

Assessing noncognitive domains of respiratory therapy applicants: Messick's framework appraisal of the multiple mini-interview

Marco Zaccagnini, RRT/CCAA, BHSc, MSc (c)

M Zaccagnini. Assessing noncognitive domains of respiratory therapy applicants: Messick's framework appraisal of the multiple mini-interview. *Can J Respir Ther* 2019;55:31–35. doi: 10.29390/cjrt-2019-002.

Educators who assess incoming applicants into a health professional training program are looking for a wide array of cognitive and noncognitive skills that best predict success in the program and as a future practicing professional. While aptitude tests generally measure cognitive skills, noncognitive constructs are more difficult to measure appropriately. The traditional method of measuring noncognitive constructs has been the panel interview. Panel interviews have been described as inconsistent in measuring noncognitive domains and consistently reported as unreliable and susceptible to bias. An alternate interview method used in many health professions schools is the multiple mini-interview (MMI) that was specifically designed to assess noncognitive domains in health professions education. This paper discusses the purpose of using the MMI, how the MMI is conducted, specific domains of focus for the MMI, and the feasibility of creating an MMI. Finally, the paper uses Messick's framework on validity to guide the consideration of the MMI.

Key Words: *multiple-mini interview; admission; education; interview; educational measurement*

INTRODUCTION

The level of education of applicants to respiratory therapy programs can range from basic high school diplomas to international medical degrees. Initially, the applicants are screened based on prerequisite foundation courses and ranked by estimates of academic success (i.e., grade point average) that have been associated with academic success [1, 2]. The top-ranked applicants are expected to present themselves to the educational institution and take part in a panel interview. The panel interview is the long-standing assessment method used at most traditional health professions schools to assess candidates' personality characteristics [3]. In respiratory therapy, the characteristics or constructs the educators are assessing include many noncognitive traits, personality traits, and affective skills such as adaptability, social skills, professionalism, and empathy. Such personality characteristics are important in determining who will succeed in training, they have been associated with professional development and career satisfaction, and are also linked to improved patient care outcomes [4, 5]. Such characteristics outline the basic expectations of entry-to-practice Registered Respiratory Therapists (RRTs) [6].

The typical panel interviews in health professions education admissions are comprised of practicing health professionals, faculty members, students, and members of the community. Panel interviewers engage in an unstructured or semi-structured dialogue with a candidate for varied amounts of time, typically ranging from 15 to 60 min [7, 8]. There have been many inconsistencies reported in academic literature regarding the use of a panel interview to measure noncognitive domains; it has been reported to be consistently unreliable and susceptible to bias [9–11]. The question then arises, how can we improve the selection process to best assess applicants' noncognitive characteristics to choose

candidates who will likely best contribute to the medical society and flourish as health professionals?

An alternative method to the traditional panel interview that has been adopted by many health professions is the multiple mini-interview (MMI), created at McMaster University, Hamilton, Ontario, in 2002 [12]. The MMI generally consists of 10 stations, with one interviewer at each station. The process is as follows: each interviewee has 2 min to read a question, followed by the interviewee providing their response for the remaining 6–8 min. MMIs thus allow applicants the opportunity to interact with multiple interviewers [9]. Available systematic reviews of the MMI process in medical schools have reported on its reliability, acceptability, and feasibility of implementation [10] along with the ability to predict clinical performance of successful applicants [13].

METHODOLOGY

Below we consider the MMI as an alternative to traditional panel interviews and summarize the domains the MMI assesses. Messick's framework on validity [14] will guide consideration of MMIs. This will include examination of prior works that pertained to reliability, validity, efforts to minimize bias and utility, and the feasibility of integrating the MMIs into the application process of a respiratory therapy training program.

