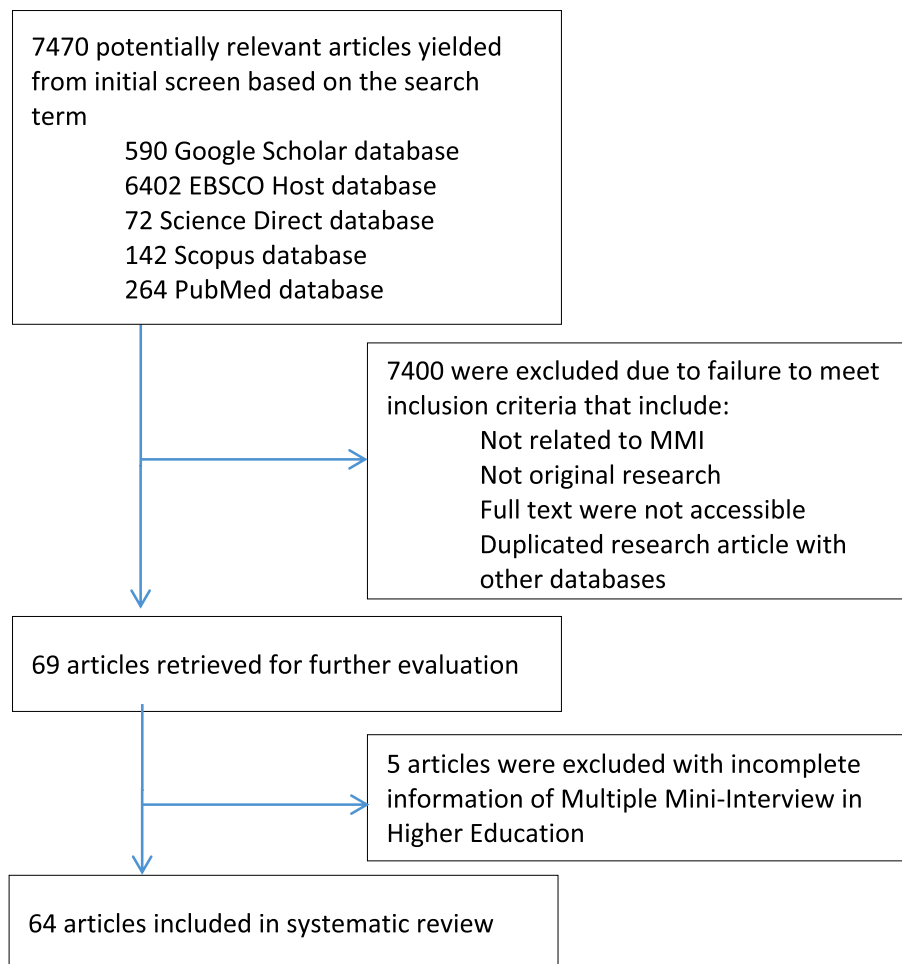


Figure 1: Study flow chart.

Table 1: The content of MMI stations.

Sources, university, country	Number of MMI station	Number of days taken	Times required per station	Competencies assessed
Eva K. W. et al., ³ 2004, McMaster University, Canada	10	4	8 min (iv), 2 min (break)	Critical thinking, ethical decision making, communication skills, knowledge of healthcare system
Eva K. W. et al., ⁴ 2004, McMaster University, Canada	10	—	8 min	Not available
Eva K. W. et al., ¹⁰ 2004, McMaster University, Canada	9	2	8 min (iv), 2 min (break)	Scholarship-critical thinking, healthcare advocacy, professionalism-ethical decision making, collaboration
Moreau K. et al., ¹¹ 2005, McMaster University, Canada	12	2	8 min (iv), 2 min (break)	Communication, collaboration, professional ethics
Reiter H. I. et al., ¹² 2006, McMaster University, Canada	Study 1 : 9 Study 2 : 12 Study 3 : 7	—	—	Not available
Brownell K et al., ¹³ 2007, University of Calgary, Canada	10	2	—	Compassion and Empathy, Honesty and Integrity, Ability To Tolerate Ambiguity, Reflective, Respect For Others
Harris S. and Owen C., ¹⁴ 2007, McMaster University, Canada	10	—	5.5 min	Giving instruction, taking instruction, emotional communication, problem solving, resilience & maturity, enthusiasm for medicine, ethics, awareness of common issues in medicine
Lemay J.F. et al., ¹⁵ 2007, University of Calgary, Canada	10	2	8 min (iv), 2 min (break)	Advocacy, ambiguity, collegiality & collaboration, cultural sensitivity, empathy, ethics, honesty & integrity, responsibility & reliability, self-assessment
Reiter H. I. et al., ¹⁶ 2007, McMaster University, Canada	8	—	8 min	Not available
Hofmeister M. et al., ⁶ 2008, University of Alberta, Canada	10	—	8 min (iv), 2 min (break)	Teamwork, honesty, ability to accept feedback about one's self, ability to accept self-limitations, caring & compassion, responsibility taking, time management, the ability to accept professional limitations, cultural sensitivity, motivation for family medicine, goal setting
Humphrey S. et al., ¹⁷ 2008, West Midlands Deanery, UK	3	—	5 min (iv), 1 min (break)	Not available
Roberts C. et al., ¹⁸ 2008, University of Sydney, Australia	8	8	7 min (iv), 2 min (break)	Not available
Rosenfeld J. M. et al., ¹⁹ 2008, McMaster University, Canada	12	2	8 min (iv), 2 min (break)	Discussion, interpersonal skills, cooperation
Eva K. W. et al., ²⁰ 2009, McMaster University, Canada	9	—	10 min	Not available

Table 1 (continued)

Sources, university, country	Number of MMI station	Number of days taken	Times required per station	Competencies assessed
Hofmeister M. et al., ²¹ 2009, University of Calgary and University of Alberta, Canada	12	1	8 min (iv), 2 min (break)	Family medicine attributes, teamwork, disclosure of error, ethical behavior, ability to accept feedback, ability to accept self-limitations, caring, taking responsibility, time management, ability to accept professional-limitations, cultural sensitivity, motivation & goal-setting, ability to handle specific situation
Kumar K et al., ²² 2009, University of Sydney, Australia (6 Days), Australia and University of British Columbia (2 Days), Canada	8	8	7 min (iv), 2 min (break)	Not available
Razack S. et al., ²³ 2009, McGill University Medical School, Canada	10	—	8 min (iv), 2 min (break)	Professionalism, empathy, communication skills
Roberts C. et al., ²⁴ 2009, University of Sydney, Australia	8	9	7 min	Not available
Dore K. L. et al., ²⁵ 2010, McMaster University and University of Alberta, Canada	7	—	8 min (iv), 2 min (break)	CanMEDs competencies: Medical expert, Communicator, Collaborator, Manager, Health advocate, Scholar, Professional
Kulasegaram K. et al., ²⁶ 2010, Michael G. DeGroot School of Medicine, McMaster University, Canada	12	—	8 min	Communication, collaboration, critical thinking, ethics, personal statement, understanding of the health care system
Roberts C. et al., ²⁷ 2010, University of Sydney, Australia	8	—	—	Not available
Jones, P. E. & Forister, J. G., ²⁸ 2011, McMaster University, Canada	8	—	7 min	Boundary recognition, responsibility, honesty, integrity, professionalism
O'Brien A. et al., ²⁹ 2011, St George's University of London, UK	4	2	5 min	Presentation skills, work experience, ethical thinking, professionalism
Uijtdehaage S. et al., ³⁰ 2011, David Geffen School of Medicine, University of California, US	Study 1 : 13 (1 rest station) Study 2 : 12	Study 1 : 3 days Study 2 : -	8 min (iv), 2 min (break)	Commitment to & experience with underserved populations, cultural sensitivity, leadership potential, maturity, being an effective team member
Yen W. et al., ³¹ 2011, Michener Institute for Applied Health Sciences, Toronto, Canada	8	—	—	Communication skills, taking responsibility on own action, ethical decision making, inter-professional collaboration, problem solving skills, reflective practice, time & resource management skills, resolves conflict
Cameron A. J. and MacKeigan L. D., ³² 2012, University of Toronto, Canada	10	—	8 min (5 station), 6 min (5 station)	Commitment to care, critical thinking, problem solving & creativity, ethical reasoning & integrity, interpersonal skills,

(continued on next page)

Table 1 (continued)

Sources, university, country	Number of MMI station	Number of days taken	Times required per station	Competencies assessed
J. Dowell et al., ³³ 2012, Dundee University, Scotland	4 (2007) 10 (2009)	—	—	motivation, communication skills, self-awareness, team player Interpersonal skills & communication (including empathy), logical reasoning & critical thinking, moral & ethical reasoning, motivation & preparation to study medicine, teamwork & leadership, honesty & integrity
Eva K. W. et al., ³⁴ 2012, McMaster University, Canada	12	—	8 min (iv), 2 min (break)	Ethical issues, communication, collaborative tasks
Griffin B. and Wilson I., ³⁵ 2012, Australian School of Medicine, Australia	9	—	8 min	Verbal communication, empathy, motivation to study medicine
Jerant A. et al., ³⁶ 2012, University of California, Davis (UCD), US	10	—	8 min (iv), 2 min (break)	Integrity/ethics, professionalism, interpersonal communication, diversity/cultural, awareness, teamwork, ability to handle stress, problem solving
McAndrew R. and Ellis J., ³⁷ 2012, Cardiff University, UK	10	5	5 min (iv), 2 min (break)	Dentistry as a career, logic, reasoning, ethics & plagiarism, manual dexterity, breaking bad news, research & data interpretation
Fraga J. D. et al., ³⁸ 2013, Reading Health System, West Reading, US	6 (5 MMI; 1 TI)	—	8 min (MMI), 20 (TI), 2 min (break),	Professionalism, communication skills, critical thinking, ethical behavior, tolerance for uncertainty, teamwork
Husbands A. & Dowell J., ³⁹ 2013, Dundee Medical School, Scotland	10	—	—	Interpersonal skills & communication (including empathy), logical reasoning & critical thinking, moral & ethical reasoning, motivation & preparation to study medicine, teamwork & leadership, honesty & integrity
McAndrew R and Ellis J., ⁴⁰ 2013, Cardiff University Dental Hospital, UK	10	—	—	Not available
Perkins A. et al., ⁴¹ 2013, Kingston University and St George's University of London, UK	5	—	5 + 5 min (1st station), 5 min (4 station)	NMC standard of competence: professional values, communication & interpersonal skills, nursing practice & decision making, & leadership, management, team working
Raghavan M et al., ⁴² 2013, University of Manitoba, Canada	11	—	8 min (iv), 2 min (break) 10 min (1 station for writing sample station)	Not available

Table 1 (continued)

Sources, university, country	Number of MMI station	Number of days taken	Times required per station	Competencies assessed
Says F. E. et al., ⁴³ 2013, King Abdulaziz University, KSA	6	—	8–10 min	Personal, professionalism, motivation, moral & bioethics, teamwork & communication skills, behaviors
Tavares W. and Mausz J., ⁴⁴ 2013, Centennial College Simulation Centre, Canada	10	2	8 min (iv), 2 min (break)	Self-awareness, responsibility, communication, ethical & moral judgment, teamwork, conflict resolution, problem solving, critical thinking, management skills
Till H et al., ⁴⁵ 2013, University of Dundee, Scotland	10	10	—	Interpersonal skills & communication (including empathy); logical reasoning & critical thinking; moral and ethical reasoning; motivation & preparation to study medicine; teamwork & leadership; honesty & integrity
Tiller D et al., ⁴⁶ 2013, University of Sydney, Australia	9 (iMMI)	4	7 min (iv), 2 min (break)	Not available
Ahmed A. et al., ⁴⁷ 2014, Dubai Health Authority, UAE	8	2	8 min (iv), 2 min (break)	Responsibility, ethical & moral judgment, communication skills, management skills, problem solving, self-awareness, teamwork, conflict resolution
Andrades M. et al., ⁴⁸ 2014, Aga Khan University, Pakistan	8	—	7 min	Safe doctor, communication skills, professionalism, problem solving, team approach, ethical issues, reasons for selecting family medicine, commitment to the program
Barbour M. E. et al., ⁴⁹ 2014, University of Bristol, UK	10	7	5 min (iv), 2 min (break)	Communication skills, teamwork, work experience, community contributions, reason to study at bristol, reason to study in dentistry, numeracy & data-handling, ethics, research in dentistry, dexterity, professionalism
Callwood A. et al., ⁵⁰ 2014, University in the South East of England, UK	8	—	5	Motivation to become a midwife, awareness of midwifery philosophy and the role of the midwife, respect for difference & diversity, honesty & integrity, kindness, compassion & empathy, intellectual curiosity & reflective nature, advocacy, respect for privacy &

