Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programmes: a BEME systematic review

BEME Review 37

Eliot L Rees,[1,2] Ashley W Hawarden,[2] Gordon Dent,[1] Richard Hays,[3] Joanna Bates,[4] Andrew B Hassell[1,2]

[1]School of Medicine, Keele University, UK; [2]University Hospitals of North Midlands NHS Trust, UK; [3]School of Medicine, University of Tasmania, Australia; [4]Centre for Heath Education Scholarship, University of British Columbia, Canada.

Contact:

Eliot L Rees, School of Medicine, David Weatherall Building, Keele University, ST5 5BG.

eliotlloyd.rees@nhs.net

Abstract

Background

In the eleven years since its development at McMaster University Medical School, the multiple mini-interview (MMI) has become a popular selection tool. We aimed to systematically explore, analyse and synthesise the evidence regarding MMIs for selection to undergraduate health programmes.

Methods

The review protocol was peer-reviewed and prospectively registered with the Best Evidence Medical Education (BEME) collaboration. Thirteen databases were searched through 34 terms and their Boolean combinations. Seven key journals were hand searched since 2004. The reference sections of all included studies were screened. Studies meeting the inclusion criteria were coded independently by two reviewers using a modified BEME coding sheet. Extracted data was synthesised through narrative synthesis.

Results

A total of 4,338 citations were identified and screened, resulting in 41 papers that met inclusion criteria. Thirty-two studies report data for selection to medicine, six for dentistry, three for veterinary medicine, one for pharmacy, one for nursing, one for rehabilitation, and one for health science. Five studies investigated selection to more than one profession.

MMIs used for selection to undergraduate health programmes appear to have reasonable feasibility, acceptability, validity and reliability. Reliability is optimised by including 7-12 stations, each with one examiner. The evidence is stronger for face validity, with more research need to explore content validity and predictive validity. In published studies MMIs

do not appear biased against applicants on the basis of age, gender or socio-economic status. However, applicants of certain ethnic and social backgrounds did less well in a very small number of published studies.

Performance on MMIs does not correlate strongly with other measures of non-cognitive attributes, such as personality inventories and measures of emotional intelligence.

Discussion

MMI does not automatically mean a more reliable selection process but it can do if carefully designed. Effective MMIs require careful identification of the non-cognitive attributes sought by the programme and institution. Attention needs to be given to the number of stations, the blueprint and examiner training.

More work is required on MMIs as they may disadvantage groups of certain ethnic or social backgrounds. There is a compelling argument for multi-institutional studies to investigate areas such as the relationship of MMI content to curriculum domains, graduate outcomes and social missions; relationships of applicants’ performance on different MMIs; bias in selecting applicants of minority groups; and the long term outcomes appropriate for studies of predictive validity.

Background

In selection for undergraduate health programmes applicant numbers invariably exceed the number of available places. Medical schools strive to admit applicants' who have the cognitive skills to excel at the course, have the personal attributes sought after in a physician, enhance class diversity, and, increasingly, who contribute to the school's mission. Admissions processes intended to select applicants on these criteria must be valid, reliable, robust, defensible, and transparent.

Selection for undergraduate programmes such as medicine and pharmacy implies selection to the respective profession (Medical Schools Council 2010). Many of these programmes have low attrition rates (Yates 2012; Fortin *et al.* 2015) and for some, their graduates are automatically entitled to registration with regulatory bodies; the vast majority of applicants selected will proceed to practise in the field. Assessments for selection are therefore undeniably high stakes and arguably the most important assessments within undergraduate programmes.

Nayer (1992) considers the purpose of admissions procedures to be *"…to select students who will complete the educational programme and go into professional careers, do well in the programme, perform creditably in professional practice and possess the traits of character and ethical values desired of a professional person*". Furthermore, the Liaison Committee on Medical Education (2012), which accredits MD programmes in the United States and Canada, states that *"a medical education program must select for admission medical students who possess the intelligence, integrity, and personal and emotional characteristics necessary for them to become effective physicians"*. Clearly such professional programmes should select for a combination of cognitive and non-cognitive attributes (Albanese *et al.* 2003)

In addition to evaluating the cognitive and humanistic qualities of the individual applicant, many schools also seek to ensure diversity in the entering class, such as a mix of racial backgrounds (black, Latino, or aboriginal students) or differing socioeconomic backgrounds. This diversity has been shown to increase learning, change attitudes, and increase health care provision over time to underserved populations (Gurin *et al.* 2002; Whitla *et al.* 2003). Other medical schools espouse a social accountability framework, defined by the WHO as "the obligation to direct their education, research, and service activities towards addressing the priority health concerns of the community, region, and/or nation they have a mandate to serve" (Boelen 1999). Commonly such schools formulate a social mission to train physicians who will practise in underserved, often rural, areas. This mission leads to an increased importance of admitting students from rural backgrounds. These requirements for admissions processes are embedded in accreditation standards across the western world (Boelen and Woollard 2009; Liaison Committee on Medical Education, 2012).

For the purposes of admissions, cognitive abilities can be assessed using previous academic achievements and performance on admissions tests. In the United Kingdom (UK), academic achievement in national exams set at the end of secondary school (e.g. A levels), has been demonstrated to be highly predictive of academic performance in medical school, accounting for 65% of variance in undergraduate and postgraduate examination performance (McManus *et al.* 2013). In the countries where graduate entry to health programmes is more usual, such as the USA and Australia, national admissions tests such as the Medical College Admission Test (MCAT) and Graduate Australian Medical Schools Admissions Test (GAMSAT) provide a common assessment of cognitive ability irrespective of the undergraduate programme of study, and deliver similar results: they are highly predictive of academic achievement throughout medical school (Donnon *et al.* 2007; Puddey & Mercer 2014). Using these standardised assessments may, however, have unintended consequences for the social missions of schools. Davis *et al.* (2013) demonstrated that black and Latino applicants performed significantly worse on the redesigned MCAT than white applicants. Differential prediction analyses suggest that this difference in performance is not due to test bias and that other factors may be at play (Davis *et al.,* 2013). This differential performance did not result in smaller proportions of black or Hispanic applicants receiving offers of study, as other criteria were also considered in the selection process (Davis *et al.,* 2013). Medical schools enact policies and processes in order to achieve the class diversity and the long term outcomes they are looking for.

While quantitative measures of applicants' academic ability can be drawn upon, and medical schools can implement policies to enhance the admissions of select groups of students, individual attributes of applicants are more difficult to quantify. Written personal statements, individual interviews, panel interviews, references, and combinations thereof (Cleland *et al.* 2012) have all been tried, to no avail: each of these approaches is fraught with low reliability (Salvatori 2001; Kulatunga-Moruzi & Norman 2002). Within the UK the majority of schools have traditionally used a combination of academic ability, a personal statement, and a reference to shortlist applicants for a panel interview (Parry *et al.* 2006). Panel interviews, however, do not offer sufficient reliability to ensure the correct applicants are selected (Kreiter *et al.* 2004).

One problem inherent in panel interviews is context specificity: how an applicant behaves in one situation is not predictive of how they behave in others (Eva 2003; van der Vleuten 2014). In order to improve the reliability of judgements of an applicant's personal attributes multiple independent observations need to be made in multiple encounters, in different contexts, exploring different attributes.

In an attempt to demonstrate the predictive validity of cognitive and non-cognitive admissions measures, Meredith *et al.* (1982) arguably developed the precursor of the MMI. They investigated the ability of four individual 30-minute interviews in combination with measures of academic ability to predict clerkship performance and clinical knowledge. They found that the sum of the interview scores predicted subjective ratings provided during clinical clerkships.

In an attempt to overcome the limitations of panel interviews, Eva *et al.* (2004c) developed the multiple mini-interview (MMI). Based on the principles of the objective structured clinical examination (OSCE), an MMI involves applicants rotating through a series of stations each designed to assess one or more personal attributes. Each station typically consists of a task, a series of questions or unstructured discussion of a topic. Stations are observed by trained interviewers and assessed on pre-defined marking schedules (Eva *et al.* 2004c). Since the development of MMIs at McMaster University a number of schools internationally have adopted the approach in their admissions processes. Given the increasing popularity of this form of selection process, together with the not inconsiderable resource it might require relative to the admissions processes it is designed to replace or augment, and change management within a school adopting more traditional methods, it seems timely to consider systematically the evidence surrounding MMIs as a means of selection to health programmes.

Aim

Through this review we sought to explore, analyse and synthesise the evidence regarding the utility of MMIs for selection to undergraduate health programmes.

<u>Methods</u>

This is a systematic review reported in accordance with the STORIES statement (Gordon & Gibbs 2014)

Utility of assessments

Though usually associated with testing applicants enrolled in a programme, the principles of assessments are equally important in making decisions as to who should be admitted to the programme in the first place (Prideaux *et al.* 2011). Van der Vleuten (1996) defined the utility of assessments as a multiplicative function of the following variables: reliability, validity, educational impact, acceptability and cost. These factors, therefore, need to be considered when determining whether to adopt an assessment technique, though one might argue that cost should be included within a broader consideration of the feasibility of the tool.

Review question

The overall question for this systematic review was: *what is the evidence regarding the utility of multiple mini-interviews for selection to undergraduate health programmes?* Through consideration of the review question a number of sub-questions were addressed:

- How acceptable are MMIs to applicants, faculty and society?
- How feasible are MMIs?
- How valid are MMIs?
- How reliable are MMIs?

In addition, we describe an overall picture of the current variability of MMIs in use internationally.

Search strategy

A search strategy was developed with guidance from a liaison librarian for health. The following 13 electronic databases were searched through 34 terms and their Boolean combinations (Table 1 to be found in the Supplementary Materials Section): Education Research Information Centre (ERIC), Medline, Web of Science, EMBASE, Cumulative Index to Nursing and Allied Health (CINAHL), British Education Index (BEI), PsychINFO, British Nursing Index (BNI), Applied Social Sciences Index and Abstracts (ASSIA), Australian Education Index, Health Business Elite, Health Management Information Consortium (HMIC), and AMED Allied and Complementary Medicine. The limits imposed were: English language, human, 2004 to present.

The reference lists of all included papers were screened for additional relevant publications. Finally, the contents since 2004 of the following key journals were hand searched: *Advances in Health Sciences Education: Theory and Practice, Medical Education, Nurse Education Today, Medical Teacher, American Journal of Pharmaceutical Education, Journal of Rehabilitation Research and Development* and *Academic Medicine*. The initial search was performed in April 2013 and updated in April 2014.