MMI tool

Purpose

Noncognitive abilities such as professionalism and empathy of future clinicians contribute to both academic and clinical performance outcomes and must, therefore, be adequately explored before the applicant's acceptance into the health profession [15]. An extensive review by

Department of Anesthesia & Critical Care, McGill University Health Centre, Montréal, QC, Canada

Correspondence: Marco Zaccagnini, RRT/CCAA, BHSc, MSc (c), Department of Anesthesia & Critical Care, McGill University Health Centre, Royal Victoria & Montréal General Hospital, 1001 Decarie Blvd, Montréal, QC H4A 3J1, Canada. Tel: (514) 969-7204. Email: marco.zaccagnini@mcgill.ca

Published online at <https://www.cjrt.ca> on 13 March 2019



This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited and the reuse is restricted to noncommercial purposes. For commercial reuse, contact editor@csrt.com

Morris [15] highlighted many shortcomings with using the panel style of interview. Weaknesses of the panel interview include the incurred costs because of the substantial time commitment required of staff and other interviewers as well as the expense and inconvenience to the applicants from the travelling, accommodations, and time lost from school or work [16]. The high subjectivity of interviewers' scores, the varying structure of interviews, the lack of interviewing training and the accompanying biases have a major influence on the validity and reliability of traditional interviews to assess noncognitive features of applicants accurately. The MMI was developed by McMaster's School of Medicine in 2002 to address the concern that panel interviews did not adequately predict the noncognitive performances of medical school applicants. Eva et al. [17] developed the MMI process in an attempt to dilute the impact of individual examiners and allow for a more standardized and generalizable performance rating to be assigned to each applicant [9].

Procedure

The logistics of MMIs are affected by the educational resources available (i.e., space, cost). The most common method of conducting MMIs consists of candidates rotating through multiple stations. Each interview station assesses a single, predetermined construct domain that reflects the objectives of the institution. Each station is kept short (8 min. or less) and is conducted simultaneously in a circuit. With each toll of the bell, the participant moves onto the next interview session in the circuit [18]. Typically, a single interviewer per interview station is sufficient to assess a candidate on the construct [17]. Each interviewer stays in the same interview room throughout the day, as candidates rotate through. The interviewer then scores each candidate based upon the same interview scenario throughout the interview process (Figure 1).

Domains

The noncognitive domains chosen to become the focus of McMaster's pilot MMI included communication skills, ethical decision-making, critical thinking, collaboration, and knowledge of the health care system. These domains were selected merely based on the preferences of the research team [19]. Other educational institutions chose domains that best represented the school's mission statement or cultural context [20, 21]. As the MMI became formally adopted in McMaster's medical school application process, there was an increased motivation by McMaster to outline and validate the desired noncognitive domains systematically. Reiter et al. [19] chose to blueprint the desired noncognitive

domains against three prominent medical education frameworks that emphasized the essential competencies of a physician. The Accreditation Council for Graduate Medical Education, the CanMEDS 2000, and the Core Committee of the Institute for International Medical Education all provided outlines, which included competencies that overlapped significantly with one another. These roles and competence domains derived from a much larger number of cognitive and personal qualities cited in the literature as desirable when choosing medical students [19]. Reiter then narrowed down the competencies to a series of seven characteristics and surveyed key stakeholders to assess which were most relevant for McMaster's medical students [19]. The conclusion supported that stakeholders ranked ethical nature, communication, and intelligence as the three most desirable domains in health professionals. The authors then used these domains as a base to build MMI scenarios for McMaster's undergraduate medical program. The authors cautioned that the domains outlined in their study are best used as a guide to developing admissions protocols.

The objective of the MMI is to reveal and assess noncognitive domains that are particularly crucial to the educational institution or health profession in question [19]. The RRT curriculum is founded on many domains of competency that are considered essential components of practice. Domains such as professional behaviour, collaborating with the interdisciplinary team, and effective communication are skills that professionals use in daily practice and significantly overlap with those outlined by the three prominent medical education frameworks listed prior, particularly CanMEDS [6, 22]. To use the MMI to assess future RRT candidates appropriately, key stakeholders (i.e., practicing RRTs, educators, managers etc.) would have to blueprint the MMI stations to mirror the outlined, desirable, noncognitive characteristic of entry-to-practice RRTs [6].