(continued on next page)

Table 1 (continued)

Sources, university, country	Number of MMI station	Number of days taken	Times required per station	Competencies assessed
Eva K. W. and Macala C., ⁵¹ 2014, University of British Columbia, Canada	12	—	8 min (iv), 2 min (1st station), 3 min (break)	dignity, initiative, problem solving & teamwork CanMEDs competencies: medical expert, communicator, collaborator, manager, health advocate, scholar, professional
Hissbach J. C. et al., ⁵² 2014, Hamburg University, Germany	12 (2009) 9 (2010)	1	5 min (iv), 1.5 min (break)	Empathy, communication skills, self-regulation
Hopson L. R. et al., ⁵³ 2014, University of Michigan, US	8	—	—	Adaptability, hardworking problem solving communication skills, teamwork, altruism ethical, aware of issues facing medicine, compassionate, drive to excel
Kelly M. et al., ⁵⁴ 2014, Clinical Science Institute, National University of Ireland, Ireland	10	2	7 min	Irish medical council's eight domains of professional practice: patient safety and quality of patient care, communication & interpersonal skills, collaboration & teamwork, management (including self-management), scholarship, professionalism, clinical skills, relating to patients
Kelly M. E. et al., ⁵⁵ 2014, National University of Ireland Galway, Ireland	10	—	7 min	Not available
Liao SC. et al., ⁵⁶ 2014, National Cheng Kuang University, Taiwan	7	—	—	Empathy; respect for life; crisis management; initiative; insightfulness; integrity, communication skills
Oliver T. et al., ⁵⁷ 2014, Ontario Veterinary College, Canada	8	—	10 min	Ethical & moral, interpersonal, intrapersonal, professional
Oyler D. R. et al., ⁵ 2014, University of Kentucky, US	4	—	7 min	Critical thinking, teamwork, ethical reasoning & integrity, communication & interpersonal skills
Roberts C. et al., ⁵⁸ 2014, University of Sydney, Australia	6	—	8 min (iv), 2 min (break)	Vocation/motivation, communication, organisation/personal management, personal attributes
Sebok S. S. et al., ⁵⁹ 2014, Queen's University, Canada	8	—	—	Communication, critical thinking, maturity, effectiveness, empathy, professionalism, resolution, integrity
Stowe C. D. et al., ⁶⁰ 2014, University of Arkansas for Medical Sciences College of Pharmacy, US	5 (pilot) 3 (full implementation)	1 1	8 min (iv), 2 min (break)	Pilot: critical-thinking, rapport/empathy, ethics/professionalism, knowledge of pharmacy, personal attributes

Table 1 (continued)

Sources, university, country	Number of MMI station	Number of days taken	Times required per station	Competencies assessed
Alweis R. L. et al., ⁶¹ 2015, Northeastern United States Internal Medical Residency, US	6 (5 MMI; 1 TI)	—	8 min (MMI), 16 min (TI), 2 min (break)	Full implementation: rapport/empathy, ethics/professionalism, personal attributes Professionalism, team player, constructive response to stress, capacity for self-reflection, capacity for empathy, adaptability/tolerance of uncertainty, and the ability to incorporate feedback
Burkhardt J. C. et al., ⁶² 2015, University of Michigan, US	8	—	—	Patient care, medical knowledge, diagnostic skills, communication skills, procedural skills, professionalism
Cox W. C. et al., ⁶³ 2015, University of North Carolina, US	7	3	6 min (iv), 2 min (break)	Integrity, adaptability, empathy, critical thinking, teamwork
Sebok S. S and Syer M. D., ⁶⁴ 2015, Canadian Medical School, Canada	8	—	8 min (iv), 2 min (break)	Communication, critical thinking, maturity, effectiveness, empathy, professionalism, resolution, integrity
Yoshimura H. et al., ⁶⁵ 2015, Tokyo Bay Urayasu-Ichikawa Medical Centre and Gifu University, Japan	5	3	10 min (5 min for PBQ; 5 min for SQ), 1 min (break)	Patient care and procedural skills, practice-based learning & improvement, interpersonal & communication skills, professionalism
Abdul Rahim & Yusoff, 2016 ⁶⁶ School of Medical Sciences, Universiti Sains Malaysia, Malaysia	5	3	5 min (with judges) and 2 min (preparation)	Language proficiency, general conduct, critical thinking, ethical awareness, communication skills, knowledge of health care system, standard interview question

MMI is free of gender, age, previous experience, prior knowledge, and cultural bias^{14,19,23,31,48,54,57,60}; MMI is a fair assessment and scoring sheet which allowed them to differentiate between applicants^{5,18–20,47,48,51,52,58,60}; neither aboriginal-specific rater training nor aboriginal rater assignment is required¹¹; violations of MMI security do not unduly influence applicant performance ratings¹²; MMI provides sufficient time for students to present ideas⁴⁸; MMI is at least as cost-efficient as many other personal interview formats,⁵⁰ MMI eases interviewer anxiety associated with having to judge candidates unfavourably⁵¹; and MMI was not stressful.²⁷ Conversely, MMI requires a greater number of rooms⁵⁰; station scores provided by student interviewers were slightly higher than those of faculty member or practitioner interviewers²⁵; student interviewers were less lenient^{26,30} and had more unexpected ratings³⁰; students preferred a mixed format, rather than

MMI alone⁵⁹; cultural specificity of some stations and English-language proficiency were seen to disadvantage international students³⁷; applicants with introverted personalities may fare less well in the MMI process⁶²; and raters were unable to distinguish between the various non-cognitive attributes.⁴⁵ Overall, MMI was consistently judged to be more favourable than unfavourable by both candidates and examiners (Table 2).

Relation to other variables. MMI correlated with OSCE performance^{4,16,66}; MMIs were predictors of success in assessment scores^{66,67}; there was a fair correlation with the Graduate Australian Medical School Admissions Test (GAMSAT) subsection 'Reasoning in Humanities and Social Sciences'¹⁵; MMI measured more variation in non-cognitive traits²¹; no personality variable correlated significantly with the MMI total score⁶³; rural attribute domains were not significant predictors of MMI scores⁵⁷;

Table 2: Data synthesis of the selected studies.

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Eva K. W. et al., ³ 2004	<ul style="list-style-type: none"> • Pilot study design • McMaster University (Canada) • N = 117 • Medical School 	<ul style="list-style-type: none"> • Development of an innovative admission protocol (MMI) that is intended to take advantage of this lesson in the context of student admission 	<ul style="list-style-type: none"> • Both applicants and examiners were positive about the experience and the potential for this protocol • The reliability of the MMI was observed to be 0.65. 	<ul style="list-style-type: none"> • Response process • Internal structure
Eva K. W. et al., ⁴ 2004	<ul style="list-style-type: none"> • Cohort study design* • McMaster University (Canada) • Medical School • N = 45 	<ul style="list-style-type: none"> • Assess the validity of this new admission protocol by examining the relationship between pre-clerkship performance, the MMI and the traditional admission protocol used by the Undergraduate MD program at McMaster University • Evaluation tools used: OSCE, PPI 	<ul style="list-style-type: none"> • MMI was the best predictor of OSCE performance and GPA was the most consistent predictor of performance on multiple-choice question examinations of medical knowledge. 	<ul style="list-style-type: none"> • Relation to other variable
Eva K. W. et al., ¹⁰ 2004	<ul style="list-style-type: none"> • Experimental study design* • McMaster University (Canada) • N = 54 • Health Sciences Faculty 	<ul style="list-style-type: none"> • Assess the consistency of ratings assigned by health sciences faculty members relative to community members during an innovative admissions protocol called the MMI 	<ul style="list-style-type: none"> • Overall test reliability was found to be 0.78 • Study suggested that admissions committees should distribute their resources by increasing the number of interviews to which candidates are exposed rather than increasing the number of interviewers within each interview 	<ul style="list-style-type: none"> • Internal structure
Moreau K. et al., ¹¹ 2005	<ul style="list-style-type: none"> • Observational Study Design* • McMaster University (Canada) • N = 12 applicants (5 aboriginal, 7 non-aboriginal) • Medical School 	<ul style="list-style-type: none"> • Evaluate whether any suggestion of bias existed in application of the MMI in its assessment of aboriginal medical school applicants. 	<ul style="list-style-type: none"> • Recommended that MMI stations be vetted by aboriginally sensitive personnel, but neither aboriginal-specific rater training nor aboriginal rater assignment is required • Interviewer type and interviewee type were both non-significant contributors to the scores observed, $p > 0.2$ and $p > 0.8$ • Overall reliability of this 12-station MMI was 0.70 	<ul style="list-style-type: none"> • Response process • Internal structure
Reiter H. I. et al., ¹² 2006	<ul style="list-style-type: none"> • Case Study Design* • McMaster University (Canada) • N= Study 1 : 57 <p>Study 2 : 384 Study 3 : 38</p> <ul style="list-style-type: none"> • Medical applicant (Study 1 and Study 2); Rehabilitation Sciences applicant (Study 3) 	<ul style="list-style-type: none"> • Determine the impact of security violations on perceived competence levels in 3 different studies (low-stakes in research study, high-stakes in admission, high-stakes in dual application) 	<p><i>Study 1:</i></p> <ul style="list-style-type: none"> • Overall test generalizability of this 9-station MMI with 2 examiners per station was 0.78 <p><i>Study 2:</i></p> <ul style="list-style-type: none"> • Overall test-retest reliability of this 12-station MMI with 1 examiner was 0.70. 	<ul style="list-style-type: none"> • Response process • Internal structure

Brownell K et al., ¹³ 2007	<ul style="list-style-type: none"> • Acceptability and Feasibility design • University of Calgary (Canada) • N = 281 • Medical Admission 	<ul style="list-style-type: none"> • Introducing the MMI as a replacement for the more traditional interview process and its acceptability by applicants and interviewer 	<ul style="list-style-type: none"> • Mean performance on exposed station equaled 4.92 (SD = 1.36), mean performance to which applicants were naive equaled to 4.94 (SD = 0.65): revealed no effect of prior exposure. 	<ul style="list-style-type: none"> • Response process
Harris S. and Owen C., ¹⁴ 2007	<ul style="list-style-type: none"> • Australian National University (Australia) • N = 115 candidates • Medical Admission 	<ul style="list-style-type: none"> • Describe the development and pilot testing of a set of admissions instruments based on the McMaster University MMI. • Designed to assess desirable, non-cognitive characteristic in order to inform final decisions on candidate selection for entry to medical school 	<p><i>Study 3:</i></p> <ul style="list-style-type: none"> • Overall test-retest reliability of this 7-station MMI with 1 examiner per station was 0.70 for OTs and 0.68 for PTs. <p><i>Overall result:</i></p> <ul style="list-style-type: none"> • No statistically significant differences in MMI performances were detected. • Predictable violations of MMI security do not unduly influence applicant performance ratings. • Interviewers indicated that they had adequate time to assess applicant, the MMI was a fair assessment and scoring sheet allowed them to differentiate between applicants. • Applicants indicated that they received well information beforehand for the MMI, MMI was free of gender and cultural bias and there was sufficient time to present ideas at the stations. • 97 candidates were deemed satisfactory and 18 were excluded on the basis of their interview from further consideration for admission • This method proved to be an efficient process to determine suitability of candidates to the course. • For the retained and rejected candidates, they had significantly different total scores and mean scores for each station. 	<ul style="list-style-type: none"> • Response process • Consequences
Lemay J.F. et al., ¹⁵ 2007	<ul style="list-style-type: none"> • Observational Study design* • University of Calgary (Alberta, Canada) • N = 281 • Medicine School • Observational Study design 	<ul style="list-style-type: none"> • Develop an MMI assessment that would measure several non-cognitive attributes 	<ul style="list-style-type: none"> • Cronbach's alpha for the stations ranged from 0.97 to 0.98 • The correlations between stations, based on total scores, ranged from 0.042 to 0.360. 	<ul style="list-style-type: none"> • Internal structure • Consequences