Selection criteria

For this review we were interested in primary research relating to the use of MMIs in the admissions process for undergraduate health professional programmes. All formats of MMI were included, regardless of whether they involved group stations. In order to maximise the number of relevant studies and outcomes measured we studied admissions to all undergraduate health professions programmes. We defined this as admissions to health profession programmes of initial training regardless of applicants' qualifications on application. Applications to postgraduate programmes and postgraduate training programmes were excluded on the basis that applicants had already been pre-selected to enter an undergraduate programme, by some other means. Graduate entry programmes were included as they still provide a primary healthcare qualification and conduct their own admissions processes that are similar to other programmes.

No study was excluded from the review purely on the basis of study design, although studies had to provide primary data to be included (either quantitative or qualitative). Studies that were purely descriptive were excluded, as were commentary and opinion pieces.

As MMIs were developed by Eva and colleagues at McMaster University in 2004, only studies since (and including) were included.

A summary of the inclusion and exclusion criteria can be found in Table 2, to be found in the Supplementary Materials Section.

Screening and selection of studies

All papers underwent an initial screening process by one reviewer which prioritised sensitivity over specificity, so only articles with titles that indicate they were obviously irrelevant and were in no way related to health professions education were excluded, for example 'Outcome of adolescent pregnancy at a university hospital in Jordan'. The abstracts of the remaining articles were independently assessed by two reviewers against the inclusion and exclusion criteria. If both reviewers agreed to include the paper it was retrieved and progressed to the coding stage; if both reviewers agreed to exclude the paper the article was moved to an excluded article database. In the case of disagreement the full paper was retrieved and assessed against the inclusion and exclusion criteria.

Data extraction

Full articles were retrieved for all remaining studies and coded by reviewer pairs on an adapted BEME coding sheet (Supplementary file 1, to be found in the Supplementary Materials Section).

Data extracted included: details of the citation, evaluation methods, institution of study, country of study, profession, study aim, details of the MMI used, authors' key findings and summary notes for review questions.

Authors were not contacted for further information regarding interventions. Where information was not available it is indicated as 'not reported'.

A pilot study was conducted to ensure reviewers were coding consistently. All reviewers independently extracted data and assessed the methodological quality of five papers in two rounds (two papers then three papers). Reviewers met to discuss data extracted and ensure consistency.

The provisional coding sheet was piloted with reviewers coding two articles independently before meeting to discuss amendments to be made to the coding sheet and consistency of data extraction. The coding sheet was revised in order to ensure all relevant data was captured. Reviewers then independently coded a further three articles, after which the coding sheet was finalised.

Assessment of methodological quality

Papers were assessed for methodological quality, independently by two reviewers, using three criteria rated on a 5-point Likert scale: appropriateness of study design, implementation of study, and appropriateness of data analysis. Additionally, each paper was rated on a 5-point global score for study quality. These scores were summed to give a total score for methodological quality out of 20. Free text comments were also made to justify high or low quality scores. The review group met to discuss the methodological quality of included papers and any discrepancies between quality scores were discussed until consensus was achieved. Kappa values were 0.73 ($P<0.001$) and 0.94 ($P<0.001$) for assessment of methodological quality and strength of findings, respectively.

Data synthesis / analysis

Insufficient data were available for a meta-analysis, as studies had neither a comparator nor an effect size. A narrative review was therefore performed.

Results

Search results

The database search yielded 4,335 articles (Table 3, to be found in the Supplementary Materials Section). Hand searching and reference searching identified a further 1 and 2 articles respectively. 1,903 duplicates were excluded. A further 2,114 papers were excluded through title screening as they were considered to be irrelevant (e.g. not pertaining to admissions). Three hundred and twenty-one abstracts were reviewed against inclusion and exclusion criteria. Two hundred and seventy did not meet inclusion criteria. Nineteen full papers were screened against inclusion criteria, ten of which were included. The full papers of 41 articles were retrieved and independently coded using a modified BEME coding sheet.

Figure 1(to be found in the Supplementary Materials Section) illustrates the included and excluded papers.

Methodological quality

The majority of papers (34 of 41) reported studies in which MMI had been used to inform selection decisions. While providing evidence for their feasibility, this has consequences for their ability to draw conclusions regarding predictive validity: the range of MMI scores of admitted applicants is decreased as only the highest scoring applicants receive offers for study. Therefore, these studies were not able to detect if students who score poorly on their MMI would perform poorly on assessments during the programme. Early studies in which MMIs were conducted concurrently to the institutions' regular admission processes, and in which the scores did not contribute to selection, were able to study predictive validity using a full range of MMI scores. These studies have, however, followed a small cohort of students who participated in both standard interviews (for selection purposes) and MMIs (for research purposes), and have provided much of the evidence to support claims for predictive validity of MMIs. One study was neither limited by range restriction nor by small cohort size, providing arguably the strongest evidence of predictive validity (Eva et al 2012).

A further limitation imposed by researching MMIs that have been used for selection decisions relates to the study of acceptability to applicants. All but one of these studies has used questionnaires issued to applicants after their MMI. While they were reassured that their response to the questionnaire would not affect any decisions regarding admission, they may not have felt confident in rating the process negatively for fear of adverse consequences. Furthermore, these applicants had chosen to apply to an institution that uses MMI for selection, and hence their views on MMIs may not be representative of all medical applicants.

Due to test security considerations, many of the papers lack sufficient detail regarding the content and process of the MMI to allow extensive interpretation of the results. MMIs are essentially an assessment method, and like all assessment methods their validity depends on the way in which they are implemented and the content they assess.

Finally, the research in all aspects of MMIs has to date been almost exclusively conducted within single institutions.

Summary of included papers

Of the 41 included papers 32 (78%) report data from MMIs for selection to undergraduate medicine, 6 (15%) for dentistry, 3 (7%) for veterinary medicine, 1 (2%) for pharmacy, 1 (2%) for nursing, 1 (2%) for rehabilitation, 1 (2%) for therapy and hygiene, and 1 (2%) for allied health sciences. Five included studies (12%) report findings from selection to more than one profession.

The 41 included papers report data from 20 institutions, with a maximum of 10 papers from a single institution (McMaster University Medical School).

Supplementary file 2 (to be found in the Supplementary Materials Section) reports details of all included studies.

Summary of MMIs in use at different institutions

The mean number of stations per MMI is 9.2 (mode: 10; range: 5 to 12). The stations last a mean of 7.3 (mode: 8; range: 5 to 10) minutes. Fourteen institutions use one interviewer per station, 2 use two, 3 use either one or two per station, and two have not reported the number of interviewers.

Feasibility

MMIs have been reported to be feasible (Brownell *et al.* 2007), and some schools have even found them to be 'logistically simpler' than other interview methods such as panel interviews as they required fewer interviewers and less time commitment per interviewer (Harris & Owen 2007; Brownell *et al.* 2007), though it should be noted that many MMIs incur the extra logistics of organising simulated patients for communication stations. Organisations with experience of delivering OSCEs will also be familiar with the logistical challenge of moving candidates between stations. Brownell *et al.* (2007) report the ability to conduct all of the admissions interviews over just a few days as advantageous. Challenges posed through the introduction of MMIs include: recruiting sufficient interviewers, space, organisation and station development (Dowell *et al.* 2012). Cameron & MacKeigan (2012) recommend using established station banks during the implementation of MMIs. Eva *et al.* (2004c) suggest organisations with experience of organising OSCEs may use similar facilities and processes. Tiller *et al.* (2013) describe feasibly running an internet based-MMI (using Skype videoconferencing software) for international applicants.

When compared to their more traditional interview process Brownell *et al.* (2007) found that an MMI process allowed them to interview more applicants within a given time utilising fewer interviewers, with each of them having to dedicate less time to the process (average time devoted for MMI was 9 hours, time devoted for traditional interview was 8-14 hours).

Acceptability

<u>Applicants</u>

The studies investigating acceptability of MMIs to applicants reveal their views on the information provided regarding the MMIs, the timing of the stations, how stressful the MMI was, the ability to recover from stations and applicants' overall preference between MMI and traditional panel interviews.

Information. Applicants to the School of Medicine at the University of Calgary indicated that the information they received prepared them for the MMI (mean rating = 4.06 / 5; Brownell *et al.* 2007). McAndrew & Ellis (2012) report 91% and 71% of applicants being satisfied with the information provided on the structure and content of the MMI, respectively.

Timing. Applicants to the School of Medicine at the University of Calgary felt that there was sufficient time to present their ideas at stations (mean rating = 3.64 / 5; Brownell *et al.* 2007), though the duration of their stations was not reported. For admission to the School of Pharmacy at the University of Toronto, 47% of applicants felt six-minute stations were 'just right' and 50% 'a bit short', for the eight-minute stations 50% indicate they were 'just right' and 43% 'a bit long' (Cameron & MacKeigan 2012). Kumar *et al.* (2009) report that applicants to the School of Medicine, Sydney University, commented that stations were not long enough which resulted in a pressure to speak more quickly. Twenty-three percent of applicants to the School of Medicine, University of California, Los Angeles (UCLA), were of the opinion that the eight-minute per station MMI was well timed, with 47% indicating they were too short and 30% indicating they were too long. (Uijtedhaage *et al.* 2011).

Stress. A third of applicants to the University of Dundee Medical School indicated they felt the MMI was more stressful than traditional panel interviews (Dowell *et al.* 2012). Applicants to the University of Calgary Veterinary School also found the MMI process more stressful (mean score 3.5/5 on Likert scale; Hecker *et al.* 2009) than traditional panel interviews. Razack *et al.* (2009) reported that applicants found the MMI more stressful than the traditional interview (3.78 *vs* 3.39 on 6-point Likert scale, $F=6.04$, $P=0.016$). Forty-four percent of applicants to the School of Medicine, UCLA indicated that they found the MMI stressful (Uijtdehaage *et al.* 2011). However, applicants undergoing MMI at the School of Medicine, University of Calgary were neutral when asked if they found the MMI stressful (Mean rating 2.89/5 (SD 1.11); Brownell *et al.* 2007).