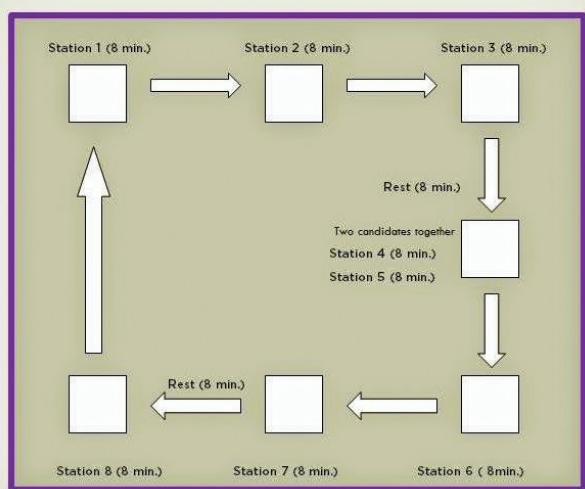
MMI psychometric properties

Validity

Validity is a dynamic process whereby we construct arguments with either weak or strong evidence to support our assessment decisions [23]. Similar to hypothesis testing, validity evidence is collected to support or refute the validity hypothesis [24]. There is conflicting evidence surrounding the use of panel interviews as a valid measure of assessing non-cognitive constructs in health professions education [1]. While some investigators report positive correlations between pre-admission interview ratings and in-course academic performance, others conclude that the interview is not predictive of success [15]. Salvatori [14] provided a comprehensive review of a range of admission tools used by health education professionals and found that unstructured interviews provided conflicting predictive validity of outcomes. When considering the exuberant cost involved in training today's health professionals and the public's expectation of professional accountability [25, 26], admission committees must include evidence-based selection criteria to ensure that the best-qualified candidates are selected. Qualified candidates must encompass both academic strength and clinically relevant constructs. Examples of these constructs include the previously outlined noncognitive domains, personality traits, and affective skills. While panel interviews had shown a slight improvement in inter-rater reliability and validity scores when interviewers were scripted and had more structure [27, 28], the MMI further improves validity scores because of their format. The MMI format was specifically designed to assess intangible constructs. To argue for evidence of the validity of the MMI, Messick's [14] framework will be used to analyze the construct validity of the MMIs. Messick's [14] framework for construct validity includes content validity, response validity, internal structure validity, relationship to other variables, and consequential validity.

Content evidence of validity refers to the logical-empirical relationship of content tested to the achievement domain [29]. MMI stations that are constructed around the academic institution's "blueprints" or desired competencies have a higher content validity. Blueprinting ensures there is an optimal match between the MMI testing stations and the desired characteristics to be tested from the candidates [9]. To ensure

FIGURE 1
Multiple mini-interview set-up.



content validity for selecting RRTs, the MMI stations would have to be constructed alongside our professional blueprint that specifically outlines the desired core competencies of entry-to-practice RRTs [6].

Response process validity is defined as evidence of data integrity. All sources of error associated with the test administration should be controlled or eliminated as much as possible [29]. Examples of areas of response process include quality control of assessors (i.e., how to assess individuals and what is the purpose of assessment) [30] and ensuring student familiarity with the format [31]. Stakeholders who are involved in the RRT interviews must be made aware of what each MMI scenario is attempting to assess. Additionally, incoming applicants must also be made aware of the MMI interview structure, so they may appropriately prepare for the interview.

Internal structure, as a source of validity evidence, relates to the statistical or psychometric characteristics of the examination questions or performance prompts, the scale properties (i.e., reproducibility and generalizability), and the psychometric model used to score and scale the assessment. Reliability is defined as the reproducibility of assessment data or scores, over time or occasions [29]. Reliability does not lay in the test itself, but rather, the outcome of the test. When we look at the reliability of the MMI, we are assessing if the MMI was able to consistently differentiate the constructs that were tested in each station [32]. Some examples of estimating reliability include test–retest reliability and inter-rater reliability.

Test–retest reliability assesses the degree to which test scores are consistent from one test administration from the next. Simply put, a candidate's response should score the same on the same question if given at different time intervals. Although test–retest is fundamental in reliability estimates, providing a test to candidates and then retesting them with a similar test is logistically challenging to carry out [32]. The MMI has the ability to test the same applicant on two parallel forms of a test for a particular construct indicating good test–retest reliability [9]. Gafni et al.'s [33] research on failed applicants who retook the same MMI in the same institution also reported good test–retest correlation, suggesting that MMI performance does not vary between attempts [33].

The inter-rater reliability is another measure of validity most useful when calculating scores of behaviours by rater observation. Inter-rater reliability assesses the degree to which test scores are consistent when measurements are measured by different raters with the same, standardized method [32]. Unstructured, nonstandardized panel interviews are the most common method chosen to conduct candidate interviews. The panel interview is void of any standardization, and inter-rater reliability is grossly inconsistent [34]. The existent literature on unstructured interviews indicates ambiguous results [1]. The panel interview inter-rater reliability consistently reports lower than acceptable averages, and inconsistencies may be a result of potential interview bias. The literature indicates that interviewers consistently scored candidates higher when personality preferences matched [35]. In an attempt to control for interviewer bias, it is best to employ the individual opinions of multiple observers.