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Reiter H. I. et al., ¹⁶ 2007	<ul style="list-style-type: none"> • Michael G DeGroot School of Medicine at McMaster University • N = 117 candidates for MMI (Spring 2002); 45 enrolled in clerkship evaluation and MCCQE (2005/2006) 	<ul style="list-style-type: none"> • Comparing performance on the various admissions measures used by McMaster University's medical school with in-program performance during clerkship evaluation and national licensing examination. 	<ul style="list-style-type: none"> • Correlations between 5 admission tools and in-program assessment exercise confirmed that: • Only the MMI was statistically predictive of OSCE performance (standardized $\beta = 0.4$, $P < 0.05$) • Only the MMI was statistically predictive of clerkship performance, measured with both the average ratings assigned by clerkship directors (standardized $\beta = 0.7$, $P < 0.001$) and encounter card ratings provided by clinical preceptors (standardized $\beta = 0.5$, $P < 0.01$) • Correlation between each admissions tool and performance on the subscales of the MCCQE Part I revealed that: • Only the MMI was statistically predictive of CLEO or PHELO performance (standardized $\beta > 0.4$, $P < 0.01$) • Only the MMI was predictive of CDM performance (standardized $\beta = 0.35$, $P < 0.05$) • The MMI and uGPA were equally predictive of overall test performance (standardized $\beta > 0.3$, $P < 0.06$) 	
Hofmeister M. et al., ⁶ 2008	<ul style="list-style-type: none"> • University of Alberta, Edmonton, Canada • N = 71 • Physician, medical educators 	<ul style="list-style-type: none"> • Assess the acceptability of the MMI to both international medical graduate (IMG) applicants to family medicine residency training in Alberta, Canada 	<ul style="list-style-type: none"> • 69 applicants responded to applicant acceptability survey. 13 (18.8%) indicated that they had been interviewed for residency training position before and 26 (37.7%) indicated they had not; 30 (43.5%) did not respond. • Five themes were identified: expression of appreciation (n = 16), content (n = 15), format (n = 14), requests for more information on MMI (n = 5) and miscellaneous (n = 5) 	<ul style="list-style-type: none"> • Response process

Humphrey S. et al., ¹⁷ 2008	<ul style="list-style-type: none"> • West Midlands Deanery, UK • N = 96 • Senior House Officers (SHOs) to a UK regional paediatric training programme 	<ul style="list-style-type: none"> • Assess candidates' and interviewers' perceptions of the use of a MMI for selection of senior house officers (SHOs) to a UK regional paediatric training programme 	<ul style="list-style-type: none"> • Both candidates and interviewers were positive about the fairness of MMI (Mean score of 4.0 and 4.4, respectively) • Gender, age and previous experience of MMIs did not account for differences in candidate responses ($P > 0.05$) • Total of 86% candidates were international medical graduates who preferred the format more than UK graduates did ($P = 0.01$) • Interviewers mainly experienced consultants who agreed that the multi-station format was better than traditional interview (mean score 4.8) and represented a reliable process (mean score 4.4). 	<ul style="list-style-type: none"> • Response process • Internal structure
Roberts C. et al., ¹⁸ 2008	<ul style="list-style-type: none"> • University of Sydney • N = 485 candidates; 155 interviewer and 21 questions taken from a pre-prepared bank • Medical candidates 	<ul style="list-style-type: none"> • Establish whether interviewers can make reliable and valid decisions about applicants when selecting candidates for entry to a graduate-entry medical program, using a pre-professionalism framework and the MMI format. • Wanted to know which features of the MMI were most useful in guiding admissions committees to focus their resources in making robust decisions about candidates. 	<ul style="list-style-type: none"> • For a single MMI question and 1 interviewer, 22% of the variance between scores reflected candidate-to-candidate variation. The remaining 78% would reflect unwanted factors. • The generalizability coefficient with this test format was 0.70 • The reliability for an 8-question MMI was 0.7; to achieve 0.8 would require 14 questions. • A disattenuated correlation with the Graduate Australian Medical School Admissions Test (GAM-SAT) subsection 'Reasoning in Humanities and Social Sciences' was 0.26 • The mean inter-question correlation was 0.23 (SD = 0.15), giving a typical range of inter-question correlations of 0.08–0.38 (range – 0.22 to 0.58) 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable
Rosenfeld J. M. et al., ¹⁹ 2008	<ul style="list-style-type: none"> • Observational Study design* • McMaster University (Canada) • N = 400 applicant for MMI • N = 400 applicant for Traditional Interview • Medical School 	<ul style="list-style-type: none"> • Determine the cost to ensure that the tools can be feasibly implemented • Provide answers to frequently asked questions regarding the logistics involved in mounting a MMI 	<ul style="list-style-type: none"> • MMI: 72 h of creation time for 24 MMI Stations, \$50 per station = \$1200 • MMI: N = 192 (96 per each of 2 days of interviews) thus 66.7 h per 400 candidates 	<ul style="list-style-type: none"> • Response process

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Eva K. W. et al., (20) 2009	<ul style="list-style-type: none"> • McMaster University, Canada • Undergraduate N = 34 • Postgraduate N = 22 • Medical School 	<ul style="list-style-type: none"> • Tests of the validity of the MMI selection process, comparing MMI scores with those achieved on a national high-stakes clinical skills examination • Explore the stability of candidate performance and the extent to which so-called 'cognitive' and 'non-cognitive' qualities should be deemed independent of one another 	<ul style="list-style-type: none"> • TI: N = 192 (48 per each of 4 days of interviews), thus 400 h per 400 candidates • MMI: Admissions coordinator 154 to 278 staff hours (11–19 staff spread over 2 days of interview) and 8 actors for each 2 days • MMI: 8 clinics with 12 rooms provided in kind 4800 score sheets (\$240) • TI: 16 clinics rooms provided in kind 1200 score sheets (\$60) <p><i>Overall:</i></p> <ul style="list-style-type: none"> • MMI is at least as cost efficient as many other personal interviews format even though MMI's disadvantage is the requirement of a greater number of rooms. • Median reliability of eight administrations of the MMI in various cohorts was 0.73 • Correlation between performance on them MMI and number of stations passes on an objective structured clinical examination-based licensing examination was $r = 0.43$ ($P < 0.05$) in postgraduate sample and $r = 0.35$ ($P < 0.05$) in undergraduate sample • Correlation between 'cognitive' and 'non-cognitive' assessment instruments increased with time in training 	<ul style="list-style-type: none"> • Internal structure • Relation with other variable
Hofmeister M. et al., (21) 2009	<ul style="list-style-type: none"> • University of Calgary and University of Alberta • N = 71 International medical graduate • Family Medicine Residency 	<ul style="list-style-type: none"> • Develop and assess the evidence for the validity and reliability of the MMI in the assessment of professionalism in IMG candidates for family medicine residency education at the Universities of Calgary and Alberta. 	<ul style="list-style-type: none"> • The reliability as indicated by the generalizability coefficient associated with average station scores was 0.70 with one interviewer per station • There were no statistically significant differences in total MMI scores or mean station sum scores based on session, track, applicant age, gender, years since medical 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable

Kumar K et al., (22)
2009

- University of Sydney, Australia (6 Days), University of British Columbia, Vancouver, Canada (2 Days)
- N = 442 Candidate; 75 Interviewer
- Medical selection

- Reports on a qualitative analysis of participants' experiences and perceptions of a high-stakes process for selection into graduate-entry medical school

- school completion or language of medical school.
- There were low, non-significant correlations with OSCE overall ($r = 0.15$), MCCEE ($r = 0.01$) and MCCQE I ($r = 0.06$) scores and a higher non-significant correlation with MCCQE II scores ($r = 0.33$)

6 major and sub-themes pertaining to participants' experiences:

- One-to-one interviews
- Improved the quality of candidates –interviewer interaction
- Reduced the stress
- Multiple assessment opportunities
- Ease interviewer anxiety associated with having to judge candidates unfavorably (I)
- Gave better chance of discriminating amongst candidates (I)
- Gave candidates chance to redeem a 'bad first impression with one person' and regain their composure (C)
- Standardized scenario-based interviews
- Make it harder for rehearsal and coaching of responses and thus gave insight into the 'genuine' ability of candidates (Candidates)
- Potentially reduced interviewer subjectivity (Interviewers)
- The mini-interview
- Short time (C)
- Limits the quality of interaction within the interview and lacks opportunity to impress the interviewers (C)
- What is being measured?
- Assessed 'reasoning skills and the ability to unpick a dilemma/situation' (C)
- Assessing communication skills and were concerned that decisions based on social interaction skills rather than the reasoning capabilities of the candidate (I)
- What else should be measured?

- Response process

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Razack S. et al., (23) 2009	<ul style="list-style-type: none"> • Pilot Study design* • McGill University Medical School, Canada • Faculty of Medicine • N = 100 	<ul style="list-style-type: none"> • Assess the acceptability of the MMI in our context and to compare this new tool with our current selection tools, including the traditional interview 	<ul style="list-style-type: none"> • Lack of opportunity to discuss specific personal qualities of candidates (C) • Include one station to explore the candidate's commitment to medicine and to university which they had chosen (C) • MMI rated more highly than traditional interview on fairness, imposition of stress and effectiveness as a measurement tool. • MMI also i) allowed them to be competitive; ii) was enjoyable and iii) was often a favourite part of their interview experience. 	<ul style="list-style-type: none"> • Response process
Roberts C. et al., (24) 2009	<ul style="list-style-type: none"> • University of Sydney, Australia • Medicine and Dentistry Admission • N = 207 Interviewers; 686 candidates 	<ul style="list-style-type: none"> • Establish the conceptual equivalence (concept in this case being entry-level reasoning skills in professionalism) • Investigate whether there were any systematic differences in outcome among equally able candidates from different subgroups caused by DIF • 4-facet Rasch Model 	<ul style="list-style-type: none"> • SD by facets were: interviewer stringency or leniency, 0.52; candidate ability, 0.75; and MMI question difficulty, 0.27. • The spread of candidates was 1.44 times that of the interviewers, therefore variance was 2.08 times that of the interviewers • All the questions were well within the predetermined range of 0.5–1.7 • All 39 questions had a good fit to the IRT model • Of 195 checklist items, none were found to have significant DIF after visual inspection of expected score curves. 	<ul style="list-style-type: none"> • Response process • Internal structure
Dore K. L. et al., (25) 2010	<ul style="list-style-type: none"> • Observational Study design* • McMaster University: obstetrics-gynecology and pediatrics • University of Alberta (internal medicine) • N = 484 candidates (across 2 years) 	<ul style="list-style-type: none"> • Investigates the reliability and acceptability of the MMI at the postgraduate level, given the differences in the applicant pool and process 	<ul style="list-style-type: none"> • Overall reliability of the 7-station MMI across all years and program was at least moderately acceptable, 0.55–0.72 • This 7-station were used for feasibility and until acceptability was established • Estimating overall reliability for 10-station, the range of reliability would increase to 0.64–0.79 • 88% candidates believed they could accurately portray themselves during the MMI 	<ul style="list-style-type: none"> • Response process • Internal structure