Recovery. Applicants reported appreciating the opportunity to recover from poor performance on previous stations. As each station was scored independently they recognised their performance on one would not affect their performance on the next, each station offered a 'clean slate' (Eva *et al.* 2004c; Kumar *et al.* 2009), although others felt that poor performance could not be forgotten and affected their performance on subsequent stations (McAndrew & Ellis 2012).

Preference. 74% of applicants to Dundee Medical School (Dowell *et al.*, 2 012) and 65% of applicants to Cardiff University School of Dentistry (McAndrew & Ellis 2012) preferred the MMI to traditional interviews they had experienced. Applicants to McGill found the MMI more enjoyable than traditional interview (4.96 *vs* 4.66 on 6-point Likert scale, F=3.65, *P*=0.06; Razack *et al.* 2009)

Assessors

Interviewers indicated that they enjoy participating in MMIs (Eva *et al.* 2004c), though they can be tiring (Eva *et al.* 2004c, McAndrew & Ellis 2012). Findings in studies of interviewer acceptability have described the perceived fairness, the timing and willingness to participate in MMIs. Seventy-one percent of interviewers at the School of Nursing, Kingston University and St George's University of London stated a preference for MMIs over traditional interviews (Perkins *et al.* 2013)

Fairness. Interviewers indicated that they perceive the MMI to be a fair selection tool (Brownell *et al.* 2007; Kumar *et al.* 2009; Razack *et al.* 2009; Dowell *et al.* 2012) but some have concerns regarding how stressful it may be for applicants (Kumar *et al.* 2009; Dowell *et al.* 2012).

Timing. Interviewers at Calgary Medical and Veterinary Schools indicated they had sufficient time to assess applicants (Brownell *et al.* 2007; Hecker *et al.* 2009). At the School of Pharmacy, University of Toronto 69% felt six minutes was 'just right' and eight minutes 'a bit long' (Cameron & MacKeigan 2012). Interviewers at McMaster Medical School agreed that eight minutes was more than enough to assess applicants (Eva *et al.* 2004c).

Future participation. Eighty-nine percent of interviewers at the School of Medicine, UCLA (Uijtdehaage *et al.* 2011), 94% of interviewers at Dundee Medical School (Dowell *et al.* 2012), 99% of interviewers at Calgary Medical School (Brownell *et al.* 2007), and 100% of interviewers at School of Pharmacy, University of Toronto (Cameron & MacKeigan 2012) indicated that they would be willing to participate in MMIs in the future, although a desire for further training has been identified (Eva *et al.* 2004c; Dowell *et al.* 2012).

Reliability

In studies of reliability three coefficients are typically reported: Cronbach's alpha, intra-class correlation, and generalizability. Cronbach's alpha represents the correlation between constituents of the overall assessment (Schuwirth & van der Vleuten, 2011). Within MMIs the Cronbach's alpha is typically the correlation between scores assigned on different stations.

Intra-class correlation refers to the correlation between a group of pairs of scores (Bartko 1966). For MMIs the intra-class correlation coefficient could calculate the correlation between two examiners or two rating scales each on a group of stations.

Generalizability (G) refers to the contribution to the overall variance in scores that can be attributed to the variable under investigation (Bloch & Norman, 2012). In the context of MMIs the G coefficient is the proportion of variance in MMI score that is attributable to differences in applicants' non-cognitive abilities.

<u>Internal reliability</u>

Correlations between items within stations have been consistently very high. Eva *et al.* (2004b) reported inter-item (intra-station) correlations of 0.96; Lemay *et al.* (2007) reported Cronbach's alphas for stations ranging from 0.97 to 0.98; Oliver *et al.* (2014) reported a correlation of 0.87 between oral communication scores and problem evaluation scores derived from all 8 stations within Calgary Veterinary School's MMI.

The inter-station reliability of MMIs has been shown to be reasonably high, ranging from 0.59 (Uijtedhaage *et al.* 2011) to 0.87 (Hecker *et al.* 2009) (Table 4, to be found in the Supplementary Materials Section). Studies have investigated the effects of number of stations, number of raters per station, duration of stations, and format of stations on the reliability of MMIs.

Number of stations. Since early in the development of MMIs at McMaster it has been reported that the number of stations is the main determinant of internal reliability. Generalizability analyses have repeatedly indicated that MMIs with greater numbers of stations will have greater reliability (Eva *et al.* 2004c; Roberts *et al.* 2008; Sebok *et al.* 2013). Figure 2 (to be found in the Supplementary Materials Section) illustrates reliability coefficients for 32 admissions cycles at 13 institutions using MMIs for undergraduate selection. For MMIs with seven or more stations there does not appear to be any increase in measured reliability with increasing station numbers.

Number of raters per station. Early work by Eva *et al.* (2004b) using generalizability and decision studies, concluded that whilst increasing the number of raters per station does improve reliability, greater improvements are seen when the numbers of stations is increased, and therefore it is more appropriate to utilise raters individually in stations. This finding was corroborated by Roberts *et al.* (2008) and Hecker & Violato (2011). Few institutions have since employed more than one rater per station.

Duration of stations. Dodson *et al.* (2009) studied 175 applicants for entry to Deakin University School of Medicine. Raters in half of the 10 stations in their MMI scored applicants at five minutes and then again at eight minutes. The G coefficients of the five and eight minute scorings were 0.75 and 0.78, respectively. Cameron & MacKeigan (2012) calculated intra-class correlation coefficients for five 6-minute stations and five 8-minute stations in their 10-station MMI for entry to pharmacy at the University of Toronto, finding them to be 0.66 and 0.54, respectively.

MMIs by Skype In an effort to reduce costs associated with mounting an MMI at an international site for international applicants, Tiller *et al.* (2013) introduced an internet-based iMMI that utilised Skype. The generalisability of the iMMI for international applicants and the in person MMI for local applicants were reported as 0.76 and 0.70, respectively.

Inter-rater reliability

Most MMIs employ one rater per station, thus reports of inter-rater reliability are limited. Hecker & Violato (2011) reported an inter-rater reliability of 0.52 for the two raters on their seven-station MMI. Sebok *et al.* (2013) found inter-rater reliabilities of 0.41 to 0.69 for stations scored by faculty members and students.

Research has focused on correlating scores from different groups of raters within MMIs, rater training, and on the effect of interviewer stringency.

Inter-group ratings. Cameron & MacKeigan (2012) reported that student interviewers gave slightly higher mean ratings than faculty members or practitioners. Eva *et al.* (2004b) investigated the reliability of ratings assigned by faculty members and community members by occupying three stations with two faculty members each, three stations with two community members each, and three stations with one faculty member and one community member each. The generalizability of the community member manned stations was highest (0.58), followed by the faculty member manned stations (0.46), with the faculty and community member stations having the lowest generalisability (0.31), suggesting faculty and community members' assessments of applicants differed. They also found a non-significant difference between the mean scores assigned by faculty members (4.66/5) and the scores assigned by community members (4.96/5) ($F_{1,53}$=3.972, *P*=0.06). This finding is contradicted by that of Hecker & Violato (2011) who found a non-significant main effect of interviewer type, faculty member (mean score: 10.33/15) *vs* community veterinarian (mean score: 10.06/15), on MMI scores ($F_{1,1428}$=3.18, *P*=0.075)

Rater training. Several authors have identified rater training as an; area in need of development to improve the reliability of their MMIs (Eva *et al.* 2004c Sebok *et al.* 2013). Griffin & Wilson (2010) observed that when they changed from information-based rater training to skills-based rater training that involved rating simulated interviewees the proportion of variance in their MMI scores attributable to differences between raters was reduced from 20.2% to 7.0% (t=4.42, *P*=0.004).

Interviewer stringency. Three studies have reported using multi-faceted Rasch modelling (MFRM) to adjust for rater stringency or leniency within MMIs (Roberts *et al.*

2009; Roberts *et al.* 2010; Till *et al.* 2013). Till *et al.* (2013) found that using 'fair scores' (those adjusted for rater stringency) would alter the admissions decision for between 3.1 and 4.2% of applicants for undergraduate medicine at Dundee.

Test-retest reliability

Since it is less common for applicants to be interviewed more than once at the same institution using the same assessment, test-retest reliability evidence for MMIs is limited. The use of selection centres to run MMIs in Israel, however, has enabled analysis of re-applicants on a considerable scale. Gafni *et al.* *(*2012) have reported 405 applicants repeating MOR and 230 repeating MIRKAM (MOR and MIRKAM are Hebrew acronyms for two different MMI protocols used at the selection centres for different groups of schools). They have reported test-retest correlations – adjusted for range restriction as only those with low MMI score would have retaken – of 0.72 and 0.65 for total MOR and total MIRKAM scores, respectively. This moderate test-retest reliability for these MMIs suggests that performance does not vary considerably between attempts.

Validity

Within assessments, validity refers to the extent to which the test measures what it intends to measure (Schuwirth & van der Vleuten 2011). There are several ways of describing validity, including: face validity, content validity, discriminant validity, bias and predictive validity. Messick (1995) suggests that each of these types of validity shouldn't be considered separately, rather that different aspects of validity contribute to an overall unified validity of the assessment. It is not sufficient to have evidence of one aspect of validity, neither is it necessary to have evidence of all; rather a judgement about the overall validity of an assessment can be made based on the accumulation of evidence across the aspects. When examining validity, one should recognise that validity is a property of the meaning of the test scores generated by an assessment, rather than that of the assessment method itself. These scores depend on the items within the assessment, the persons taking the assessment and the context within which the assessment is taken (Messick 1995). Therefore, if MMIs are designed carefully to measure non-cognitive attributes, they would be expected to show divergent correlation to cognitive measures. However, depending on the content of each school's MMI, the non-cognitive attribute measured might be very different (e.g. communication ability *vs* ethical reasoning), and therefore the predictive ability of each MMI might be very different from school to school. Thus the construct validity of each MMI is not necessarily transferable to another school's MMI where different attributes are valued and therefore assessed. That being said, if evidence accumulates across different institutions and across different aspects of validity, then one can reasonably conclude that MMIs can be designed to be valid assessments. For the sake of clarity we have split the results regarding validity in to the different aspects, will later discuss how this informs a judgement on the unified validity of MMIs.