The inter-rater reliability increases when raters were carefully selected, they received interview training, there were multiple observers, and interview questions were blueprinted against the medical school's goals or objectives [36]. These strategies are the foundational pillars of the MMI format, which assisted in the widespread acceptability of the MMIs. The early MMIs inter-rater reliability was initially called into question because of the single interviewer per station format, which may still be at risk for interviewer bias [1]. Eva et al. [13] recognized the bias and used a statistical theory for evaluating reliability of behavioural measurements (Generalizability Statistical Theory) [37] to conclude that increasing both the number of raters and the number of stations significantly improves the MMI's reliability. Multiple-rater MMI reliability has consistently been reported in other contexts as well [20, 30]. Finally, another source of internal structure validity is validating the item difficulty–discrimination. Item difficulty is the measure of the proportion of examinees who answered the item correctly and item discrimination is the measure of how well an item is able to distinguish between examinees who are knowledgeable and who are not [38]. One such strategy is to

TABLE 1

Summary of Messick's framework for construct validity

Messick framework	Multiple mini-interviews (MMIs)
Content validity	<ul style="list-style-type: none"> • MMI stations are created based on the academic institution's desired characteristics • Each station tests a single construct
Response process validity	<ul style="list-style-type: none"> • Data integrity • Controlled potential sources of error (test administration) • Interviewers know what construct they are assessing • Applicants aware of MMI structure
Internal structure	<ul style="list-style-type: none"> • Statistical/psychometrics of questions or prompts <ul style="list-style-type: none"> ◦ Test–retest reliability ◦ Inter-rater reliability ◦ Item difficulty–discrimination
Relationship to other variables	<ul style="list-style-type: none"> • Convergent correlation • Divergent validity
Consequential validity	<ul style="list-style-type: none"> • Impact of results on applicants • Impact of results on stakeholders • Impact of results on public

create a quality-controlled question bank to help assessment designers guarantee a degree of validity [39].

When high-stakes decisions are to be made using an instrument, there must be a high degree of reliability to make defensible decisions. In the case of selecting proficient candidates for a competitive health education program, admission committees must accurately assess all desired constructs reliably. Panel interviews have been shown to have ambiguous results, whereas the MMIs held more promise with stronger reliability [40, 41]. Modifying the MMIs' format to include a larger number of assessors and scenarios could further improve the reliability.

Relationship to other variable validity refers to how well the MMI construct scenarios correlates with other variables. That is, convergent correlations are when a different type of assessment method is used to assess the same construct, and both results are similar. In contrast, divergent validity refers to another assessment method that evaluates a different construct unrelated to the first. In noncognitive measures, certain studies of MMIs have suggested a moderate correlation between scenarios testing the same construct [33, 42]. Additionally, MMI scores have low correlations with scores on traditional measures of cognitive ability (i.e., grade point average) suggesting they measure different constructs [42].

Finally, the consequential area of validity refers to the impact on examinees from the assessment scores, decisions, and outcomes, and the impact of assessments on teaching and learning. The consequences affect many stakeholders including examinees, faculty, patients, and society. These consequences can be positive or negative, intended or unintended [29]. A test must reliably safeguard the public from individuals who will cause harm if they can proceed and practice. Likewise, the exam should not fail and withhold appropriate applicants who will eventually provide a much-needed service to the public as health professionals.

Consequences of the MMIs in reported literature have consistently shown that those accepted in medical school by the MMIs outperform applicants accepted by traditional admission processes in the educational context and in high-stakes licensure exams [20, 40, 41]. The commonly used, unstructured interview that the RRT admission committee uses offers little predictive validity of objectively assessing future performance of applicants [43, 44] (Table 1).