Kulasegaram K. et al., (26) 2010	<ul style="list-style-type: none"> • Observational Study design* • Michael G. DeGroote School of Medicine, McMaster, Hamilton, Canada • N = 152 applicants • Medical School 	<ul style="list-style-type: none"> • Examine the possible association between personality and non-cognitive performance since there is ample evidence that the MMI is a valid predictor of non-cognitive outcomes in medical school 	<ul style="list-style-type: none"> • 77% indicated that specialized medical knowledge was not needed to complete the stations. • While 90% of interviewers believed they could reasonably judge candidates' abilities. • Significant differences in GPA scores in favour of the non-volunteers, [$F = 15.68$ ($p < 0.0001$)], the difference was only 0.05 (is not sufficient to bias computed correlation). • Correlation among the subscores of the Neo-5: ranged from -0.40 to $+0.30$ (Mean = 0.035) with the largest negative correlation between neuroticism and conscientiousness • No personality variable correlated significantly with MMI total score. The mean correlation was 0.09 and the largest (non-significant) correlation was 0.20. 	<ul style="list-style-type: none"> • Relation to other variable
Roberts C. et al., (27) 2010	<ul style="list-style-type: none"> • Historical Study design* • University of Sydney, Australia • N = 207 interviewer; 686 candidates • Medicine and Dentistry Admission • Using the same candidates and interviewer in Questions Bank in MMI article (Roberts C. et al., 2009) 	<ul style="list-style-type: none"> • Focuses on the performance of interviewers and how this impacts on the reliability and acceptability of the MMI, and discusses possible strategies to moderate any negative influences • Many-Facet Rasch Model (MFRM) has the capability to both identify these sources of error and partially adjust for them within a measurement model that may be fairer to the candidate • FACETS Software (Many-Facet Rasch Model) 	<ul style="list-style-type: none"> • Decision of 207 interviewers had an acceptable fit to the MFRM model. • Interviewer stringency/leniency and question difficulty were anchored by the measurement model at 0.00 logits and candidate ability was allowed to float • Interviewers are more variable than MMI questions and the spread of interviewers is nearly 3.5 times that of MMI questions. • Reliability of separation index in terms of stringency/leniency gave a reliability of 0.91 • High reliability indicates that the interviewers are meaningfully separated according to their levels of severity from lenient to stringent with a high degree of confidence • Item reliability for four behavioral interviews was 0.64 compared to 0.77 for ten MMI stations • Person reliability was 0.72 for 85 non-extreme scoring persons 	<ul style="list-style-type: none"> • Response process • Internal structure
Jones, P. E. & Forister, J. G., (28) 2011	<ul style="list-style-type: none"> • Cohort study design • McMaster University (Canada) • N = 176 (N = 93: Behavioral Interview Format, N = 83: MMI) • Physician Assistant Program 	<ul style="list-style-type: none"> • Compare the experiences with behavioral-based and MMI outcomes with two consecutive cohorts of applicants at one public US PA program. • Rasch Analysis Software 	<ul style="list-style-type: none"> • Item reliability for four behavioral interviews was 0.64 compared to 0.77 for ten MMI stations • Person reliability was 0.72 for 85 non-extreme scoring persons 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
O'Brien A. et al., (29) 2011	<ul style="list-style-type: none"> • Pilot Study design** • St George's University of London, UK • Undergraduate streams for medicine • N = 47 candidates <p>(MBBS 5–21; MBBS 4–26)</p>	<ul style="list-style-type: none"> • Design a modified version of an MMI and to evaluate its potential reliability, feasibility and acceptability, with a view to assessing whether such an approach could be justified at St George's • Assess its concurrent validity compared to the standard interviews (Sis) 	<p>behavioral interview as compared to 0.57 for 83 in MMI</p> <ul style="list-style-type: none"> • BI and MMI had similar model fit. • BI did not adequately measure differences in applicant characteristics but MMI measured more variation in non-cognitive traits. • No difference in performance between 5-year streams and 4-years stream: • MBBS 5 mean score was 3.75 (SD 0.54) • MBBS 4 mean score was 3.73 (SD 0.74) • No difference for the interview scores: MBBS 5 mean 3.51 (SD 0.52) and MBBS 4 mean 3.56 (SD 0.57) • MBBS 4 candidates, Cronbach's alpha for MMI was 0.69 • MBBS 5 candidates, Cronbach's alpha was 0.73 • MBBS 4 applicants performed just as well on the MMI as they did on the traditional interview, with MBBS 5 applicants performing better on the MMI 	<ul style="list-style-type: none"> • Internal structure
Uijtdehaage S. et al., (30) 2011	<ul style="list-style-type: none"> • Observational Study design* • David Geffen School of Medicine, University of California, Los Angeles • N = 76 applicants (2009); N = 78 applicants (2010) • Medical undergraduate 	<ul style="list-style-type: none"> • Determine the reliability of the MMI, potential bias in scores, and the degree of acceptance by the interviewers and applicants 	<ul style="list-style-type: none"> • 2010 cohort rated the MMI process as less stressful (average 3.7 in 2010 versus 4.2 in 2009) • The preliminary reliability of the MMI in 2009 was 0.58 – lower than reported in previous studies – but improved in 2010 to 0.71 after an easy station was replaced with a more challenging. • Applicants indicated that MMI process was free from cultural bias (average 6.3) or gender bias (average 6.6) in 2009 	<ul style="list-style-type: none"> • Response process • Internal structure
Yen W. et al., (31) 2011	<ul style="list-style-type: none"> • Michener Institute for Applied Health Sciences, Toronto, Canada • N = 196 candidates • Health Science 	<ul style="list-style-type: none"> • Investigate the potential for the EQ-i to serve as a proxy measure to the MMI • Bar-On EQ-i emotional intelligent instrument 	<ul style="list-style-type: none"> • MMI had a moderate reliability estimate of 0.75, and EQ-i had a lower estimate of 0.65 • MMI total score and the EQ-i total score were not found to be 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable

Cameron A. J. and MacKeigan L. D., (32) 2012	<ul style="list-style-type: none"> • Pilot Study design • Leslie Dan Faculty of Pharmacy (LDFP), University of Toronto • N = 30 Candidates • Pharmacy 	<ul style="list-style-type: none"> • Determine specific nonacademic attributes to be assessed in an MMI designed for admission to the Leslie Dan Faculty of Pharmacy • Assess the feasibility (resources and procedures) and acceptability of the MMI to candidates (interviewees) and interviewers • Determine optimal station duration • Assess the discriminant validity of the MMI • Assess the reliability of the overall MMI score 	<p>significantly correlated ($p = 0.14$) nor were relationships found at the subscale level ($p > 0.003$)</p> <ul style="list-style-type: none"> • Correlational analysis suggests that a relationship does not exist between the MMI and the EQ-i • The ICC for the overall 10-station score was 0.77 (95% CI, 0.63–0.88); for five 6-min stations was 0.66 (95% CI, 0.42–0.82) • The Pearson r coefficient for MMI and PPA was -0.025 • MMI and PCAT composite score was 0.042 • Coefficient for PPA and PCAT composite was 0.370 ($p = 0.048$) • Station scores provided by student interviewers were slightly higher than those of faculty member or practitioner interviewers • The ICC for the MMI was 0.77 and correlations with PPA and PCAT composite were negligible 	<ul style="list-style-type: none"> • Response process • Internal structure • Relation to other variable
Dowell J. et al., (33) 2012	<ul style="list-style-type: none"> • Dundee University, UK • Dundee Medical School • N = 452 (2009); 477 (2010) 	<ul style="list-style-type: none"> • Describe Dundee medical school's progression from a traditional interview to a full-scale MMI, psychometric properties, views of applicants and assessors and to identify areas of further research 	<ul style="list-style-type: none"> • Questionnaire confirmed that the process was acceptable to all parties • Cronbach's alpha reliability was satisfactory and consistent • The range of correlations between station scores for 2009 and 2010 were 0.057–0.363 and -0.061 to 0.308, respectively • Using MFRM fair scores would change the selection outcome of 6.2% and 9.6% mod candidates in 2009 and 2010 respectively. • Student was less lenient, made more use of the full range of the rating scales and were just as reliable as staff • Candidates accepted by the MMI-admissions process had higher scores in the Canadian national licensing examinations than those who were rejected for: • Part I (mean total score, 531 [95% CI, 524–537] vs. 515 [95% CI, 507–522]; $P = 0.003$) 	<ul style="list-style-type: none"> • Response process • Internal structure
Eva K. W. et al., (34) 2012	<ul style="list-style-type: none"> • Cohort Study design • McMaster University (using MMI in 2004 or 2005) • N = 751 (Part I); N = 623 (Part II) • Medical school admission 	<ul style="list-style-type: none"> • Determine whether students deemed acceptable through a revised admission protocol using 12-station MMI outperform others on the 2 parts of the Canadian national licensing examination (MCCQE) 	<ul style="list-style-type: none"> • Candidates accepted by the MMI-admissions process had higher scores in the Canadian national licensing examinations than those who were rejected for: • Part I (mean total score, 531 [95% CI, 524–537] vs. 515 [95% CI, 507–522]; $P = 0.003$) 	<ul style="list-style-type: none"> • Consequences

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Griffin B. and Wilson I., (35) 2012	<ul style="list-style-type: none"> • Cross-sectional Study design* • Australian School of Medicine • 2006: N = 364 • 2007: N = 336 • 2008: N = 337 • Medical School 	<ul style="list-style-type: none"> • Assess how MMI scores relate to both the five factors and their associated facets. • Using UMAT (a cognitive ability test used by medical schools across Australia and New Zealand); UAI (a percentile score based on high school academic performance and similar construct to GPA) 	<ul style="list-style-type: none"> • Part II (mean total score, 563 [95% CI, 556–570] vs. 544 [95% CI, 534–554]; $P = 0.007$) • Those who matriculated at McMaster did not outperform those who matriculated elsewhere for: <ul style="list-style-type: none"> • Part I (mean total score, 524 [95% CI, 515–533] vs. 546 [95% CI, 535–557]; $P = 0.004$) • Part II (mean total score, 557 [95% CI, 548–566] vs. 582 [95% CI, 569–594]; $P = 0.003$) • Extraversion and conscientiousness were correlated with MMI scores in all 3 years, and agreeableness was significantly related in 2 of the 3 years ($p < 0.05$) • 4 of 6 facets of extraversion and conscientiousness were consistently correlated with MMI scores, but 2 facets (excitement seeking and cheerfulness under extraversion, and orderliness and dutifulness under conscientious) were unrelated. • Suggest that MMIs and personality questionnaires are not interchangeable so we would therefore not recommend that personality tests to be used as a screening tool for MMIs. 	<ul style="list-style-type: none"> • Relation to other variable
Jerant A. et al., (36) 2012	<ul style="list-style-type: none"> • Observational Study design • University of California, Davis (UCD), California • N = 444 candidates • School of Medicine 	<ul style="list-style-type: none"> • Examine relationship among applicant personality, MMI performance and medical school acceptance offers 	<ul style="list-style-type: none"> • Those with extraversion scores (among 444 candidates) in the top (versus bottom) quartile had significantly higher MMI scores (adjusted parameter estimate = 5.93 higher, 95% CI: 4.27–7.59; $P < 0.01$) • In a model excluding MMI score, top (versus bottom) quartile agreeableness (AOR = 3.22; 95% CI 1.57–6.58; $P < 0.01$) and extraversion (AOR = 3.61; 95% CI 	<ul style="list-style-type: none"> • Relation to other variable

<p>McAndrew R. and Ellis J., (37) 2012</p>	<ul style="list-style-type: none"> • Observational Study design • Cardiff University, UK • N = 190 Candidates • Cardiff Dental School 	<ul style="list-style-type: none"> • Evaluate the MMI process as part of the admissions procedure for dental school • Gain a greater insight into the applicants' and participations' experiences of the MMI process and inform future research in this field 	<p>1.91–6.82; $P < 0.01$) were associated with acceptance offers.</p> <ul style="list-style-type: none"> • After adding MMI score to the model, high agreeableness (AOR = 4.77; 95% CI 1.95–11.65; $P < 0.01$) and MMI score (AOR 1.33; 95% CI 1.26–1.42; $P < 0.01$) were associated with acceptance offers. • 127 (67%) felt they had performed well with 24 (19.3%) unsure and 36 (12.8%) saying their performance was not ideal • MMI experience was considered favourable and from 137 written comments received by candidates, 39 were most positive with reference to the MMIs 	<ul style="list-style-type: none"> • Response process
<p>Reiter H. I. et al., (67) 2012</p>	<ul style="list-style-type: none"> • 6 schools that is: McMaster University, University of Saskatchewan, University of Calgary, University of British Columbia, Dalhousie University, and University of Alberta • Data for admissions ending in 2008 and 2009. • N = 5253 (across six schools) • Medical students 	<ul style="list-style-type: none"> • Investigates whether the MMI is diversity-neutral and if so, whether applying it with greater weight would dilute the anticipated negative impact of diversity-limiting admissions measures. 	<ul style="list-style-type: none"> • There was a positive correlation between MMI scores and age ($P < 0.05$); the correlation was neutral with respect to gender, size of community of origin and income level, and there was a negative correlation ($P < 0.05$) for those with self-declared aboriginal status. • This means that MMI scores were unrelated to gender, size of community of origin and income level. They correlated positively with age and negatively with aboriginal status. • GPA and MCAT correlated negatively with age and aboriginal status, GPA correlated positively with income level, and MCAT correlated positively with size of community of origin. 	<ul style="list-style-type: none"> • Response process • Relation to other variable
<p>Fraga J. D. et al., (38) 2013</p>	<ul style="list-style-type: none"> • Retrospective Cohort study • Reading Hospital Institutional Review Board • N = 237 applicants; 17 faculty interviewer • Internal medical residency 	<ul style="list-style-type: none"> • Investigate reliability and acceptability of five-station MMI model for resident selection into an internal medicine residency program in the United States 	<ul style="list-style-type: none"> • G-coefficients are reported for each station and ranged from a minimum of 0.97 for the Last Call station to a high of 0.98 for Overloaded Census station • G-coefficient for combinations of candidate within station = 0.96 • G-coefficient for station within interviewer = 0.98 	<ul style="list-style-type: none"> • Response process • Internal structure