Face validity is the extent to which the test appears at face value to test what it is designed to. Content validity refers to the extent to which the content of a test represents all of the areas the test claims to assess. For example, an MMI that claims to assess non-cognitive attributes of applicants, but only has stations assessing ethical decision-making would have poor content validity. Construct validity describes how the assessment relates to other assessments of similar or different constructs. Weak correlations (discriminant) should be seen between tests that intend to measure different constructs (e.g. cognitive tests and non-cognitive tests) and stronger correlations (convergent) should be seen between two tests that report to measure the same construct (e.g. two different tests of communication skills) (Schuwirth & van der Vleuten 2010). Bias refers to the whether attributes other than those designed to be assessed (e.g. socioeconomic status, ethnicity, gender) affect performance on the assessment.

Predictive validity refers to the extent to which performance on the assessment is associated with performance on future assessments or practice, i.e. the extent to which this test can predict future performance.

<u>Face validity</u>

Cameron & MacKeigan (2012) surveyed interviewers (n=30) and applicants (n=30) at the School of Pharmacy, University of Toronto regarding the face validity of their stations and found that 93% and 97% agreed or strongly agreed that the stations were relevant to their pharmacy training. Dowell *et al.* (2012) surveyed assessors (n=116) at a station level on how well they achieved what they set out to do; seven of ten stations received ratings of 'very well' or 'moderately well' from 75% of respondents. Applicants to the School of Medicine, UCLA indicated that they felt the MMI process was free of cultural or gender bias (Uijtdehaage *et al.* 2011).

<u>Content validity</u>

The content of each MMI is determined by the non-cognitive attributes defined by the admitting institution as important for admission, and testable by an MMI station.

Blueprinting. Authors report ensuring content validity by creating a blueprint of the specific non-cognitive attributes agreed to by the admitting institution. Stations are developed based on this blueprint. (Eva *et al.* 2004c; Cameron & MacKeigan 2012). For example, Harris & Owen (2007) utilised Q methodology (Brown 1996) to determine the attributes most valued by stakeholders at the Australian National University, and identified six factors: love of medicine and learning, groundedness, self-confidence, balanced approach, mature social skills and realism. A ten station MMI was then developed to assess these factors.

Number of factors assessed. There is wide variation between institutions as to how many distinct factors they consider their MMI process to be assessing. Roberts *et al. (*2009) argue that their MMI, used for selection at University of Sydney, measures one concept: 'entry-level reasoning skills in professionalism'. Oliver *et al.* (2014) report that a 2-factor model (oral communication and problem evaluation) best explains the variance in their MMI, though they note the two constructs were highly correlated at 0.87. Hecker *et al.* (2009) performed three factor analyses, one on each of the three measures scored within stations: non-cognitive attributes, communication skills and critical thinking skills. The non-cognitive attributes scores were combined with grade point average (GPA) and age and resulted in a three-factor solution; 'moral and ethical values', 'interpersonal ability' and 'academic ability'. Communication skills scores all loaded on to one factor, and critical thinking loaded on to two separate critical thinking factors, suggesting a total of six factors. Finally, Lemay *et al.* (2007) report the analysis of their MMI to reveal a 10 factor solution, with each station forming a single factor.

Effect of preparation. Some authors have investigated the effect of coaching on applicant performance in the MMI. Griffin *et al.* (2008) investigated the effect of previous interview experience and coaching on MMI performance in a sample of 287 applicants. Students who had reported being coached performed no differently to those who had not on total MMI score (3.54 *vs* 3.56 , *P*=0.72); however, coached students performed significantly worse on the communication skills station (3.81 *vs* 4.01, *P*=0.044). They found no difference in total MMI scores between students who had attended interviews at other universities and those who were 'interview naive'. Seventeen applicants repeated their MMI a year following rejection and saw an increase in their ranking (interview z score) between attempts (-0.72 to 0.00, t=4.14, *P*=0.001).

Reiter *et al.* (2006) observed that access to station details (for one of two pilot stations) in advance of an MMI did not result in improved performance over those who did not have access (4.92 *vs* 4.94, t(383)=0.24, *P*>0.8).

Convergent and discriminant correlations with external variables

Divergent correlations

Cognitive measures. MMI scores have been reported to have low correlations with measures of past academic performance such as GPA (r=0.006 to 0.06; Kulasegaram *et al.* 2010; Eva *et al.* 2012), pre-pharmacy average (r=-0.025; Cameron & MacKeigan 2012) or Universities Admission Index (r=-0.03 to 0.11; Griffin & Wilson 2012). Small correlations have been reported for pre-admissions measures of cognitive ability such as: Graduate Australian Medical School Admission Test (section 1, reasoning in humanities and social sciences, r=0.20; section 2, written communications, r=0.20; section 3, reasoning in biological and physical sciences, r=0.12; Roberts *et al.* 2008), Undergraduate Medicine and Health Sciences Admission Test parts 1 (logical reasoning ability; r= -0.11 to 0.01), 2 (interpersonal understanding; r= 0.13 to 0.22), and 3 (non-verbal reasoning; r= -0.11 to -0.06; Griffin & Wilson 2012), UK Clinical Aptitude Test (β=0.00, *P*=0.28; O'Brien *et al.* 20110), Medical College Admissions Test (r=0.10; Kulasegaram *et al.* 2010), and Pharmacy College Admission Test (r=0.042; Cameron & MacKeigan 2012).

Non-cognitive measures. Small (and frequently non-significant) correlations have been reported between MMI scores and other admissions measures of non-cognitive ability: personal interview (r=0.185; Eva *et al.* 2004c), simulated tutorial (r=0.317; Eva *et al.* 2004c), and autobiographical sketch (r=0.014 to 0.170; Eva *et al.* 2004c; Eva *et al.* 2012)

Convergent correlations

Non-cognitive measures. Total scores for the two MMIs (MOR and MIRKAM) co-ordinated by the National Institute for Testing and Evaluation in Jerusalem are highly correlated (r=0.75; Gafni *et al.* 2012) suggesting they are measuring similar constructs. Furthermore, moderate correlations have been reported between MOR and MIRKAM scores and a judgement and decision making questionnaire (MOR: r=0.53, MIRKAM: r=0.46; Gafni *et al.* 2012) and strong correlations with a biographical questionnaire (MOR: r=0.72, MIRKAM: r=0.72; Gafni *et al.* 2012).

Personality. Four studies have investigated the associations between MMI score and personality types, three of which have used the NEO big five personality types: neuroticism, extroversion, openness to experience, agreeableness, and conscientiousness (McCrae & Costa, 1994). Though the studies have used different tools to assess personality, each tool gives a value for extent to which the respondent meets that personality trait; correlations between these values and MMI scores have been investigated. Kulasegaram *et al.* (2010) reported no associations between total MMI score and any of the NEO big five. Jerant *et al.* (2012) reported that the only personality trait significantly associated with MMI score was extroversion (r=0.35, $P<0.01$). Griffin & Wilson (2012) found extroversion (0.19 to 0.30, $P<0.002$), agreeableness (0.14 to 0.19, $P<0.002$) and conscientiousness (0.20 to 0.25, $P<0.002$) to be correlated with total MMI score. Oliver *et al.* (2014) investigated the associations between MMI and extroversion and emotionality, and reported extroversion was associated with higher MMI score (r=0.22, $P<0.05$) but emotionality was not (r=-0.01, n/s). Within these analyses of multiple correlations, Kulasegaram *et al.* (2010) and Griffin & Wilson (2012) applied Bonferroni corrections, whereas Jerant *et al.* (2012) and Oliver *et al.*

(2014) did not appear to, so the associations reported in the latter two studies are more likely to have suffered type 1 errors.

Emotional intelligence. Yen *et al.* (2011) reported no correlation between a validated self-report measure of emotional intelligence, defined as ‘a type of social intelligence that involves the ability to monitor one’s own thinking and actions’ (Salovey & Mayer 1990), and total MMI score for 196 applicants for admission to the health sciences programme at the Michener Institute for Applied Health Sciences. Cherry *et al.* (2014) warn against using assessments of emotional intelligence in admissions, and suggest, rather, that attention should be paid to developing students’ emotional intelligence within the curriculum.

Consequences as validity evidence

Bias

Test bias arises “when deficiencies in a test itself or the manner in which it is used result in different meanings for scores earned by members of different identifiable subgroups” (American Educational Research Association *et al.* 1999). Some authors have investigated whether certain groups perform less well on their MMIs. It is difficult to be definitive about whether MMIs may bias against certain groups from the available evidence. List references here for some “authors”

Male vs female. Jerant *et al.* (2012) reported a positive association between MMI score and female sex (P<0.01), though no significant differences were found in five other studies (Eva *et al.* 2004c; Hecker *et al.* 2009; O’Brien *et al.* 2011; Uijtdehaage *et al*. 2011; Reiter *et al.* 2012). Griffin & Wilson (2010) found that interviewers were not more lenient towards interviewees of the same gender.

Age   O'Brien *et al.* (2011) found no correlation between age and MMI score on either the undergraduate or graduate entry medicine programs at St George's University of London. Reiter *et al.* (2012), however, reported a slight, but highly significant, positive correlation between age and MMI score in the 2008 admissions cycle (r=0.124, n=786, *P*=0.001) but no correlation in the 2009 cycle (r=0.054, n=1306, *P*=0.052). Jerant *et al.* (2012) compared MMI scores between age groups and found applicants aged 19 to 21 performed significantly less well (*P*=0.02) than those aged 25 to 39.

Socioeconomic factors.   Moreau *et al.* (2006) reported that aboriginal interviewers or interviewees made no difference to MMI scores, although Reiter *et al.* (2012) found a negative correlation between aboriginal status and MMI score in both the 2008 (aboriginal: n=45, z=-0.69; other applicants: n=1635, z=0.02; F=20.8, *P*<0.001) and 2009 (aboriginal: n=51, z=-0.31; other applicants: n=1947, z=0.00; F=4.0, *P*=0.04) admissions cycles. Applicants who had graduated from rural high schools achieved significantly lower MMI scores than those from urban high schools in a study of applicants to the University of Manitoba, Canada (4.4 vs 4.6, t=2.96, *P*=0.003; Raghavan *et al.* 2013). Likewise, applicants with rural connections (self-reported) performed worse on MMI than those without (4.4 vs 4.6, t=2.44, *P*=0.015; Raghavan *et al.* 2013). Uijdehaage *et al.* (2011) found no difference in MMI score for (self-reported) economically disadvantaged applicants; findings from Reiter *et al.* (2012) corroborate this.