FEASIBILITY

When attempts are made to implement a novel tool within practice or an existing curriculum, a significant discussion point is the tool's feasibility. Feasibility is often a point of concern for many stakeholders, predominantly those with administrative duties. Implementing the MMIs at face value seems more resource-intensive than the panel interview counterpart. Although there is some truth that the initial creation of stations,

blueprinting institutional objectives, creating evaluation forms, and simulation-based scenarios take longer, each component can subsequently be modified and reused in the following years. Rosenfeld et al. [18] provided a synopsis of the cost-efficiency of MMIs relating to the administrative duties, miscellaneous expenses, interview training, assessors, staff, infrastructure requirements, and their experience creating the McMaster MMI process. Eva et al. [9] further elaborated on the McMaster process of interviewing 400 medical school applicants; excluding the expenses generated for the initial project framework and organization, the reported costs of the actual MMI operation on the interview day is about \$35 per candidate. Hissbach et al. [45] discussed the slight modification of their existing MMI process that cut costs by approximately 50%. Similar results exist in other medical domains, Corelli et al. [46] noted a cost-effectiveness per candidate interviewed for pharmacy school of \$75.30 with the MMI format compared with \$136.34 with a standard interview [46].

Regarding logistics, Brownell et al. [47] reported many advantages of including the MMIs compared with panel interviews in medical school admissions. With the MMIs, interviews were completed in 2 days rather than the average 8 days. They required less personnel and it did not cost any more compared with previous years. Finlayson et al. [48] also reported on the success of implementing MMIs in a medical residency training process, claiming the process is cost-effective, required less time from volunteers, held good inter-rater reliability, and participants simply enjoyed the process more. While challenges and the absolute cost will vary according to the education institutions, frameworks exist to guide educators regarding the logistical requirements and anticipated budgets [18].

CONCLUSION

Noncognitive abilities are an essential aspect for RRTs and all other health care professionals. Rather than assuming candidates will eventually develop these skills, admission committees are well positioned to assess these noncognitive domains early on and select appropriate candidates who will exemplify these desired traits. Selection committees often rely on a panel interview to expose noncognitive abilities. Panel interviews are consistently reported as unreliable in detecting noncognitive abilities and predicting future performance. The MMIs are a commonly used alternative interview tool to better assess noncognitive domains of a candidate with strong validity and reliability.

Despite the strong validity and reliability, admission committees cannot simply duplicate an MMI from one school or health domain and expect it to work within a different context. The MMIs commonly reported validity and reliability is context specific and requires a “ground-up” approach when implementing. Educators interested in including MMIs within their admission criteria are expected to actively validate their (context-specific) MMI validity. Validity is a dynamic process that uses an exam tool to collect evidence to support or refute the assessor’s hypothesis about an applicant. Using Messick’s [28] framework to assess the MMIs, construct validity has repeatedly been favourable, and the MMI’s feasibility of curricular integration has been reported to be relatively easy. For the MMIs to be justly included in assessing RRT applicants, the station scenarios and the constructs tested must be tailored and blueprinted against the RRT’s educational institutions goals, objectives, and the desired noncognitive abilities of an entry-to-practice RRT.

Developing and implementing MMIs for respiratory therapy schools in Canada is worthy of consideration. To facilitate this one might establish a multicentered working group responsible for blueprinting MMI stations to RRT entry-to-practice constructs, developing training material for academic institutions and interviewers, facilitating pilot implementation at certain institutions, and of course evaluating the interview validity. We must be vigilant in continuously assessing and modifying the RRT-specific MMIs to maintain the validity in testing incoming applicants. Finally, the MMI must adhere to the institution’s logistical allowance to be sustainable. The adoption of an RRT-specific MMI for assessing applicant’s noncognitive abilities could be a unique way to avoid the pitfalls associated with

traditional panel interviews and could lead to more reliable and professional students who, ultimately, will become strong professionals.

FUNDING

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

COMPETING INTERESTS

The author has completed the ICMJE uniform disclosure form at <http://www.icmje.org/conflicts-of-interest/> and declares: no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work.