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Husbands A. & Dowell J., (39) 2013	<ul style="list-style-type: none"> • Cohort Study design* • Dundee Medical School, UK • N = 140 (1st year students; 128 (2nd students) in 2009, N = 150 (1st year students) in 2010 	<ul style="list-style-type: none"> • Address which aspects of the selection process can be justified in terms of predictive validity of knowledge-based and OSCE examinations in early medical school • Establish if MMIs are useful in the UK 	<ul style="list-style-type: none"> • G-coefficient for interviewer within candidate = 0.95 • Applicants indicated that they agreed with the statements that 'the MMI was fair' more strongly than with 'a traditional interview is fair' (5.12 vs. 4.07, $p < 0.001$) • Applicants had higher agreement that 'the MMI is effective at evaluating non-cognitive skills' than for 'the traditional interview is effective evaluating non-cognitive skills' (5.05 vs. 3.41, $p < 0.001$) • No difference in perceived stressfulness of the MMI compared to the traditional interview (3.06 vs. 3.18, $p = 0.32$) • Generalizability data showed that even with only five stations, the reliability of the process was high enough for high-stakes decisions such as admissions (>0.9 for candidate within station) 	<ul style="list-style-type: none"> • Relation to other variable • Consequences
McAndrew R. and Ellis, J. (40) 2013	<ul style="list-style-type: none"> • Observational Study design* • Cardiff University Dental Hospital, UK • N = 235 (BDS interviewees); N = 62 (Therapy and hygiene interviewees) 	<ul style="list-style-type: none"> • Identify any new themes for consideration during the MMI process highlighted by a post-interview evaluation questionnaire in a new cohort of undergraduate dental students and undergraduate dental care professionals (hygienist and therapist) 	<ul style="list-style-type: none"> • MMI scores significantly correlated with six of 10 examination sittings, with magnitudes ranging from 0.24 to 0.50 (unrestricted), accounting for between 5.70% and 25.00% of variance in students' examination scores. • Multiple regressions also confirmed that the MMIs remained the most consistent predictor of success, accounting between 5% and 17% of the variance in assessment scores alone or in combination with candidates' gender. • Binomial statistically calculated and chi-squared tests of independence of categorical variables on nominally scaled data revealed statistically significant differences ($p > 0.001$) and both groups of students were positive 	<ul style="list-style-type: none"> • Response process

Perkins A. et al., (41) 2013	<ul style="list-style-type: none"> • Kingston University and St George's University of London, UK • N = 890 candidates • Nursing Programme 	<ul style="list-style-type: none"> • Compare and contrast the findings between groups • Further detailing and quantification of applicants' perception towards the MMI as a selection tools by analyzing the responses from the different applicants cohorts • Assist in identifying those candidates who demonstrate the potential to achieve the Nursing and Midwifery Council (NMC) domains 	<p>regarding the selection process ($p > 0.001$)</p> <ul style="list-style-type: none"> • 54 separate free text comments recorded but qualitative analysis failed to identify any themes that had not been previously identified. • Positive response from candidates with 65% replying that it was "a better experience" compared with traditional interviews. • Interviewers likewise responded positively with 71% noting "a better experience" • Unsolicited feedback indicated that some would have preferred to have had greater opportunity to discuss nursing issues with their interviewees. • Rural high school graduates scored significantly lower (mean of 4.4 on scale of 1–7; $p, 0.05$) than urban high school graduates (4.6) • Among rural-attribute domains, those with rural community service alone had the highest MMI scores (4.9) while those with rural connections alone had the lowest scores (4.3; $p = 0.016$) • GPA (mean of 4.0) of rural high school graduates was similar to that of graduates of urban high schools (4.1) • GPA and MCAT scores in a multiple linear regression model, rural-attribute domains were not significant predictors of an applicant's MMI score • Percentage of men students performed well in station (personal character) about 95% • Percentage for women is about 90% were excellent, 9% were acceptable and about 1% was poor. • 94% and 71% of women and men students, were excellent in team 	<ul style="list-style-type: none"> • Response process
Raghavan M. et al., (42) 2013	<ul style="list-style-type: none"> • Observational Study design* • University of Manitoba, Canada • Faculty of Medicine • N = 1257 applicants (2008–2011) 	<ul style="list-style-type: none"> • Seek an association between MMI scores and applicants' place of high school graduation • Find associations between MMI scores and applicant attributes in the following three domains: rural connections, employment in rural areas and rural community service such as volunteer and leadership activities • Understand MMI performance of applicants from Aboriginal communities 	<ul style="list-style-type: none"> • Evaluate the applicant's non-cognitive strength and area for improvement • Evaluate the applicant's preparation for medical school <ul style="list-style-type: none"> • identify the outstanding students to be selected for different faculty committees, students support unit, as their skills will be looked after 	<ul style="list-style-type: none"> • Response process • Relation to other variable
Says F. E. et al., (43) 2013	<ul style="list-style-type: none"> • Observational Study design* • King Abdulaziz University, KSA • N = 352 participant (Men = 174; women = 178) • Faculty of Medicine 	<ul style="list-style-type: none"> • Evaluate the applicant's non-cognitive strength and area for improvement • Evaluate the applicant's preparation for medical school <ul style="list-style-type: none"> • identify the outstanding students to be selected for different faculty committees, students support unit, as their skills will be looked after 	<ul style="list-style-type: none"> • Response process • Internal structure 	

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
		<p>and developed more through courses and workshops</p> <ul style="list-style-type: none"> Identify below-average students who will receive the same attention in order to enhance and develop their non-cognitive skills 	<p>work and communication skills respectively</p> <ul style="list-style-type: none"> 94% and 73% of women and men students, were excellent in behaviors respectively Reliability analysis using the Cronbach alpha revealed moderate reliability index (0.38) among the performance results of men, whereas women is 0.76 (high reliability index) The reliability index among whole student performance was 0.61 	
Tavares W. and Mausz J., (44) 2013	<ul style="list-style-type: none"> Observational Study design* Centennial College Simulation Centre (Toronto, Ontario, Canada) N = 30 candidates Paramedicine 	<ul style="list-style-type: none"> Seek evidence of reliability and validity for the assessment of non-clinical attributes using the MMI in paramedic contexts Explore the association between non-clinical attributes and clinical skills in paramedicine 	<ul style="list-style-type: none"> The satisfactory level for reliability index was 0.70 Inter-station reliability for the MMI reached 0.77 Pearson's correlations between the overall MMI score and mean Simulation-Based Assessment (SBA) global rating scores reached $r = 0.31$ ($r = 0.48$) and ranged by dimension from $r = -0.11$ (-0.17) (procedural skills) to $r = 0.54$ ($r = 0.83$) (communication) 	<ul style="list-style-type: none"> Internal structure Relation to other variable
Till H. et al., (45) 2013	<ul style="list-style-type: none"> Observational Study design* University of Dundee Medical School N = 452 candidate assessed in 2009 	<ul style="list-style-type: none"> Explore whether the MMI as used at the University of Dundee Medical School is an effective instrument that can reliably separate the more able candidates from the less able candidates Explore whether the three groups of examiners who carried out the assessment process exhibited any systematic differences in their rating patterns Multifaceted Rasch Modeling 	<ul style="list-style-type: none"> MMI reliably (0.89) separated the candidates into four statistically distinct levels of non-cognitive ability. The Rasch measures accounted for 31.69% of the total variance in the ratings (candidates 16.01%, examiners 11.32% and stations 4.36%) Students rated more severely than staff and also had more unexpected ratings. Adjusting scores for examiner severity/leniency and station difficulty would have changed the selection outcomes for 9.6% of the candidates. 	<ul style="list-style-type: none"> Response process Internal structure
Tiller D. et al., (46) 2013	<ul style="list-style-type: none"> Cohort Study design* University of Sydney 	<ul style="list-style-type: none"> Determine whether the scores derived from the iMMI process were both equivalent to and as 	<ul style="list-style-type: none"> Comparison of MMI and iMMI scores for international candidates in 2009 and 2011, respectively, was 	<ul style="list-style-type: none"> Response process Internal structure

	<ul style="list-style-type: none"> • N = 293 international candidates (iMMI) • N = 571 local candidates (MMI) • Medical and dentistry 	<ul style="list-style-type: none"> • reliable as those scores derived from the in-person MMI • Wants to describe the feasibility, acceptability and cost-effectiveness of the iMMI • iMMI using Skype 	<ul style="list-style-type: none"> • no significant in terms of variance ($p = 0.338$) • No significant difference between the MMI scores for local and international candidates in 2011 ($p > 0.05$); the MMI scores for international candidates had greater variation ($p < 0.01$) • The reliability of the nine-question iMMI was 0.76 and for MMI was 0.70 • Interviewer satisfaction (iMMI circuits and the technology): rated highly with 61 (78%) giving the two most positive ratings for the interview process, 71% for satisfaction with the technology • Estimated administrative cost of conducting the iMMIs was AU\$10 145, representing a saving of AU\$51 742 or 84% from estimated cost when conducting in Vancouver in 2009 (AU\$61 887) 	
Ahmed A. et al., (47) 2014	<ul style="list-style-type: none"> • Observational Study • Dubai Health Authority, United Arab Emirates • N = 187 candidates • Dubai Residency Training Program • Family Medicine 	<ul style="list-style-type: none"> • Explore the implementation of MMI in different cultural settings 	<ul style="list-style-type: none"> • A MMI with 8 stations, produced absolute and relative reliability of 0.8 and 0.81, respectively 	<ul style="list-style-type: none"> • Internal structure
Andrades M. et al., (48) 2014	<ul style="list-style-type: none"> • Aga Khan University, Pakistan • N = 16 (interviews); N = 14 (MMI) • Family Medicine Residency Programme 	<ul style="list-style-type: none"> • Compare MMI versus semi-structured interviews for assessing non-cognitive domains in the selection of Family Medicine residents. • Determine perceptions of the interviewers and candidates for the acceptability and feasibility of MMI as a selection tool 	<ul style="list-style-type: none"> • Majority (87%) of the interviewers believed that they were able to get an accurate portrayal of the candidates. • 50% (5 out of 8) interviewers were unsure of the feasibility of conducting an MMI compared to the interviews. • All but one of the interviewers thought that interviews can be replaced by MMI. 	<ul style="list-style-type: none"> • Response process
Barbour M. E. et al., (49) 2014	<ul style="list-style-type: none"> • Observational Study design* • University of Bristol, United Kingdom • Dentistry and medicine programs 	<ul style="list-style-type: none"> • Establish whether starting station influenced total score • Determine whether gender influenced total score or MMI outcome • Assess candidate and interviewer acceptance of MMI 	<ul style="list-style-type: none"> • MMI provided an efficient means to discriminate between the performance of applicants who were all academically highly qualified, with total scores ranging from 35% to 87% of the maximum possible score 	<ul style="list-style-type: none"> • Response process • Internal structure

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
	<ul style="list-style-type: none"> • N = 414 candidates (213 candidate in 2011–2012; 201 in 2012–2013) 		<ul style="list-style-type: none"> • Female candidates performed significantly better than male candidates when assessed by total score ($p = 0.011$; mean score 94.4 for female applicants and 91.9 for male applicants) and by outcome (offer/decline; $p = 0.016$; 58.6% of female and 46.4% of male interviewees received an offer of study following interview) • Stakeholder acceptance was good, with 75% of candidates and 95% of assessors preferring MMI over traditional interviews 	
Callwood A. et al., (50) 2014	<ul style="list-style-type: none"> • Mixed Method Study design* • University in the South East of England, UK • N = 62 participants • Midwifery Studies programme 	<ul style="list-style-type: none"> • Develop, pilot and examine the reliability of MMIs in pre-registration student midwife selection in the UK 	<ul style="list-style-type: none"> • Station specific attributes assessed (with Cronbach's alpha scores) included: compassion and empathy (0.93); respect for difference and diversity (0.96); honesty and integrity (0.97); intellectual curiosity and reflective nature (0.94); advocacy (0.91); respect for privacy and dignity (0.95); team working and initiative (0.96); the role of the midwife and motivation to become a midwife (0.95) • Cronbach's alpha scores measuring internal consistency ranged from 0.91 to 0.97 	<ul style="list-style-type: none"> • Internal structure
Eva K. W. and Macala C., (51) 2014	<ul style="list-style-type: none"> • Observational Study design* • University of British Columbia, Vancouver, Canada • N = 41 medical school candidates • Faculty of Medicine 	<ul style="list-style-type: none"> • Explore the impact of MMI station design on the utility of the process for making selection decisions 	<ul style="list-style-type: none"> • Behavioral interview (BI) stations more reliably differentiated between candidates ($G = 0.77$) than did the other station types (SJ, $G = 0.69$; FF, $G = 0.66$) • The correlation between actual MMI scores and BI stations was also greatest (BI, $r = 0.57$; SJ, $r = 0.45$; FF, $r = 0.42$) • Candidates considered the FF stations to be more challenging and more anxiety-provoking than either the SJ or BJ stations 	<ul style="list-style-type: none"> • Response process • Internal structure
Hissbach J. C. et al., (52) 2014	<ul style="list-style-type: none"> • Pilot test in 2009 • Hamburg University, Germany • N = 80 (2009); N = 200 (2010) 	<ul style="list-style-type: none"> • Improve reliability and reduce costs of the subsequent procedure in 2010 	<ul style="list-style-type: none"> • The mean inter-station correlation in 2009 was 0.20 (min: -0.11; 	<ul style="list-style-type: none"> • Response process • Internal structure