Predictive validity

The ability of MMIs to predict performance in course and in clinical practice is, naturally, of interest. MMIs are designed to assess non-cognitive attributes in applicants, and therefore would not be expected to predict future academic performance (e.g. how one performs on an assessment of ethical reasoning does not necessarily predict how well they will perform on an assessment of physiology).

Written assessments. McMaster's 2002 MMI pilot (n=45) did not significantly predict performance on the personal progress inventory (PPI, a progress test administered at McMaster school of medicine), whereas undergraduate GPA and autobiographical submission did (Eva *et al.*2004a). However, this MMI which was designed to assess ethical decision-making did predict performance on three domains of the Medical Council of Canada Qualifying Examination (MCCQE) part I: considerations of the legal, ethical and organizational aspects of medicine (CLEO), population health and ethical, legal and organizational aspects of medicine (PHELO) and clinical decision making (CDM). Another study conducted at McMaster matched the MMI scores of 751 applicants (2004 and 2005) to their MCCQE part I (an assessment of clinical knowledge and clinical decision making) total scores, demonstrating the predictive ability of the MMI for cognitive outcomes. The ability of Dundee Medical School's MMI to predict performance on written assessments has been more mixed; associations have been reported for their first cohort on second semester, and second year written assessments but not on first semester written assessment. No associations were demonstrated for written assessments in the second cohort. Tables 5 and 6 (found in the Supplementary Materials Section) illustrate the associations and correlations between performance on MMI and future assessments.

Clinical assessments. The MMI piloted in the 2002 admissions cycle at McMaster (n=45), significantly predicted pre-clinical OSCE performance ($\beta=0.44$, $P<0.01$; Eva *et al.* 2004a), clinical OSCE performance ($\beta=0.4$, $P<0.05$), clerkship performance, measured by end of clerkship ratings assigned by clerkship directors ($\beta=0.7$, $P<0.001$) and clinical encounter cards provided by clinical supervisors ($\beta=0.5$, $P<0.01$; Reiter *et al.* 2007). Further, MMI scores from this population were correlated with number of stations passed on the Medical Council of Canada Qualifying Examination (MCCQE) part II (r=0.35, *P*<0.05; Eva *et al.* 2009). For McMaster applicants interviewed in 2004 or 2005, MMI significantly predicted MCCQE part II total score (β=0.21, *P*<0.001; Eva *et al.* 2012).

The MMI used to select medical students at Dundee n 2009 (n=128) significantly predicted OSCE performance in both semesters of year 1 (semester 1 β=0.18, *P*=0.034; semester 2 β=0.34, *P*<0.001), and in year 2 (β=0.30, *P*<0.001; Husbands & Dowell 2013). The MMI for the 2010 cohort (n=150) again significantly predicted semester 2 OSCE performance (β=0.33, *P*<0.001) but demonstrated no correlation with semester 1 OSCE scores (r=-0.07, *P*=0.55, adjusted for range restriction).

Foley & Hijazi (2013) reported a correlation between the MMI scores of students at Aberdeen dental school (n=75) and end of year assessments consisting of short answer questions and OSCES (r=0.18, *P*=0.001).

Oliver *et al.* (2014) demonstrated that the MMI used for selection of students to University of Calgary Veterinary School (n=60) was correlated with students' performance at building the practitioner-patient relationship (r=0.46, *P*<0.001, corrected for range restriction) and explaining and planning (r=0.28, *P*<0.05, corrected for range restriction) during a standardised clinical communication interview eight months after their MMI (see Tables 5 & 6, found in the Supplementary Materials section).

Discussion

The aim of this paper was to explore, analyse and synthesise the evidence relating to the utility of multiple mini-interviews for selection to undergraduate health programmes.

The purpose of admissions processes in the context of health professions education is to ensure the right applicants are selected both for success within the programme and for performance as healthcare professionals. Making these decisions requires more data than applicants' past academic performance. Other selection tools such as personal references (Dean's letter), individual interviews, and autobiographical statements lack the psychometric properties required to inform the selection of tomorrows' healthcare professionals. Multiple mini-interviews, when designed thoughtfully can aid selection decisions by providing reliable data on applicants' non-cognitive attributes.

Main findings

Overall, MMIs are used to assess applicants' non-cognitive attributes, although, depending on the content of the stations, there may be some overlap with cognitive assessment. When adopting MMIs in admissions, schools need to consider carefully the attributes they value as an institution and use these values to inform their station design. Therefore, what their MMI measures will depend upon the content of their stations, and will vary between schools. Much like an OSCE, station design can be good or poor, and will have effects on the psychometric properties of the MMI. Therefore, like OSCEs, schools need to become skilled and expert in designing MMI stations.

We found clear evidence of MMI feasibility, based on the research output of 20 institutions that have adopted this format for admissions interviews. We also found evidence for the feasibility of conducting an MMI through distance videoconferencing, a potential solution to the high costs for some applicants to travel for interview.

There is ample evidence of MMIs being acceptable to both applicants and interviewers. The majority of applicants prefer MMIs to traditional interviews, but would prefer longer stations. Interviewers perceive MMIs to be a fairer selection tool than panel interviews and most are willing to participate again. The potential for social desirability bias, whereby applicants may give the answers they consider the institutions will want to hear to avoid negative consequences, should be considered when interpreting the evidence regarding acceptability to applicants.

While the reliability of different MMIs varies, findings have been consistently positive with 30 out of 32 cohorts reporting reliability coefficients of greater than 0.6, and one study reporting a Cronbach's alpha of 0.87. The optimal number of stations for MMIs appears to be between 7 and 12, each with one interviewer. Increasing the number of stations has greater impact on the reliability than increasing duration of station or number of raters per station. This should be the focus of resource distribution in MMIs. Our findings regarding the reliability of MMIs are in keeping with the recent systematic review by Knorr & Hissbach (2014)

Studies of validity have reported high face and content validity for MMIs. MMI scores have low correlations with scores on measures of cognitive ability and other measures of non-cognitive performance, suggesting they are measuring different constructs. The MMIs investigated do not appear to produce results that are biased against applicants of any age or gender, nor do they bias applicants from lower socio-economic strata. However, MMIs may disadvantage aboriginal applicants, and no data is yet available about applicants from different ethnic backgrounds. MMIs also appear to disadvantage rural applicants. Certain groups do perform more poorly on some MMIs, but this does not necessarily mean that those MMIs are biased. It should, however, be ensured that the poorer performance is not attributable to any test invalidity for example construct underrepresentation or construct-irrelevant variance (Messick 1995).

Some small associations have been seen between MMI score and in-course written assessment performance. MMIs have been shown to significantly predict performance on practical clinical assessment both during the programme and in postgraduate assessments. Validity is the property of MMIs that is least likely to be able to be generalised between institutions. The validity of any MMI is very much dependent on the context and content of the stations, which in turn will depend on the attributes that the institution values in their applicants.

MMIs are designed to assess specific non-cognitive attributes, but these specific attributes may not be assessed again within the programme. Also, some MMIs are designed to assess multiple traits. These reasons may mean that strong predictive correlations are unlikely to be seen, at least on in-course written and clinical assessments.

While there is explicit evidence of the feasibility, validity and reliability of MMIs as selection tools, the other factor that contributes to an assessments' utility, as defined by Van der Vleuten (1996) is the educational impact of the assessment. There is potential for MMIs to positively influence applicants' thinking about getting into medicine, realising that a wider range of academic and personal qualities are important. This is broader than simply attending an interview skills coaching course, and such preparation might be good for the profession as it exposes applicants from very early on to the wider, more humanistic values in professional practice. No research has yet specifically addressed this issue.

Positives

MMIs appear to be a new instrument for admissions with good feasibility, reliability, and predictive value. Because they do not correlate with any existing instruments, they appear to evaluate new domains. MMIs are the first admissions instrument to demonstrate predictive value into clinical performance at undergraduate and postgraduate levels, although this remains to be more fully examined.

It appears that even when applicants are aware of the content of a station, or have been coached or have previous experience with the MMI, their performance is not affected positively. Given the intense competition for health programme places in an era when station details may find their way on to the websites, an instrument that is not susceptible to privileged diffusion of information is very useful: the performance of the applicant can be ascribed to non-cognitive abilities alone and not to other confounders.

Adopting MMIs makes the admissions process more rigorous. In order to decide on station content, schools should explicitly blueprint to their values.

Issues identified

No studies explicitly discuss any negative consequences of adopting MMIs in to their admissions protocols.

The instrument is only as good as the development of its content – the gap in the literature about details about development hinder the ability to assess and improve quality. For many programmes it is uncertain whether stations are blueprinted to attributes considered important by stakeholders, curriculum domains or graduate outcomes. The finding that aboriginal applicants and rural applicants attain significantly lower scores on MMI is worrying given that these cohorts are already underrepresented in health professions (Dhalla *et al.* 2002; Young *et al.* 2012). Schools with social missions to increase the number of aboriginal physicians, or to produce graduates who will practise in rural areas will have to track the impact of MMIs on their stated missions, and may need to enhance other admissions policies in order to achieve their missions. The possibility of an unconscious bias towards urban situations should be explored. The finding that applicants from lower socioeconomic strata perform equally well is a positive finding, given the struggle to increase the diversity of the class, and the difficulty in identifying these applicants. The scores assigned to applicants by interviewers of different stakeholder groups have low to moderate correlations; it is therefore important that all are represented within an MMI.

Strengths and limitations

This systematic review benefits from a comprehensive search strategy including studies of selection to all health professions with a focus on undergraduate programmes. The international membership of the review group has ensured findings have been interpreted for different international contexts. In a recent systematic review that sought to explore the evidence for the reliability, validity, acceptability and feasibility of the MMI in the selection of health profession students, Pau *et al.* (2013) concluded that MMIs were feasible and acceptable. This BEME systematic review builds on the review conducted by Pau and colleagues through inclusion of further studies and more in-depth synthesis of data. This review also compliments the recent review by Knorr & Hissbach (2014) by focussing specifically on undergraduate programmes and including evidence regarding feasibility and acceptability.

The findings of this review should be interpreted in the context of the limitations of included studies, which have been described under assessment of methodological quality. In brief, they include the presence of small studies in single institutions with relatively short follow-up periods. In addition, there are relatively few non-medicine studies so these are underrepresented in this review. Further, these findings are not necessarily transferable to selection for postgraduate training as applicants for these programmes will have been pre-selected to the programme through undergraduate admission protocols.