REFERENCES

1. Prideaux D, Roberts C, Eva K, et al. Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 conference. *Med Teach* 2011;33(3):215–23. doi:10.3109/0142159X.2011.551560.
2. York TT, Gibson C, Rankin S. Defining and measuring academic success. *Pract Assess Res Eval* 2015;20(5):1–20. Available at: <http://pareonline.net/getvn.asp?v=20&n=5>. Accessed March 1, 2018.
3. BeMo. Panel, traditional or conversational med school interview tips. 2018. Available at: <https://bemoadademicconsulting.com/blog/how-to-ace-panel-traditional-conversational-med-school-interview> (Accessed March 10, 2018).
4. Levy JJ, Richardson JD, Lounsbury JW, Stewart D, Gibson LW, Drost AW. Personality traits and career satisfaction of health care professionals. *Individ Differ Res* 2011;9(4):238–49. doi:10.1080/13678860802261215.
5. Koenig TW, Parrish SK, Terregino CA, Williams JP, Dunleavy DM, Volsch JM. Core personal competencies important to entering students’ success in medical school: What are they and how could they be assessed early in the admission process? *Acad Med* 2013;88(5):603–13. doi:10.1097/ACM.0b013e31828b3389.
6. The National Alliance of Respiratory Therapy Regulatory Bodies. National Competency Framework. 2016. Available at <http://narrtb.ca/download/ncf-part-i-entry-to-practice-2016/?wpdmdl=33>. Accessed March 5, 2018.
7. Switzer-Mcintyre S. Clinician’s commentary on van der Spuy et al. *Physiother Canada* 2016;68(2):186–7. doi:10.3138/ptc.2015-24-CC.
8. Toronto Career Centre. Interview techniques. 2015. Available at: https://www.studentlife.utoronto.ca/sites/default/files/Med_School_Interviews.pdf (Accessed March 10, 2018).
9. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: The multiple mini-interview. *Med Educ* 2004;38(3):314–26. doi:10.1046/j.1365-2923.2004.01776.x.
10. 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):1027–41. doi:10.3109/0142159X.2013.829912.
11. Axelson RD, Kreiter CD. Rater and occasion impacts on the reliability of pre-admission assessments. *Med Educ* 2009;43(12):1198–202. doi:10.1111/j.1365-2923.2009.03537.x.
12. McMaster University. How we select: Multiple mini-interview (MMI). 2015. Available at: [http://mdprogram.mcmaster.ca/md-program-admissions/how-we-select/multiple-mini-interview\(mmi\)](http://mdprogram.mcmaster.ca/md-program-admissions/how-we-select/multiple-mini-interview(mmi)) (Accessed February 1, 2018).
13. 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 Suppl):S40–2. doi:10.1097/00001888-200410001-00012.
14. Messick S. The psychology of educational measurement. *J Educ Meas* 1984;21(3):215–37.
15. Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Adv Heal Sci Educ* 2001;6(2):159–75. doi:10.1023/A:1011489618208.
16. Morris JG. The value and role of the interview in the student admissions process: A review. *Med Teach* 1999;21(5):473–81. doi:10.1080/01421599979149.
17. 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):602–9. doi:10.1097/00001888-200406000-00021.