	<ul style="list-style-type: none"> • Medical school admission 		<ul style="list-style-type: none"> max: 0.44) and 0.19 in 2010 (min: 0.07; max: 0.32) • The overall reliability of the initial 2009 HAM-Int procedure with 12 station and an average of 2.33 raters per station was ICC = 0.75 • In 2010 the ICC remained stable at 0.76, despite the reduction of the process to nine stations and 2.17 raters per station • Costs were cut down from \$915 to \$495 per candidate. • With the 2010 modalities, we could have reached an ICC of 0.80 with 16 single rater stations (\$570 per candidate). 	
Hopson L. R. et al., (53) 2014	<ul style="list-style-type: none"> • University of Michigan, Ann Arbor, Michigan • N = 71 participant • Emergency Medicine Resident Selection 	<ul style="list-style-type: none"> • Determine the perspective of participants on the MMI in a stimulated EM interview situation • Determine if the MMI would provide new information to assess residency candidates 	<ul style="list-style-type: none"> • Participant responses showed a negative effect on a decision to interview at this program on both pre (mean 2.7, SD 0.7) and post (mean 2.8, SD 0.9) • EM grades correlated with MMI performance ($F[1, 66] = 4.18$; $p < 0.05$) with honors students having higher scores • Participants preferred a traditional interview (mean difference = 1.36; $p < 0.01$) • Mixed format (traditional interview and MMI) was preferred over a MMI alone (mean difference = 1.1; $p < 0.01$) 	<ul style="list-style-type: none"> • Response process • Relation to other variable
Joshi N. K. et al., (68) 2014	<ul style="list-style-type: none"> • Media social such as Twitter, ALiEM Blog, YouTube • Data is taken for the first 14 days of the event • Able to attract 712 unique readers from 41 countries 	<ul style="list-style-type: none"> • Organize and summarize the responses from the global education community and propose potential solutions and recommendations 	<ul style="list-style-type: none"> • Non-cognitive abilities are already adequately elicited with the current interview process • Several comments centered on the time and resources that are required to recruit and train MMI facilitators, select and execute the MMI stations and assess applicant performance. • Contributors' comments were consistent in expressing the importance of an interview process that is fair, unbiased, equitable and reliable, and elicits information about non-cognitive qualities. 	<ul style="list-style-type: none"> • Response process
	<ul style="list-style-type: none"> • Mixed method study design 			<ul style="list-style-type: none"> • Response process <p>(continued on next page)</p>

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Kelly M. E. et al., (55) 2014	<ul style="list-style-type: none"> • National University of Ireland Galway • N = 109 participate (EU origin = 64, Non-EU = 45) • International Medical students 	<p>Aims of this study were to run an experimental MMI in an internationally diverse student population to establish its</p> <ol style="list-style-type: none"> Establish the fairness with respect to age, gender, socioeconomic group and candidate background Predictive validity in year one assessment outcomes Stakeholder (MMI candidates and assessors) acceptability 	<ul style="list-style-type: none"> • Non-EU students and those for whom English was not a first language achieved significantly lower scores on MMI than their EU and English speaking counterparts (difference in mean 11.9% and 12.2% respectively, $P < 0.001$) • MMI score was associated with English Language proficiency (IELTS) ($r = 0.5$, $P < 0.01$) • Correlations emerged between First Year results and IELTS ($r = 0.44$; $p = 0.006$; $n = 38$) and EU school exit exam ($r = 0.52$; $p < 0.001$; $n = 56$) • MMI predicted EU student OSCE performance ($r = 0.27$; $p = 0.03$; $n = 64$) • MMI was considered a highly authentic assessment that offered a deeper understanding of the applicant than traditional tools, with an immediate relevance to clinical practice. • Cultural specificity of some stations and English language proficiency were seen to disadvantage international students. 	<ul style="list-style-type: none"> • Relation to other variable
Kelly M. et al., (54) 2014	<ul style="list-style-type: none"> • Feasibility Study design* • Clinical Science Institute, National University of Ireland, Galway, Ireland • N = 241 eligible students • Medical program 	<ul style="list-style-type: none"> • Establish the feasibility of running a MMI in an Irish setting 	<ul style="list-style-type: none"> • Cronbach Alphas for the 10 individual stations range from 0.74 to 0.80. Overall Cronbach Alpha of MMI was 0.78 • 90% either agreed or strongly agreed that the content of the MMI was relevant to their understanding of the practice of medicine • MMI was considered almost on a par with academic achievement as suitable grounds for selecting medical students MMI 73%; Academic achievement 79%; whereas other tools were less favored. 	<ul style="list-style-type: none"> • Response process • Internal structure
Liao SC. et al., (56) 2014	<ul style="list-style-type: none"> • National Cheng Kuang University (NCKU), Taiwan • N = 122 	<ul style="list-style-type: none"> • Strengthen the evaluation of applicants' interpersonal skills 	<ul style="list-style-type: none"> • Correlation between each of the seven MMI stations and the group 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable

	<ul style="list-style-type: none"> • Department of Medicine 	<ul style="list-style-type: none"> • Construct a more holistic and valid medical student admission interview 	<p>interview were all positive, ranging from 0.15 to 0.42</p> <ul style="list-style-type: none"> • The correlation coefficients were all significant (all $p < 0.05$), the only exception being that for station 5 (insightfulness) ($p = 0.10$) • Cronbach's alpha for the seven MMI stations was 0.54 • Cronbach's alpha for the MMI combined with the group interview rose to 0.63 • Combining the group interview with the MMI increased the internal consistency of the entire interview and made it a more valid interview 	
<p>Oliver T. et al., (57) 2014</p>	<ul style="list-style-type: none"> • Ontario Veterinary College (OVC) • N = 186 candidates • Doctor of Veterinary Medicine (DVM) program 	<ul style="list-style-type: none"> • Investigate if different MMI scores measure distinct non-cognitive attributes • Determine if MMI scores (construct specific or total MMI scores) are related to conceptually relevant personality measures and conceptually relevant scores in a standardized clinical communication interview 	<ul style="list-style-type: none"> • The reliability for the MMI was $E_p^2 = 0.73$, suggesting that scores for participants on oral communication and problem evaluation were generally consistent across raters and stations. • Standardized factor loadings ranged from 0.55 to 0.71 for oral communication, and from 0.32 to 0.58 for problem evaluation (all with $p < 0.001$). • The correlated uniqueness values for scores within the same station ranged from 0.48 to 0.67, and more importantly the correlation between the oral communication and problem evaluation factors was extremely high ($r = 0.87$) • These result suggest that, while there is a good fit and rational for a 2-factor model over a 1-factor model, there are method (station) and trait (attributes measured) effects present which limit the ability to conclude we are assessing two independent factors • Total MMI score had a weak but significant correlation with extraversion, and significant correlations with building the relationship and explaining and planning. 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Oyler D. R. et al., ⁵ 2014	<ul style="list-style-type: none"> • Pilot Study Design • University of Kentucky (United States) • N = 38 • Pharmacy Admission 	<ul style="list-style-type: none"> • Develop and implement a new interview process designed to assess the soft and hard skills necessary for success in postgraduate year 1 (PGY1) • Assess candidate and interviewer acceptance of the MMI as a part of the traditional pharmacy residency interview process • Research Electronic Data Capture (REDCap) 	<ul style="list-style-type: none"> • Candidates agreed that MMI allowed them to convey their abilities effectively but did not agree that it was more effective than traditional interview • Candidate did not agree that MMI caused less anxiety than traditional interviews • MMI was more effective than traditional interview in assessing candidates' abilities, skills and thought processes. 	<ul style="list-style-type: none"> • Response process
Roberts C. et al., (58) 2014	<ul style="list-style-type: none"> • Experimental Study design** • University of Sydney, Sydney, Australia • N = 1382 candidates • Medical school 	<ul style="list-style-type: none"> • Reliability and validity of non-cognitive characteristics of candidates when selecting them into general practice specialty training using MMI • Explored the concurrent validity of the MMI with the Situational Judgment Test (SJT) 	<ul style="list-style-type: none"> • For a single MMI question and one assessor, 28% of the variance between scores was due to candidate-to-candidate variation. • The varying views that interviewer had for particular candidates accounted for 40% of the variance in scores. • The generalizability co-efficient for a six question MMI was 0.7; to achieve 0.8 would require ten questions. • A disattenuated correlation with the SJT ($r = 0.35$), and in particular a raw score correlation with the subdomain related to clinical knowledge ($r = 0.25$) demonstrated evidence for construct and concurrent validity. 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable
Sebok S. S. et al., (59) 2014	<ul style="list-style-type: none"> • Experimental Study design (Nested design) • Queen's University, Kingston, Ontario, Canada • N = 455 school applicants • Medical Program 	<ul style="list-style-type: none"> • Examine the psychometric properties of the MMI as employed at another Canadian medical school. • Descriptive analyses is using Generalizability theory (G theory) and Many-Facet Rasch Model (MFRM) 	<ul style="list-style-type: none"> • Generalizability analyses: • Estimated variance component for applicants was 0.47 (16.3%) • The largest variance component was ps, the interaction between applicants and stations (1.21, 41.7%) • Large residual error variance ($pr:s$) of 1.17, or 40.3% of the total variance, are confounded by the interactions between applicants, raters, station and other unexplained sources of error 	<ul style="list-style-type: none"> • Internal structure

Stowe C. D. et al., (60) 2014	<ul style="list-style-type: none"> • Pilot study • University of Arkansas for Medical Sciences College of Pharmacy, United States • N = 13 volunteered (pilot phase); N = 224 (Full implementation) 	<ul style="list-style-type: none"> • Outline the development and implementation of the MMI at the University of Arkansas for Medical Sciences (UAMS) College of Pharmacy 	<ul style="list-style-type: none"> • There was little variability amongst stations (0.04, 1.3%) and raters nested within stations (0.01, 0.4%), indicating the stations had similar variability and the raters were scoring examinees consistently • Many-Facet Rasch analyses • Most of the applicants are situated in the 0–1 region on the logit scale, indicating that they were all fairly proficient • Both faculty and student raters are positioned around the 0 logit mark, which verify that they were equivalent in terms of their severity • Applicant reliability coefficient for the medical school applicants across the eight stations range from 0.84 to 0.92, which indicates heterogeneity among the applicants • Applicant reliability coefficients for the included attributes five (communication, critical thinking, maturity, empathy and professionalism) ranged from 0.66 to 0.76 • MMI meets some of the requirements for psychometric quality; however, there are still ongoing issues that need to be addressed. • The analyses were also able to identify problematic applicants, raters, stations and items 	<ul style="list-style-type: none"> • Internal structure
Zaidi N. B. et al., (69) 2014	<ul style="list-style-type: none"> • Pharmacy • Experimental Design Study (Nested design) • One of United States Medical School • N = 15 applicants 	<ul style="list-style-type: none"> • This study explore whether items, defined as specific attributes on an MMI evaluation form, are assessed consistently across MMI stations regardless of station scenario. • Generalizability (G) Theory 	<ul style="list-style-type: none"> • Applicant (<i>p</i>) represents only 6% of total variance. • The estimated variance components from the G study suggest that the greatest amount of variance is attributable to the main effect of the scenario (<i>s</i>) facet and the 	<ul style="list-style-type: none"> • Internal structure

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Alweis R. L. et al., (61) 2015	<ul style="list-style-type: none"> • Historical Study design** • at Northeastern United States Internal Medical Residency • N = 13 interviewers; N = 240 candidates • Qualitative study 	<ul style="list-style-type: none"> • Aimed to gain a better understanding of these biases from the perspective of the interviewer 	<p>interaction between <i>scenario</i> and <i>applicant (ps)</i>.</p> <ul style="list-style-type: none"> • These two variance (<i>ps</i>) components account for 77% of the total variance. • The <i>item</i> facet (<i>i</i>) represents the lowest estimated variance, estimating only 0.6% of the total variance in MMI scores. • The <i>scenario-item</i> interaction (<i>si</i>) accounts for only 1.4% of the total variance. • Low estimate of variance attributable to the <i>item</i> facet is reinforced by a high Cronbach's alpha (0.97) for the seven items, which suggests very high internal consistency among the attributes measured by this MMI • Five major themes regarding interviewer perceptions were sub grouped into specific applicant characteristics, personality factors, cultural factors, perception of prior preparation and concerns with norming. • Subjects noted that they felt that applicants with introverted personalities may have fared less well in the MMI process. • Many felt that the system itself could be biased toward extroverts who may be able to respond more quickly in situations that require spontaneous answers. • Seven interviewers noted that applicants unfamiliar with national systems and cultural norms may have been at a disadvantage. • Five interviewers voiced concern that scores might be biased by norm-referencing, or comparing the applicants to those coming before or after them on a specific interview day. 	<ul style="list-style-type: none"> • Response process