Implications for future research

There is ample evidence regarding feasibility and internal reliability of MMIs. Likewise, with the exception of exploring significantly different cultures, or sub-groups of interest, further studies of acceptability to applicants or interviewers are unnecessary. Future research should focus on:

- Exploring the relationship of MMI content to curriculum domains, graduate outcomes, and social missions.
- Investigating test-retest reliability of re-applicants who were not successful in their first round of application.
- Comparing applicants' scores on MMIs in different institutions, particularly if the attributes that the MMIs are designed to assess are similar or disparate.
- Determining the effect of interviewer training on reliability.
- Further exploring the performance of minority groups. Bias should be investigated through differential prediction analyses of different subgroups performance on MMIs, and would do well to be multi-institutional. It would also be of interest to explore whether any groups of minority applicants perceive there to be bias or construct irrelevance in MMIs.
- Continuing to study predictive validity using longer follow up periods with larger cohorts, and using behavioural outcomes such as Multi-Source Feedback. The outcomes that schools use to predict should reflect the content of their MMIs When investigating predictive abilities of MMIs used for selection investigators should correct for range

restriction as there is likely to be no follow up data for applicants with the lowest MMI scores.

- Exploring the educational impact on applicants of adopting MMIs in to selection processes.

Summary

In summary MMIs used as a selection process for health profession programmes appear to have reasonable validity, reliability, and acceptability. The evidence is stronger for face validity, with more research needed to explore content validity and predictive validity. Further research is needed in more institutions in more national contexts and with longer follow up periods to strengthen the evidence base particularly with regard to predictive validity and performance of minority groups.

References

Albanese, M.A., Snow, M.H., Skochelak, S.E., Huggett, K.N. & Farrell, P.M. (2003) Assessing personal qualities in medical school admissions. *Academic Medicine.* 78; 313-321

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. 2nd ed. Washington, DC: American Educational Research Association; 1999.

Bartko, J.J. (1966) The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*. 19; 3-11.

Bloch, R. & Norman, G. (2012) Generalizability theory for the perplexed: a practicalintroduction and guide: AMEE Guide No. 68. *Medical Teacher*. 34; 960-992

Boelen, C. (1999) Adapting health care institutions and medical schools to societies' needs. *Academic Medicine*. 74; S11-S20

Boelen, C. and Woollard, B. (2009) Social accountability and accreditation: a new frontier for educational institutions. *Medical Education*. 43; 887-94

Brown, S.R. (1996) Q methodology and qualitative research. *Qualitative Health Research.* 6; 561-567

Brownell, K., Lockyer, J., Collin, T. & Lemay, J.F. (2007) Introduction of the multiple mini interview into the admissions process at the University of Calgary: acceptability and feasibility. *Medical Teacher.* 29; 394-396

Cameron, A.J. & MacKeigan, L.D. (2012) Development and pilot testing of a multiple mini-interview for admission to a pharmacy degreed program. *American Journal of Pharmaceutical Education.* 76; 1

Cherry, M.G., Fletcher, I., O'Sullivan, H. & Dornan, T. (2014) Emotional intelligence in medical education: a critical review. *Medical Education.* 48; 468-478

Cleland, J., Dowell, J., McLachlan, J., Nicholson, S. & Patterson, F. (2012) *Identifying best practice in the selection of medical students.* General Medical Council available from: http://www.gmc-uk.org/Identifying_best_practice_in_the_selection_of_medical_students.pdf_51119804.pdf

Davis, D., Dorsey, J.K., Franks, R.D., Sackett, P.R., Searcy, C.A. & Zhao, X. (2013) Do racial and ethnic minority group differences in performance on the MCAT exam reflect test bias? *Academic Medicine.* 88; 593-602

Dhalla, I.A., Kwong, J.C., Streiner,D.L., Baddour, R.E., Waddell, A.E. & Johnson, I.L. (2002) Characteristics of first-year medical students in Canadian medical schools. *Canadian Medical Association Journal*. 166; 1029-1035

Dodson, M., Crotty, B., Prideaux, D., Ward, A. & de Leeuw, E. (2009) The multiple mini-interview: how long is long enough? *Medical Education.* 43; 168-174

Donnon, T., Paolucci, E.O. & Violato, C. (2007) The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Academic Medicine*. 82; 100-106

Dowell, J., Lynch, B., Till, H., Kumwenda, B. & Husbands, A. (2012) The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Medical Teacher*. 34; 297-304

Eva, K.W. (2003) On the generality of specificity. *Medical Education.* 37; 587-588

Eva, K.W., Reiter, H.I., Rosenfeld, J. & Norman, G.R. (2004a) The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine.* 79; S40-S42

Eva, K.W., Reiter, H.I. Rosenfeld, J. & Norman, G. (2004b) The relationship between interviewers' characteristics and ratings assigned during a multiple mini-interview. *Academic Medicine*. 79; 602-609

Eva, K.W., Rosenfeld, J., Reiter, H.I. & Norman, G.R. (2004c) An admissions OSCE: the multiple mini-interview. *Medical Education.* 38; 314-326

Eva, K.W., Reiter, H.I., Trinh, K., Wasi, P., Rosenfeld, J. & Norman, G.R. (2009) Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education.* 43; 767-775

Eva, K.W., Reiter, H.I., Rosenfeld, J., Trinh, K., Wood, T.J. & Norman G.R. (2012) Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *JAMA.* 308; 2233-2040

Foley, J.I. & Hijazi, K. (2013) The admissions process in a graduate-entry dental school: can we predict academic performance? *British Dental Journal.* 214; E4

Fortin, Y., Kealey, L., Slade, S. & Hanson, M.D. (2015) Investigating Canadian medical school attrition metrics to inform socially accountable admissions planning. *Medical Teacher.* DOI:10.3109/0142159X.2015.1045847

Gafni, N., Moshinsky, A., Eisenberg, O., Zeigler, D. & Ziv, A. (2012) Reliability estimates: behavioural stations and questionnaires in medical school admissions. *Medical Education.* 46; 277-288

Gordon, M. & Gibbs, T. (2014) STORIES statement: publication standards for healthcare education evidence synthesis. *BMC Medicine.* 12; 143

Griffin, B., Harding, D.W., Wilson, I.G. & Yeomans, N.D. (2008) Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school. *The Medical Journal of Australia.* 189; 270-273

Griffin, B.N. & Wilson, I.G. (2010) Interviewer bias in medical student selection. *Medical Journal of Australia.* 193; 343-346

Griffin, B. & Wilson, I. (2012) Associations between the big five personality factors and multiple mini-interviews. *Advances in Health Sciences Education: Theory and Practice.* 17; 377-388

Gurin, P., Dey, E.L., Hurtado, S. & Gurin, G. (2002). Diversity and higher education: theory and impact on educational outcomes. *Harvard Educational Review*. 72; 330-366

Harris, S. & Owen, C. (2007) Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School. *Medical Education*. 41; 234-241

Hecker, K., Donnon, T., Fuentealba, C., Hall, D., Illanes, O., Morck, D.W. & Muelling, C. (2009) Assessment of applicants to the veterinary curriculum using a multiple mini-interview method. *Journal of Veterinary Medical Education*. 36; 166-173

Hecker, K. & Violato, C. (2011) A generalizability analysis of a veterinary school multiple mini-interview: effect of number of interviewers, type of interviewers, and number of stations. *Teaching and learning in medicine.* 23; 331-336

Husbands, A. & Dowell, J. (2013) Predictive validity of the Dundee multiple mini-interview. *Medical Education.* 47; 717-725

Jerant, A., Griffin, E., Rainwater, J., Henderson, M., Sousa, F., Bertakis, K.D., Fenton, J.J. & Franks, P. (2012) Does applicant personality influence multiple mini-interview performance and medical school acceptance offers? *Academic Medicine.* 87; 1250-1259

Kirkpatrick, D.L. (1967) Evaluation of training. In: Craig R, Bittel L, eds. *Training and Development Handbook.* (pp87-112) New York: McGraw-Hill

Knorrr, M. & Hissbach, J. (2014) Muliple mini-interviews: same concept, different approaches. *Medical Education.* 48; 1157-1175

Kreiter, C.D., Yin, P., Solow, C. & Brennan, R.L. (2004) Investigating the reliability of the medical school admissions interview. *Advances in Health Sciences Education.* 9; 147-159

Kulasegarem, K., Reiter, H.I., Wiesner, W., Hackett, R.D. & Norman, G.R. (2010) Non-association between Neo-5 personality tests and multiple mini-interview. *Advances in health sciences education.* 15; 415-423

Kulatunga-Moruzi, C. & Norman, G.R. (2002) Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teaching and Learning in Medicine.* 14; 34-42

Kumar, K., Roberts, C., Rothnie, I., du Fresne, C. & Walton, M. (2009) Experiences of the multiple mini-interview: a qualitative analysis. *Medical Education.* 43; 360-367

Lemay, J.F., Lockyer, J.M., Collin, V.T. & Brownell, A.K.W. (2007) Assessment of non-cognitive traits through the admissions multiple mini-interview. *Medical Education.* 41; 573-579

Liaison Committee on Medical Education (2012) Functions and structure of a medical school: standards for accreditation of medical education programs leading to the M.D. degree [Online] Available from: http://www.lcme.org/publications/functions.pdf [Accessed: 13/08/2013]

McAndrew, R. & Ellis, J. (2012) An evaluation of the multiple mini-interview as a selection tool for dental students. *British Dental Journal.* 212; 331-335

McAndrew, R. & Ellis, J. (2013) Applicants' perceptions on the multiple mini-interview process as a selection tool for dental and therapy and hygiene students. *British Dental Journal.* 215; 565-570

McCrae, R. M., & Costa, P. T. (1994). A contemplated revision of the NEO five-factor inventory. *Personality and Individual Differences*, 36; 587–596

McManus, I.C., Dewberry, C., Nicholson, S., Dowell, J.S., Woolf, K. & Potts, H.W.W. (2013) Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: meta-regression of six UK longitudinal studies. *BMC Medicine.* 11; 243

Medical Schools Council. (2010) *Guiding principles for the admission of medical students.* [Online] Available from: http://www.medschools.ac.uk/AboutUs/Projects/Documents/Revised MSC Admissions Principles 2010 (final).pdf. [Accessed 13/08/13]

Meredith, K.E., Dunlap, M.R. and Baker, H.H. (1982) Subjective and objective admissions factors as predictors of clinical clerkship performance. *Journal of Medical Education.* 57; 743-751

Messick, S. (1995) Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific enquiry into score meaning. *American Psychologist.* 50; 741-749

Nayer, M. (1992) Admission criteria for entrance to physiotherapy schools: how to choose among many applicants. *Physiotherapy Canada.* 44; 41-46

O'Brien, A., Harvey, J., Shannon, M., Lewis, K. & Valencia, O. (2011) A comparison of multiple mini-interviews and structured interviews in a UK setting. *Medical Teacher.* 33; 397-402

Oliver, T., Hecker, K., Hausdorf, P.A. & Conlon, P. (2014) Validating MMI scores: are we measuring multiple attributes? *Advances in Health Sciences Education.* 19; 379-392

Parry, J., Mathers, J., Stevens, A., Parsons, A., Lilford, R., Spurgeon, R. and Thomas, H. (2006) Admissions processes for five year medical courses at English schools: review. *British Medical Journal.* 332; 1005-1009.