18. Rosenfeld JM, Reiter HI, Trinh K, Eva KW. A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Adv Heal Sci Educ* 2008;13(1):43–58. doi:10.1007/s10459-006-9029-z.
19. Reiter HI, Eva KW. Reflecting the relative values of community, faculty, and students in the admissions tools of medical school. *Teach Learn Med* 2005;133417(17):4–8. doi:10.1207/s15328015tlm1701_2.
20. 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. doi:10.3109/0142159X.2012.652706.
21. Pau A, Chen YS, Lee VKM, Sow CF, De Alwis R. What does the multiple mini interview have to offer over the panel interview?. *Med Educ Online* 2016;1(21):1–5. doi:10.3402/meo.v21.29874.
22. Frank J, Snell L, Sherbino J, Editors. *CanMEDS 2015 – Leader. CanMEDS physician competency framework*. 2015. Available at: http://www.royalcollege.ca/portal/page/portal/rc/common/documents/canmeds/framework/canmeds2015_framework_series_IV_e.pdf. Accessed March 15, 2018.
23. Sibbald M. Issues with validity. 2018. Available at: <https://avenue.cllmcmaster.ca/d2l/le/content/227032/viewContent/1845650/View>. Accessed March 15, 2018.
24. Cook DA, Hatala R. Validation of educational assessments: A primer for simulation and beyond. *Adv Simul* 2016;1(1):31. doi:10.1186/s41077-016-0033-y.
25. Wass V, Vleuten C Van Der, Shatzer J, Jones R. Assessment of clinical competence. *Med Educ Quartet* 2001;357:945–9.
26. Official Languages Community Development Bureau Overview of the cost of training health professionals 2008. Health Canada, Ottawa. doi:978-1-100-11971-7.
27. Edwards J, Johnson E, Molitor J. The interview in the admission process. *Acad Med* 1990;65(3):167–77.
28. Powis D. Select medical students. *BMJ* 1998;317(7166):1149–50. doi:10.3109/01421597909010576.
29. Downing SM. Validity: On the meaningful interpretation of assessment data. *Med Educ* 2003;37(9):830–7. doi:10.1046/j.1365-2923.2003.01594.x.
30. Campagna-Vaillancourt M, Manoukian J, Razack S, Nguyen LHP. Acceptability and reliability of multiple mini interviews for admission to otolaryngology residency. *Laryngoscope* 2014;124(1):91–6. doi:10.1002/lary.24122.
31. 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 Pharm Educ* 2015;79(4):1–10. doi:10.5688/ajpe79453.
32. Meghan M. Reliability & generalizability-theory. 2018. Available at: <https://avenue.cllmcmaster.ca/d2l/le/content/227032/viewContent/1845652/View>. Accessed: March 21, 2018.
33. Gafni N, Moshinsky A, Eisenberg O, Zeigler D, Ziv A. Reliability estimates: Behavioural stations and questionnaires in medical school admissions. *Med Educ* 2012;46(3):277–88. doi:10.1111/j.1365-2923.2011.04155.x.
34. Kreiter CD, Yin P, Solow C, Brennan RL. Investigating the reliability of the medical school admissions interview. *Adv Heal Sci Educ* 2004;9(2):147–59. doi:10.1023/B:AHSE.0000027464.22411.0f.
35. Quintero AJ, Segal LS, King TS, Black KP. The personal interview: Assessing the potential for personality similarity to bias the selection of orthopaedic residents. *Acad Med* 2009;84(10):1364–72. doi:10.1097/ACM.0b013e3181b6a9af.
36. Poole A, Catano VM, Cunningham DP. Predicting performance in Canadian dental schools: The new CDA structured interview, a new personality assessment, and the DAT. *J Dent Educ* 2007;71(5):664–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17493975>. Accessed March 23, 2018.
37. Shavelson RJ, Webb NM. Generalizability theory. 2005:599–612. Available at: https://web.stanford.edu/dept/SUSE/SEAL/Reports_Papers/methods_papers/GTheoryAERA.pdf. Accessed March 23, 2018.
38. Professional Testing Inc. Step 9. Conduct the item analysis. 2006. Available at: http://www.proftesting.com/test_topics/steps_9.php (Accessed January 1, 2019).
39. 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):350–9. doi:10.1111/j.1365-2923.2009.03292.x.
40. 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):767–75. doi:10.1111/j.1365-2923.2009.03407.x.
41. Eva KW, Reiter HI, Rosenfeld J, Trinh K, Wood TJ, Norman GR. Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *Jama* 2012;308(21):2233. doi:10.1001/jama.2012.36914.
42. 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):443–55. doi:10.3109/0142159X.2016.1158799.
43. Wilkinson D, Zhang J, Byrne G, et al. Medical school selection criteria and the prediction of academic performance. Evidence leading to change in policy and practice at the University of Queensland. *Med J Aust* 2008;188:349–54. doi:10.5694/j.1326-5377.2008.tb01653.x.
44. Ferguson E, James D, Madeley L. Factors associated with success in medical school: Systematic review of the literature. *Br Med J* 2002;324(7343):952–7. doi:10.1136/bmj.324.7343.952.
45. 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):1–10. doi:10.1186/1472-6920-14-54.
46. Corelli RL, Muchnik MA, Beechinor RJ, et al. Perceptions and cost-analysis of a multiple mini-interview in a pharmacy school admissions process. *Am J Pharm Educ* 2015;79(9):4–11.
47. Brownell K, Lockyer J, Collin T, Lemay JF. Introduction of the multiple mini interview into the admissions process at the University of Calgary: Acceptability and feasibility. *Med Teach* 2007;29(4):394–6. doi:10.1080/01421590701311713.
48. Finlayson HC, Townson AF. Resident selection for a physical medicine and rehabilitation program: Feasibility and reliability of the multiple mini-interview. *Am J Phys Med Rehabil* 2011;90(4):330–5. doi:10.1097/PHM.0b013e31820f9677.