Burkhardt J. C. et al., (62) 2015	<ul style="list-style-type: none"> • Historical Study design** • University of Michigan, Ann Arbor, Michigan • N = 71 out of 72 interns • Emergency Medicine Residency Admission 	<ul style="list-style-type: none"> • Investigate whether it provides additional information regarding future first-year resident performance that can be useful in resident selection 	<ul style="list-style-type: none"> • An individual's score on the MMI correlated with overall performance ($p < 0.05$) in the single linear logistic regression. • MMI correlated with ACGME individual competencies patient care and procedural skills at a less robust level ($p < 0.1$), but not with any other outcomes. 	<ul style="list-style-type: none"> • Relation to other variable
Cox W. C. et al., (63) 2015	<ul style="list-style-type: none"> • Observational Design Study • University of North Carolina (UNC) • N = 253 • Chapel Hill Eshelman School of Pharmacy 	<ul style="list-style-type: none"> • Describe the development, implementation and evaluation of the Multiple Mini-Interview (MMI) within a doctor of pharmacy (PharmD) admissions model 	<ul style="list-style-type: none"> • Analyses suggest that MMI stations assessed different attributes as designed, with Cronbach's alpha for each station ranging from 0.90 to 0.95. • All correlations between MMI station scores and academic indicators (GPA & PCAT) were negligible. • No significant differences in average station scores were found based on age, gender, or race. 	<ul style="list-style-type: none"> • Internal structure • Relation to other variable
Sebok S. S and Syer M. D., (64) 2015	<ul style="list-style-type: none"> • Canadian Medical School • N = 455 Applicants who were shortlisted; N = 6 raters (3 student raters, 3 faculty raters) 	<ul style="list-style-type: none"> • Investigated how raters distinguish between different non-cognitive attributes in the context of performance-based assessments to better understand unexplained rater variance • Tools: Many Facet Rasch Model; Hierarchical Clustering 	<ul style="list-style-type: none"> • At times, raters were unable to distinguish between the various non-cognitive attributes • Applicant reliability values ranged from 0.84 to 0.92, which suggests heterogeneity within the sample. • Station 4 had a Rasch reliability of separation value of 0.94, which denotes substantial differences between the faculty and student raters. • Both faculty and student raters show the non-cognitive attribute of critical thinking as more distinct from communication and empathy. • Thus the placement of the attribute clusters against the center scale, which in this instance is approximately 0.30, indicates overall lower levels of dissimilarity and less variability within this station compared with station 4. 	<ul style="list-style-type: none"> • Response process • Internal structure
Yoshimura H. et al., (65) 2015	<ul style="list-style-type: none"> • Observational Study design • Tokyo Bay Urayasu-Ichikawa Medical Centre's, Urayasu City, Japan and Gifu University • N = 26 medical graduates 	<ul style="list-style-type: none"> • Investigated reliability and acceptability of the post-graduates admissions MMI with PBQ and SQ interview formats within MMI stations. 	<ul style="list-style-type: none"> • PBQ and SQ formats had generalizability coefficients of 0.822 and 0.821, respectively • One examiner per station, seven stations could produce a reliability 	<ul style="list-style-type: none"> • Response process • Internal structure

(continued on next page)

Table 2 (continued)

Source and year	Study design, Location, field of study, Number candidate	Objective & instruments	Results	Validity evidence
Abdul Rahim & Yusoff, 2016 (66)	<ul style="list-style-type: none"> • School of Medical Sciences, Universiti Sains Malaysia • N = 447 • Medical Student Admission 	<ul style="list-style-type: none"> • This paper described the implementation of the MMI and reported the preliminary evaluation data on its validity evidence. 	<p>of more than 0.80 in both PBQ and SQ formats.</p> <ul style="list-style-type: none"> • More than 60% of both candidates and examiners felt positive about the overall candidates' ability. • All participants liked the fairness of this MMI when compared with the previously experienced SSPI. • SQs were perceived more favorable by candidates; in contrast, PBQs were perceived more relevant by examiners • Overall reliability of 5 manned stations was 0.94 • CFA confirmed unidimensional of MMI • Positive rating on feasibility and acceptability of MMI by interviewers and candidates • Area for improvement was the variation of difficulty and discrimination of the stations between sessions. <ul style="list-style-type: none"> ○ Might indicate the quality of the question, assuming that the standards of the interviewers are similar. ○ Or might indicate varying interviewer standards between sessions. 	<ul style="list-style-type: none"> • Response process • Internal structure

there was positive correlation with simulation-based assessment,²⁹ communication skills,²⁹ and strength of argument score²⁹; emergency medicine grades correlated with MMI scores⁵⁹; MMI score was associated with English language proficiency (IELTS)⁶¹; correlation between each of the seven MMI stations and the group interview were all positive³⁸; and there was a significant correlation with building relationships, explaining, and planning,³⁹ a disattenuating correlation with the SJT,⁴⁰ and, in particular, a raw score correlation with the subdomain related to clinical knowledge.⁴⁰ Conversely, there was a non-significant correlation with OSCE, MCCEE, MCCQE¹⁷; MMI total score and EQ-i total score were not found to be significantly correlated²⁴; MMI measures different attributes than do PCAT and PPA^{25,44}; extraversion, conscientiousness, and agreeableness were correlated with MMI scores⁶⁴; those with extraversion scores in the top (versus the bottom) quartile had significantly higher MMI scores⁶⁵; MMI correlated positively with age and negatively with aboriginal status⁵⁴; there was a weak correlation with extraversion³⁹; and all correlations between MMI station scores and GPA were negligible.⁴⁴ Overall, MMI positively correlated with assessment scores, and candidate background was not a contributing factor; however, it positively correlated with English proficiency and inconsistently correlated with non-cognitive traits (Table 2).

Consequences. The retained and rejected candidates had significantly different total scores and mean scores for each station⁴⁹; predictive of OSCE performance, clerkship performance,⁶⁸ CLEO or PHELO performance,⁶⁸ and CDM performance⁶⁸; successful MMI candidates had higher scores in a licensing national examination than unsuccessful MMI candidates⁶⁹; and a consistent factor determined success in medical school assessment.⁶⁶ Overall, MMI somewhat predicted the performance of candidates during medical training, in licensing national examinations, and in the workplace (Table 2).

Discussion

The primary focus of this study was to explore validity evidence of MMI in the selection of students for higher education institutions, either within or outside health profession education. We found a significant number of articles to provide evidence to support its validity in five areas: (i) **Content** is the extent to which MMI covers a specific set of items to reflect the intended attributes to be assessed; (ii) **response process** refers to the relationship between the intended construct and the comprehension of respondents while responding to the items); (iii) **internal structure** is the degree of relationship between and among items and constructs); (vi) **relations to other variables** are the relationships of MMI scores to external variables; and (v) **consequences** are any evidence to signify the measurement scores on specific intended or unintended outcomes. Taken together, MMI has demonstrated reasonable validity evidence in the five areas.

A total of 64 articles were critically appraised, and the key findings were that (i) MMI is flexible for assessing various important attributes of candidates such as professionalism, communication skills, ethics and morals, and critical

thinking and problem solving; (ii) MMI was generally acceptable to both candidates and interviewers across 11 countries; (iii) MMI was consistently reliable and stable with acceptable Cronbach's alpha across educational settings; (iv) candidates showed high performance in clinical assessment and licensing national examinations; (v) MMI was reported as a bias-free admission tool for most factors such as culture and personal background, except for English proficiency; (vi) MMI was rarely correlated with non-cognitive attributes such as personality traits and emotional intelligence; (vii) MMI was mainly carried out (80%) in the undergraduate selection process; and (viii) MMI commonly includes seven to 12 stations per circuit, with each station requiring seven to 10 min. These key findings provide evidence to support the validity of MMI as an admission tool in the higher education context. Unfortunately, despite its potential, MMI implementation has not been reported outside the health profession education context.

The earliest study reported about MMI was in 2004 at McMaster University.² About a decade later, MMI had been successfully implemented in four countries,⁷⁰ and now it has been carried out in 11 countries across different regions: America, Europe, and Asia-Pacific. It can be concluded that MMI is being accepted as a global admission tool in higher education institutions across regions, educational settings, and cultures. One possible reason for the widespread use of MMI is its flexibility, acceptability, and reliability in assessing different content areas and attributes of candidates across different educational settings.^{7,70,71} Like OSCE, MMI possesses the ability to be adapted by institutions according to their financial and facilities capacity. Interestingly, some studies discovered that even with only five stations, the reliability of MMI was high enough for high-stake decisions such as admissions.^{27,47} Moreover, studies have shown that the cost of conducting MMI was similar to other forms of personal interview,^{7,47,50,70} thus supporting its efficiency. These facts signify the validity of MMI in terms of content, response process, and internal structure and confirm the findings of previous systematic reviews.^{7,70} Unfortunately, this review clearly showed that all studies had been conducted in the health profession education context; hence, further study should be conducted outside such a context to support MMI's credentials as an admission tool in higher education. This review recommends that further research on MMI validity should be carried out in postgraduate studies and areas outside the health profession education context to verify its credentials in the student selection process.

Similar to previous systematic reviews,^{7,70} this review revealed that MMI is inconsistently and poorly correlated with non-cognitive attributes such as personality traits, rural attribute domain, and emotional intelligence. Interestingly, MMI positively correlated with various cognitive-based and performance-based assessments such as OSCE^{4,16,66}; argument ability²⁹; reasoning skills¹⁵; simulation-based examination⁵⁹; and national council examinations.¹⁵ The findings indicate that proper design of MMI is important to ensure the recruitment of the most suitable candidates into higher education institutions. Unexpectedly, MMI was reported to correlate with several non-cognitive attributes such as communication skills,²⁹

building relationships,³⁹ and English proficiency level.⁶¹ It is worth mentioning that a significant correlation with English proficiency level could disadvantage non-native and second-language candidates during MMI; perhaps further study should be carried out to verify this postulation. This review suggests that MMI consistently correlated with cognitive attributes, and further research should be carried out to test its validity outside of health profession education contexts and its correlation with important non-cognitive attributes such mental health, ethics, and professionalism.

Previous systematic reviews echoed that MMI is lacking in predictive validity evidence,⁷⁰ but more research is required to support this aspect.⁷ In this review, MMI demonstrated its ability to select candidates who demonstrated high performance during medical training,^{66,68} in licensing national examinations,⁶⁹ and in the workplace.⁶⁸ For example, MMI predicted the performance of selected candidates during OSCE, clerkship, CLEO or PHELO performance, and CDM performance and is a consistent factor in determining success in medical school assessment.^{66,68} Furthermore, Eva et al. (2012) reported that selected candidates achieved high marks in a licensing national examination.⁶⁹ Nevertheless, these results were the initial evidence to support the predictive validity of MMI, especially in terms of non-cognitive attributes and outside the health profession education context. Therefore, future work should concentrate on consequences for important non-cognitive attributes.

Several messages can be taken from this systematic review. First, more research is obviously required to explore MMI's effective educational contribution to important non-cognitive outcomes related to personal values, professional conduct, and patient care. Second, research should no longer focus on the content, response process, and internal structure because these aspects have been confirmed by many studies; therefore, efforts must focus on other validity evidence, especially the consequences and relations of MMI with important non-cognitive attributes, to justify its worth and credibility, given the intensive resources being used for its implementation. Third, there is limited multi-centre study showing that MMI is a cultural bias-free admission tool, hence future work to address this gap should be encouraged. Fourth, despite being in a technology-driven era, technology's uses in MMI are largely unexplored; therefore, any effort to leverage technology to enhance the potential of MMI should be given due consideration. Finally, MMI has not been implemented outside the health profession education context; therefore, it may be interesting to explore its validity in such a context.

Conclusions

MMI has been widely adapted by various institutions in many countries and is gaining momentum owing to its potential and credentials. MMI has demonstrated its superiority in terms of acceptability, reliability, content validity, and as a bias-free admission tool in many studies. However, more research is required to provide evidence to support its educational impact on important non-cognitive outcomes.

Source of funding

This study was not funded by any grants.

Conflict of interest

The authors have no conflict of interest to declare.

Ethical approval

Due to the nature of a review article, ethical approval is not applicable.

Consent

Informed consent was not applicable in the study.