Pau, A., Jeevaratnam, K., Chen, Y.S., Fall, A.A., Khoo, C. & Nadarajah, V.D. (2013) The multiple mini-interview (MMI) for student selection in health professions training – a systematic review. *Medical Teacher*. 35; 1027-1041

Perkins, A., Burton, L., Dray, B. and Elcock, K. (2013) Evaluation of a multiple mini-interview protocol used as a selection tool for entry to an undergraduate nursing programme. *Nurse Education Today.* 33; 465-469

Prideaux, D., Roberts, C., Eva, W., Centeno, A., McCrorie, P., McManus, C., Patterson, F., Powis, D., Tekian, A. & Wilkinson D. (2011) Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 conference. *Medical Teacher*. 33; 215-223

Puddey, I.B. & Mercer, A. (2014) Predicting academic outcomes in an Australian graduateentry medical programme. *BMC Medical Education*. 15; 14-31

Raghavan, M., Martin, B.D., Burnett, M., Aoki, F., Christensen, H., MacKalski, F, Young, D.G. & Ripstein, I. (2013) Multiple mini-interview scores of medical school applicants with and without rural attributes. *Rural and Remote Health.* 13; 2362

Razack, S., Faremo, S., Drolet, F., Snell, L., Wiseman, J. & Pickering J. (2009) Multiple mini-interviews versus traditional interviews: stakeholder acceptability comparison. *Medical Education.* 43; 993-1000

Reiter, H.I., Salvatori, P., Rosenfeld, J., Trinh, K. & Eva, K.W. (2006) The effect of defined violations of test security on admissions outcomes using multiple mini-interviews. *Medical Education.* 40; 36-42

Reiter, H.I., Eva, K.W., Rosenfeld, J. & Norman, G.R. (2007) Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education*. 41; 378-384

Roberts, C., Watson, M., Rothnie, I., Crossley, J., Lyon, P., Kumar, K. & Tiller D. (2008) Factors affecting the utility of multiple mini-interviews in selecting candidates for a graduate-entry medical school. *Medical Education.* 42; 396-404

Roberts, C., Zoanetti, N. & Rothnie, I. (2009) Validating a multiple mini-interview question bank assessing entry-level reasoning skills in candidates for graduate-entry medicine and dentistry programmes. *Medical Education.* 43; 350-359

Roberts, C., Rothnie, I., Zoanetti, N. & Crossley, J. (2010) Should candidate scores be adjusted for interviewer stringency or leniency in the multiple mini-interview? *Medical Education.* 44; 690-698

Rosenfeld, J.M., Reiter, H.I., Trinh, K. & Eva, K.W. (2008) A cost efficieny comparison between the multiple mini-interview and traditional admissions interviews. *Advances in Health Sciences Education.* 13; 43-58

Salovey, P. & Mayer, J. (1990) Emotional Intelligence. *Imagination, Cognition and Personality.* 9; 185-211

Salvatori, P. (2001) Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education.* 6; 159-175

Schuwirth, L.W. & van der Vleuten, C.P. (2010) How to design useful test: the principles of assessment. In Swanwick, T. (ed). *Understanding Medical Education.* (pp195-207) Oxford: Wiley-Blackwell

Schuwirth, L.W. & van der Vleuten, C.P. (2011) General overview of the theories used inassessment: AMEE Guide No. 57. *Medical Teacher.* 33; 783-797.

Sebok, S.S., Luu, K. & Klinger, D.A. (2013) Psychometric properties of the multiple mini-interview used for medical admissions: findings from generalizability and Rasch analyses. *Advances in Health Sciences Education.* 19; 71-84

Till, H., Myford, C. & Dowell, J. (2013) Improving student selection using multiple mini-interviews with multifaceted Rasch modeling. *Academic Medicine*. 88; 216-223

Tiller, D., O'Mara, D., Rothnie, I., Dunn, S., Lee, L. & Roberts, C. (2013) Internet-based multiple mini-interviews for candidate selection for graduate entry programmes. *Medical Education.* 47; 801-810

Uijtdehaage, S., Doyle, L. & Parker, N. (2011) Enhancing the reliability of the multiple mini-interview for selecting prospective health care leaders. *Academic Medicine.* 86; 1032-1039

van der Vleuten, C.P.M. (1996) The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education.* 1; 41-67

van der Vleuten, C.P.M. (2014) When I say… context specificity. *Medical Education.* 48; 234-235

Whitla, D.K., Orfield, G., Silen, W., Teperow, C., Howard, C. & Reede, J. (2003) Educational benefits of diversity in medical school: a survey of students. *Academic Medicine*. 78; 460-466

Yates, J. (2012) When did they leave, and why? A retrospective case study of attrition on the Nottingham undergraduate medical course. *BMC Medical Education* 12; 43

Young, M.E., Razack, S., Hanson, M.D., Slade, S., Varpio, L., Dore, K.L. & McKnight, D. (2012) Calling for a broader conceptualization of diversity: surface and deep diversity in four Canadian medical schools. *Academic Medicine.* 87; 1501-1510.

Table 1

Search terms

"Multiple Mini-interview$"	**AND**	**Admission***	**AND**	**Undergraduat***	**AND**	**"Healthcare education"**
OR MMI		**OR** Applicant*		**OR** Pre-registration		**OR** "Medical Education"
OR OSCE		**OR** Selection		**OR** Initial		**OR** Medic*
OR Station*		**OR** Candidate*		**OR** Universit*		**OR** "Nurs* Education"
OR Multiple				**OR** Student$		**OR** Nurs*
				OR School$		**OR** Physiotherap*
				OR Bachelor$		**OR** Midwif*
				OR Degree		**OR** Dent*
				OR Graduate		**OR** Pharmac*
						OR Veterinary
						OR "Occupational therap*"
						OR Dietetic$
						OR "Allied health"
						OR Audiology
						OR "speech pathology"
						OR "Clinical psychology"

Table 2

Inclusion and exclusion criteria

	Inclusion criteria	**Exclusion criteria**
Population	Applicants to: Undergraduate medicine Undergraduate nursing Undergraduate dentistry Undergraduate pharmacy Undergraduate veterinary Undergraduate midwifery Undergraduate allied health professions	Applicants to: Non-health professions courses Postgraduate courses Postgraduate training programmes
Intervention	Multiple Mini-Interviews	
Outcome	All outcomes	
Study design	Studies which provide primary data	Commentary articles
Publication date	After 2004	Before 2004
Study language	English[13]	Non-English

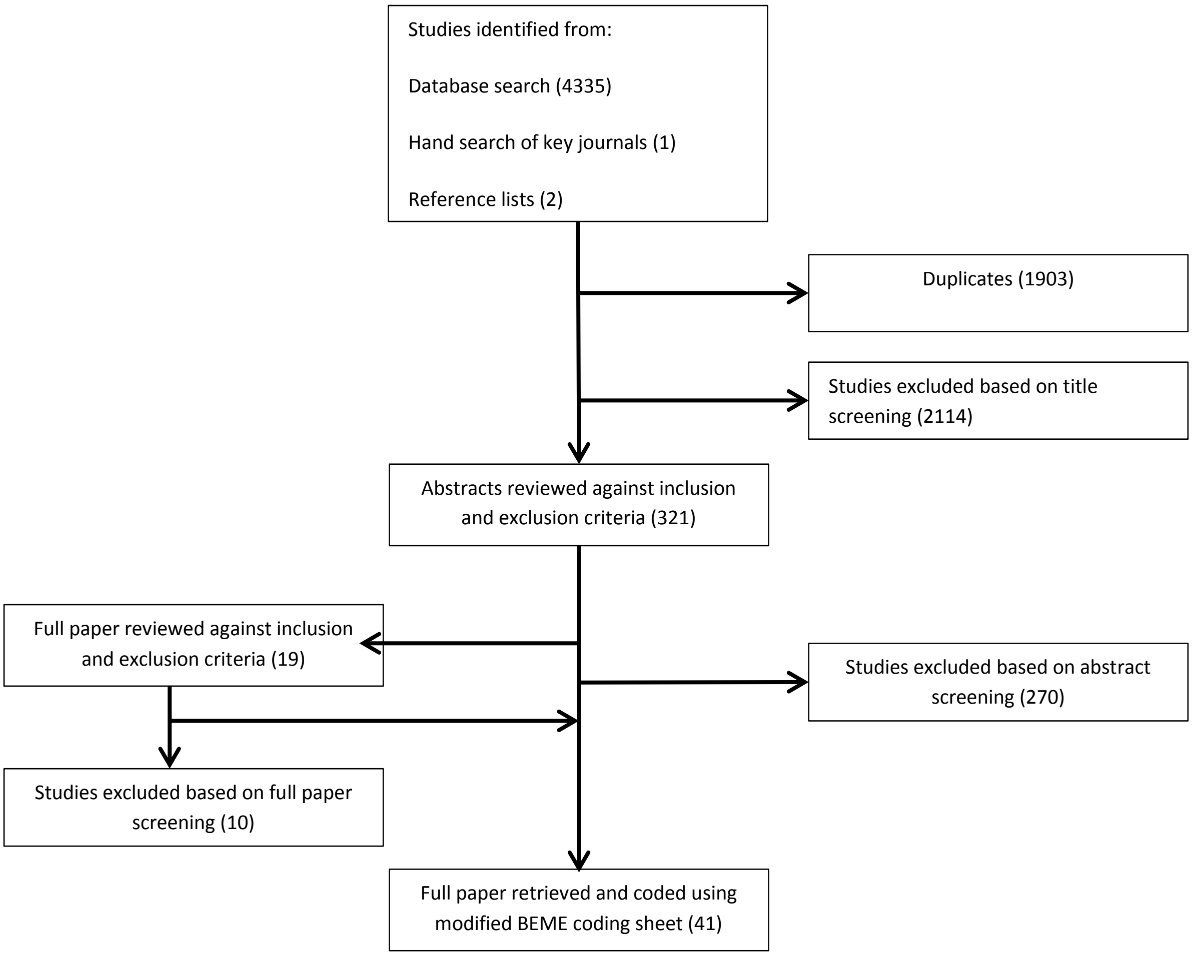

Figure 1.

Flow diagram for included studies

Table 3.

Results from electronic database search

Database:	**Results:**
Medline	821
Allied and Complementary Medicine (AMED)	0
British Nursing Index (BNI)	6
Cumulative Index to Nursing and Allied Health (CINAHL)	220
EMBASE	1750
Education Resources Information Centre (ERIC)	48
Web of Science	989
British Education Index (BEI)	30
PsycINFO	189
Applied Social Sciences Index and Abstracts (ASSIA)	247
Australian Education Index (AEI)	5
Health Business Elite	9
Health Management Information Consortium (HMIC)	21
Total (before duplicates removed):	4335
Duplicates	1903
Total (after duplicates removed):	**2432**

Table 4.

Interstation reliability

Programme	Reference	No. of stations	Raters per station	n	Method	Coefficient
School of Medicine, University of Calgary 2009	Cameron 2012	10	1	30	Internal reliability (ICC)	0.77
Veterinary Sciences, University of Calgary 2008	Hecker 2009	5	2	110	Cronbach's α	0.87
Veterinary Sciences, University of Calgary 2009	Hecker 2011	7	2	103	G study	0.79
Veterinary Sciences, University of Calgary (year not reported)	Oliver 2014	8	2	186	G study	0.73
School of Medicine, UCLA 2009	Uijtedhaage 2011	12	1	76	G study	0.59
School of Medicine, UCLA 2010	Uijtedhaage 2011	12	1	78	G study	0.71
Davis School of Medicine, University of California 2010	Jerant 2012	10	1	444	Cronbach's α	0.68
School of Medicine, Deakin University 2008	Dodson 2009	10	1	175	G study	0.78
School of Medicine, University of Dundee 2007	Dowell 2012	4	1	473	Cronbach's α	0.66
School of Medicine, University of Dundee 2009	Dowell 2012	10	1	452	Cronbach's α	0.70
School of Medicine, University of Dundee 2010	Dowell 2012	10	1	477	Cronbach's α	0.69
School of Medicine, McMaster University pilot 2002	Eva 2004c	12	1	18	G study	0.81
School of Medicine, McMaster University 2002	Eva 2004c	10	1	117	G study	0.65

School of Medicine, McMaster University 2003	Eva 2004b	9	2	57	G study	0.78
School of Rehabilitation, McMaster University 2003	Eva 2009	3	2	57	G study	0.51
School of Medicine, McMaster University 2004	Eva 2009	12	1	379	G study	0.70
School of Rehabilitation, McMaster University 2004	Eva 2009	7	1	380	G study	0.70
School of Medicine, McMaster University 2005	Eva 2009	12	1	696	G study	0.69
MIRKAM[1] 2006	Gafni 2012	8	1	413	G study	0.74
MIRKAM 2007	Gafni 2012	8	1	451	G study	0.68
MIRKAM 2008	Gafni 2012	8	1	533	G study	0.63
MIRKAM 2009	Gafni 2012	8	1	626	G study	0.61
MOR[2] 2006	Gafni 2012	8	1	565	G study	0.73
MOR 2007	Gafni 2012	8	1	645	G study	0.76
MOR 2008	Gafni 2012	9	1	695	G study	0.79
MOR 2009	Gafni 2012	9	1	757	G study	0.77
School of Medicine St George's University London (4 year graduate entry course) 2009	O'Brien 2011	8	1	21	Cronbach's alpha	0.69
School of Medicine, St George's University London (5 year undergraduate course) 2009	O'Brien 2011	8	1	26	Cronbach's α	0.73
School of Medicine, University of Sydney 2006	Roberts 2008	8	1	485	G study	0.70
School of Medicine,	Tiller 2013	9	1	293	G study	0.76

[1] MIRKAM is a national assessment centre in Israel, run by the Hebrew University in Jerusalem in collaboration with the National Institute for Testing and Evaluation, that consists of eight mini interviews and two questionnaires
[2] MOR is a national assessment centre in Israel, run by three universities in Israel in collaboration with the National Institute for Testing and Evaluation, that consists of nine mini interviews and two questionnaires

University of Sydney 2011 iMMI						
School of Medicine, University of Sydney 2011 MMI	Tiller 2013	9	1	571	G study	0.70
School of Medicine, Queen’s University 2011	Sebok 2013	8	2	485	G study	0.68

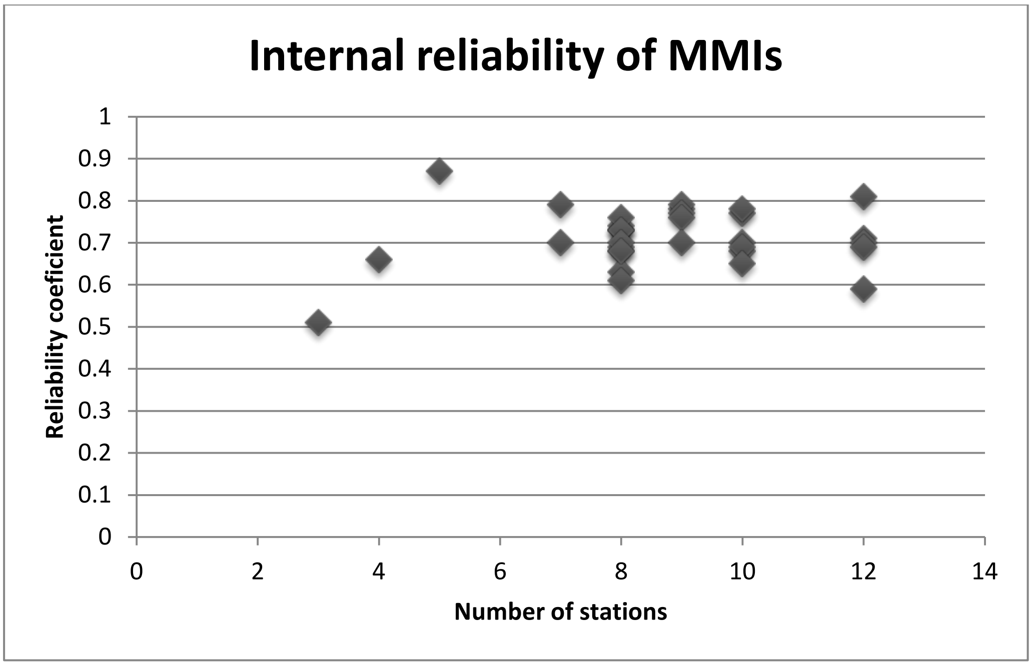

Figure 2.

Internal reliability of MMIs

Table 5

Associations between MMI scores and future assessments

School and admissions cycle	n	Outcome measure	Standardised β	p
Written assessments				
School of Medicine, McMaster University 2002	45	MCCQE part I (total score)	>0.3	<.06
School of Medicine, McMaster University 2002	45	CLEO (domain of MCCQE part I)	>0.4	<.01
School of Medicine, McMaster University 2002	45	PHELO (domain of MCCQE part I)	>0.4	<.01
School of Medicine, McMaster University 2002	45	CDM (domain of MCCQE part I)	0.35	<.05
School of Medicine, McMaster University 2004 & 2005	751	MCCQE part I total score	0.12	<.001
School of Medicine, University of Dundee 2009	128	Year 1 semester 1 written assessment	n/s	n/s
School of Medicine, University of Dundee 2009	128	Year 1 semester 2 written assessment	0.26	.002
School of Medicine, University of Dundee 2009	128	Year 2 written assessment	0.21	.018
School of Medicine, University of Dundee 2010	150	Year 1 semester 1 written assessment	n/s	n/s
School of Medicine, University of Dundee 2010	150	Year 1 semester 2 written assessment	n/s	n/s
Practical assessments				
School of Medicine, McMaster University 2002	45	Pre-clerkship OSCE performance	0.44	<.01
School of Medicine, McMaster University 2002	45	Clerkship OSCE performance	0.4	<.05
School of Medicine, McMaster University 2002	45	Clerkship ratings	0.7	<.001
School of Medicine, McMaster University 2002	45	Clinical encounter cards	0.5	<.01
School of Medicine, McMaster University 2004 & 2005	623	MCCQE part II total score	0.21	<.001
School of Medicine, University of Dundee 2009	128	Year 1 semester 1 OSCE	0.18	.034
School of Medicine, University of Dundee 2009	128	Year 1 semester 2 OSCE	0.34	<.001
School of Medicine, University of Dundee 2009	128	Year 2 OSCE	0.30	<.001
School of Medicine, University of Dundee 2010	150	Year 1 semester 1 OSCE	n/s	n/s

School of Medicine, University of Dundee 2010	150	Year 1 semester 2 OSCE	0.33	<.001

Table 6

Correlations between MMI score and future assessments

School and admissions cycle	n	Outcome measure	r	p
School of Medicine, McMaster University 2002	45	Number of stations passed on MCCWE part II	0.35	<.05
School of Dentistry, University of Aberdeen 2008 to 2011	75	End of year assessments (knowledge & skills)	0.18	.001
Veterinary Sciences, University of Calgary (Year not reported)	60	Standardised clinical communication interview; building the practitioner-patient relationship	0.46	<.001
Veterinary Sciences, University of Calgary (Year not reported)	60	Standardised clinical communication interview; explaining and planning	0.28	<.05