References

1. 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): 327–331.
2. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004; 38(3): 314–326.
3. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004; 38(3): 314–326.
4. Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Acad Med* 2004; 79(10): S40–S42.
5. Oyler DR, Smith KM, Elson EC, Bush H, Cook AM. Incorporating multiple mini-interviews in the postgraduate year 1 pharmacy residency program selection process. *Am J Health Syst Pharm* 2014; 71(4).
6. Hofmeister M, Lockyer J, Crutcher R. The acceptability of the multiple mini interview for resident selection. *Fam Med* 2008; 40(10): 734–740.
7. Rees EL, Hawarden AW, Dent G, Hays R, Bates J, Hassell AB. Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: a BEME systematic review: BEME Guide No. 37. *Med Teach* 2016; 38(5): 443–455.
8. Liberati AAD, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; 6(7): e1000100.
9. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006; 119(2): 166.e7–e16.
10. Eva KW, Reiter HI, Rosenfeld J, Norman GR. The relationship between interviewers' characteristics and ratings assigned during a multiple mini-interview. *Acad Med* 2004; 79(6): 602–609.
11. Moreau K, Reiter H, Eva K. Comparison of aboriginal and nonaboriginal applicants for admissions on the Multiple Mini-Interview using aboriginal and nonaboriginal interviewers. *Teach Learn Med* 2005; 18(1): 58–61.
12. 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.

13. Lemay JF, Lockyer JM, Collin VT, Brownell AKW. Assessment of non-cognitive traits through the admissions multiple mini-interview. *Med Educ* 2007; 41(6): 573–579.
14. Humphrey S, Dowson S, Wall D, Diwakar V, Goodyear HM. Multiple mini-interviews: opinions of candidates and interviewers. *Med Educ* 2008; 42(2): 207–213.
15. Roberts C, Walton M, Rothnie I, Crossley J, Lyon P, Kumar K, et al. Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Med Educ* 2008; 42(4): 396–404.
16. Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. Predictive validity of the multiple mini-interview for selecting medical trainees. *Med Educ* 2009; 43(8): 767–775.
17. Hofmeister M, Lockyer J, Crutcher R. The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Med Educ* 2009; 43(6): 573–579.
18. Roberts C, Zoanetti N, Rothnie I. Validating a multiple mini-interview question bank assessing entry-level reasoning skills in candidates for graduate-entry medicine and dentistry programmes. *Med Educ* 2009; 43(4): 350–359.
19. 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 post-graduate admissions. *Acad Med* 2010; 85(10): S60–S63.
20. Roberts C, Rothnie I, Zoanetti N, Crossley J. Should candidate scores be adjusted for interviewer stringency or leniency in the multiple mini-interview? *Med Educ* 2010; 44(7): 690–698.
21. Jones EP, Forister GJ. A comparison of behavioral and multiple mini-interview formats in physician assistant program admissions. *The Journal of Physician Assistant Education* 2011; 22(1): 36–40.
22. O'Brien A, Harvey J, Shannon M, Lewis K, Valencia O. A comparison of multiple mini-interviews and structured interviews in a UK setting. *Med Teach* 2011; 33(5): 397–402.
23. Uijtdehaage S, Parker N. Enhancing the reliability of the multiple mini-interview for selecting prospective health care leaders. *Acad Med* 2011; 86(8): 1032–1039.
24. Yen W, Hovey R, Hodwitz K, Zhang S. An exploration of the relationship between emotional intelligence (EI) and the Multiple Mini-Interview (MMI). *Adv Health Sci Educ* 2011; 16(1): 59–67.
25. Cameron AJ, MacKeigan LD. Development and pilot testing of a multiple mini-interview for admission to a pharmacy degree program. *Am J Pharmaceut Educ* 2012; 76(1).
26. Dowell J, Lynch B, Till H, Kumwenda B, Husbands A. The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Med Teach* 2012; 34(4): 297–304.
27. 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).
28. SAYS FE, Ayuob N, Fahmy AR, FAYEZ FE, HASANIAN M, EL DEEK B. Experience of establishment of multiple mini structure interview as part of student admission policy at faculty of medicine, King Abdulaziz University, 2011–2012. *Med Teach* 2013; 35(sup1): S74–S77.
29. TAVARES W, MAUSZ J. Assessment of non-clinical attributes in paramedicine using multiple mini-interviews. *Emerg Med J* 2013; 32(1): 70–75.
30. Till H, Myford C, Dowell J. Improving student selection using multiple mini-interviews with multifaceted Rasch modeling. *Acad Med* 2013; 88(2): 216–223.
31. Tiller D, O'Mara D, Rothnie I, Dunn S, Lee L, Roberts C. Internet-based multiple mini-interviews for candidate selection for graduate entry programmes. *Med Educ* 2013; 47(8): 801–810.
32. 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. *Med Teach* 2014; 36(8): 703–709.
33. Barbour ME, Sandy JR. Multiple mini interviews for selection of dental students: influence of gender and starting station. *J Dent Educ* 2014; 78(4): 589–596.
34. Callwood A, Cooke D, Allan H. Developing and piloting the multiple mini-interview in pre-registration student midwife selection in a UK setting. *Nurse Educ Today* 2014; 34(12): 1450–1454.
35. Eva KW, Macala C. Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Med Educ* 2014; 48(6): 604–613.
36. Hissbach JC, Sehner S, Harendza S, Hampe W. Cutting costs of multiple mini-interviews—changes in reliability and efficiency of the Hamburg medical school admission test between two applications. *BMC Med Educ* 2014; 14(1): 1.
37. Kelly M, Dowell J, Husbands A, Kropmans T, Jackson AE, Dunne F, et al. Can multiple mini interviews work in an Irish setting? A feasibility study. *Ir Med J* 2014; 107(7): 210–212.
38. Liao S-C, Hsiue T-R, Lin C-H, Huang A-M. Multiple mini-interviews combined with group interviews in medical student selection. *Med Educ* 2014; 48: 1104.
39. Oliver T, Hecker K, Hausdorf PA, Conlon P. Validating MMI scores: are we measuring multiple attributes? *Adv Health Sci Educ* 2014; 19(3): 379–392.
40. Roberts C, Clark T, Burgess A, Frommer M, Grant M, Mossman K. The validity of a behavioural multiple-mini-interview within an assessment centre for selection into specialty training. *BMC Med Educ* 2014; 14(1): 1.
41. Sebok SS, Luu K, Klinger DA. Psychometric properties of the multiple mini-interview used for medical admissions: findings from generalizability and Rasch analyses. *Adv Health Sci Educ* 2014; 19(1): 71–84.
42. Stowe CD, Castleberry AN, O'Brien CE, Flowers SK, Warmack TS, Gardner SF. Development and implementation of the multiple mini-interview in pharmacy admissions. *Currents in Pharmacy Teaching and Learning* 2014; 6(6): 849–855.
43. Zaidi NB, Swoboda C, Wang LL, Manuel RS. Variance in attributes assessed by the multiple mini-interview. *Med Teach* 2014; 36(9): 794–798.
44. Cox WC, McLaughlin JE, Singer D, Lewis M, Dinkins MM. Development and assessment of the multiple mini-interview in a school of pharmacy admissions model. *Am J Pharmaceut Educ* 2015; 79(4).
45. Sebok SS, Syer MD. Seeing things differently or seeing different things? exploring raters' associations of noncognitive attributes. *Acad Med* 2015; 90(11): S50–S55.
46. Yoshimura H, Kitazono H, Fujitani S, Machi J, Saiki T, Suzuki Y, et al. Past-behavioural versus situational questions in a postgraduate admissions multiple mini-interview: a reliability and acceptability comparison. *BMC Med Educ* 2015; 15(1): 1.
47. Abdul Rahim AF, Yusoff MSB. 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.
48. Brownell K, Lockyer J, Collin T, Lemay J-F. Introduction of the multiple mini interview into the admissions process at the University of Calgary: acceptability and feasibility. *Med Teach* 2007; 29(4): 394–396.
49. Harris S, Owen C. Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School. *Med Educ* 2007; 41(3): 234–241.
50. Rosenfeld JM, Reiter HI, Trinh K, Eva KW. A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Adv Health Sci Educ* 2008; 13(1): 43–58.

51. Kumar K, Roberts C, Rothnie I, Du Fresne C, Walton M. Experiences of the multiple mini-interview: a qualitative analysis. *Med Educ* 2009; 43(4): 360–367.
52. Razack S, Faremo S, Drolet F, Snell L, Wiseman J, Pickering J. Multiple mini-interviews versus traditional interviews: stakeholder acceptability comparison. *Med Educ* 2009; 43(10): 993–1000.
53. McAndrew R, Ellis J. An evaluation of the multiple mini-interview as a selection tool for dental students. *Br Dent J* 2012; 212(7): 331–335.
54. Reiter HI, Lockyer J, Ziola B, Courneya C-A, Eva K, Alliance CMM-IR. Should efforts in favor of medical student diversity be focused during admissions or farther upstream? *Acad Med* 2012; 87(4): 443–448.
55. McAndrew R, Ellis J. Applicants' perceptions on the multiple mini-interview process as a selection tool for dental and therapy and hygiene students. *Br Dent J* 2013; 215(11): 565–570.
56. 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): 465–469.
57. Raghavan M, Martin B, Burnett M, Aoki F, Christensen H, MacKalski B, et al. Multiple mini-interview scores of medical school applicants with and without rural attributes. *Rural Rem Health* 2013; 13(2362).
58. Andrades M, Bhanji S, Kausar S, Majeed F, Pinjani S. Multiple mini-interviews (MMI) and semistructured interviews for the selection of family medicine residents: a comparative analysis. *International Scholarly Research Notices* 2014; 2014: 1–6.
59. Hopson LR, Burkhardt JC, Stansfield RB, Vohra T, Turner-Lawrence D, Losman ED. The multiple mini-interview for emergency medicine resident selection. *J Emerg Med* 2014; 46(4): 537–543.
60. Joshi NK, Yarris LM, Doty CI, Lin M. Social media responses to the annals of emergency medicine residents' perspective article on multiple mini-interviews. *Ann Emerg Med* 2014; 64(3): 320–325.
61. Kelly ME, Dowell J, Husbands A, Newell J, Siun OF, Kropmans T, et al. The fairness, predictive validity and acceptability of multiple mini interview in an internationally diverse student population—a mixed methods study. *BMC Med Educ* 2014; 14(1): 1.
62. Alweis RL, Fitzpatrick C, Donato AA. Rater perceptions of bias using the multiple mini-interview format: a qualitative study. *Journal of Education and Training Studies* 2015; 3(5): 52–58.
63. Kulasegaram K, Reiter HI, Wiesner W, Hackett RD, Norman GR. Non-association between Neo-5 personality tests and multiple mini-interview. *Adv Health Sci Educ* 2010; 15(3): 415–423.
64. Griffin B, Wilson I. Associations between the big five personality factors and multiple mini-interviews. *Adv Health Sci Educ* 2012; 17(3): 377–388.
65. Jerant A, Griffin E, Rainwater J, Henderson M, Sousa F, Bertakis KD, et al. Does applicant personality influence multiple mini-interview performance and medical school acceptance offers? *Acad Med* 2012; 87(9): 1250–1259.
66. Husbands A, Dowell J. Predictive validity of the Dundee multiple mini-interview. *Med Educ* 2013; 47(7): 717–725.
67. Burkhardt JC, Stansfield RB, Vohra T, Losman E, Turner-Lawrence D, Hopson LR. Prognostic value of the MULTIPLE MINI-INTERVIEW for emergency medicine residency performance. *J Emerg Med* 2015; 49(2): 196–202.
68. Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Med Educ* 2007; 41(4): 378–384.
69. 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. *J Am Med Assoc* 2012; 308(21): 2233–2240.
70. 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. *Med Teach* 2013; 35(12): 1027–1041.
71. Knorr M, Hissbach J. Multiple mini-interviews: same concept, different approaches. *Med Educ* 2014; 48(12): 1157–1175.

Glossary

- BI: Behavioral Interview
 CI: Confidence Interval
 CLEO: Considerations of the Legal, Ethical and Organizational Aspects of Medicine
 DIF: Differential Item Functioning
 EM: Emergency Medicine
 EU: Group comprised from Ireland, Great Britain, Finland and Germany
 non-EU: Group comprised from Malaysia, Singapore, Canada and USA
 FF: Free Form
 GPA: Grade Point Average
 ICC: Interclass Correlation Coefficient
 IRT: Item Response Theory
 MMI: Multiple Mini Interview
 MCCEE: Medical Council of Canada Evaluating Examination
 MCCQE: Medical Council of Canada Qualifying Examination
 OT: Occupational Therapy
 OSCE: Objective Structure Clinical Examination
 PA: Physician Assistant
 PT: Physiotherapy
 PBQ: Past Behavioral Question
 PPA: Pre-pharmacy Average
 PPI: Personal Progress Inventories
 PCAT: Pharmacy College Admission Test
 PHELO: Population Health + CLEO
 SD: Standard Variation
 SJ: Situational Judgment
 SQ: Situational Question
 SEM: Standard Error of Measurement
 TI: Traditional Interview
 UAI: University Admission Index
 UMAT: Undergraduate Medical and Health Science Test
 US: United States

How to cite this article: Yusoff MSB. Multiple Mini Interview as an admission tool in higher education: Insights from a systematic review. *J Taibah Univ Med Sc* 2019;14(3):203–240.