



## Reliability and acceptability of the multiple mini-interview for selection of residents in cardiology

LUCRECIA M. BURGOS<sup>1\*</sup>, MD; ALBERTO ALVES DE LIMA<sup>1\*</sup>, PhD; JOSEFINA PARODI<sup>1</sup>, MD; JUAN PABLO COSTABEL<sup>1</sup>, MD; MARÍA NIEVES GANIELE<sup>2</sup>, MD; EDUARDO DURANTE<sup>2</sup>, MD; MARÍA DOLORES ARCEO<sup>2</sup>, MD; RICARDO GELPI<sup>3</sup>, PhD

*\*Both authors contributed equally*

<sup>1</sup>Instituto Cardiovascular de Buenos, Ciudad de Buenos Aires, Argentina; <sup>2</sup>Instituto Universitario del Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina; <sup>3</sup>Consejo Nacional de Investigaciones Científicas y Técnicas, Ciudad de Buenos Aires, Argentina

### Abstract

**Introduction:** The multiple mini-interview (MMI) model can be useful to evaluate non-cognitive domains and guide the selection process in medical residency programs. The aim of this study was to evaluate the reliability and acceptability of the MMI model for the selection of residents in a cardiology residency program.

**Methods:** We conducted an observational and prospective study. It was performed in a tertiary-care center specialized in cardiology and included candidates for the cardiology residency program in March 2018. Ten stations were developed to evaluate different non-cognitive domains. Reliability was evaluated by the generalizability G coefficient. Candidates and interviewers were surveyed to evaluate the acceptability of the MMI model.

**Results:** Nine faculty members were trained and 22 candidates were evaluated. The G study showed a relative G coefficient between 0.56 and 0.73, according to the design. 91% of the candidates stated that they preferred MMI over other types of interviews as a selection method for admission to the residency program, and all the interviewers considered they had enough time to evaluate the candidates and their strengths as future residents.

**Conclusion:** The MMI is a reliable model to evaluate candidates for a residency program in cardiology with high acceptability among residents and observers.

**Keywords:** Medical education; Cardiology; Internship and residency

*\*Corresponding author:*

Lucrecia M. Burgos, MD;  
Blanco encalada 1543,  
Zip code:1428,  
Ciudad de Buenos Aires,  
Argentina.

**Tel:** +54-011-4787-7500

**Email:** lburgos@icba.com.ar

*Please cite this paper as:*

Burgos LM, De Lima AA, Parodi J, Costabel JP, Ganiele MN, Durante E, Arceo MD, Gelpi R. Reliability and acceptability of the multiple mini-interview for selection of residents in cardiology. *J Adv Med Educ Prof.* 2020;8(1):25-31. DOI: 10.30476/jamp.2019.83903.1116.

**Received:** 18 October 2019

**Accepted:** 9 December 2019

### Introduction

The traditional interview with an interviewer or a panel interviewer in different formats is an instrument widely used in the application process of medical residency programs to evaluate non-cognitive competences, such as professionalism, teamwork and communication (1). However, variability of interviewer's skill, the questions that the interviewees must answer, interviewer bias, leniency or stringency, and context specificity limit the reliability of the interview so much that the process has been

described as an elaborate, labor-intensive lottery (2-4). Kreiter et al. published a review about the reliability of university admission interviews, concluding that there is not sufficient evidence to establish an appropriate reliability of the method. Inter-interviewer variability is a factor with considerable variation from 0.14 to 0.95 and this inconsistency is related to interview format. Thus, the reliability of the structured interviews is reported as higher than that of unstructured interviews (5). Reliability estimates may be artificially inflated by the interview

team having access to academic information on candidates and by non-verbal communication between members of the interviewing team (6). A candidate who is assigned to a permissive interviewer who influences the rest of the interview panel will score highly, whereas a candidate who is assigned to an exigent interviewer who influences the rest of the interview panel will score poorly (7). Content specificity is another important bias that further limits the reliability. Performance is more determined by the situation or context surrounding the skill in question than by the individual traits or characteristics. Turner et al. demonstrated that while inter-rater reliability was high in the oral certification examinations in Internal Medicine of the Royal College of Physicians and Surgeons of Canada, the generalizability across sessions was low (8). Thus, regardless of whether inter-rater correlation is high with structured questionnaires or well trained examiners, a single interview is not capable of determining an individual's performance in a given competence consistently. Thus, the need to generate an alternative approach emerges.

The multiple mini-interview (MMI) was developed by Eva et al. This model reduces the effect of the examiner and content specificity bias by increasing the number of interviews and using standardized questions (6). The MMI model is based on the objective structured clinical examinations (OSCEs) and is organized on several stations with an interviewer who rates the performance of the candidates in each station (9). This instrument has been validated and demonstrated generalizability and acceptability compared with traditional interview methods (2, 10-14). The MMI model has many advantages, including standardization and consistency, as all the candidates work through the same stations with the same scoring systems. The flexibility of the MMIs also allows residency programs to create stations that best reflect their specific expectations to select candidates with skills, competence and performance during the interview that better match the professional needs of the center (6). Another advantage is that the interviewers often feel more comfortable and unbiased about their ratings because they are not influenced by other interviewers. Ultimately, MMIs allow for a more comprehensive and authentic assessment of an applicant, preserving validity, acceptability, feasibility and reliability (6, 10, 12, 14, 15). This type of model has been applied in the selection process of many residency programs. Since then, the MMI model has been used as a selection tool in several medical schools and residency programs

in Canada, the United Kingdom, and Argentina (6, 13, 16-18). An experienced reporter in a congress abstract by the family medicine residency program of the *Hospital Italiano de Buenos Aires* showed that the relative G coefficient (the reliability of the ranking) was 0.72. The D study determined that a G coefficient of 0.80 could be reached with 9 stations in that population of residents (19). Six to 12-station MMI demonstrated that the tool is highly reliable (12, 20-22). Interviewers and institutions agreed that the MMI format was a reliable option to the conventional interview. MMI predicted performance outcomes during clerkship and on national licensing examination also demonstrated predictive validity for selecting medical trainees and in the context of US licensing examinations (10, 23).

The aim of this study was to evaluate the reliability and acceptability of a 10-station MMI model for the selection of residents in a cardiology residency program in Argentina.

## Methods

We conducted an observational and prospective study with psychometric analysis. The study was performed in a tertiary-care center specialized in cardiology and included candidates for the cardiology residency program in March 2018.

Firstly, the participants completed a selection test with multiple-choice questions designed to assess theoretical medical knowledge. Of the 160 applicants who took the test, the first 22 candidates ranked in descending order were selected. Candidates were invited by telephone and e-mail and were thoroughly informed about the evaluation method and agreed to participate voluntarily.

Nine interviewers of different medical specialties (cardiology, cardiology sub-specialties and clinicians) and healthcare professionals (nurses and psychologists) were trained to observe and score each candidate at each station.

Ten stations were developed with different scenarios. Station construction was guided by creation of an examination "blueprint" that fit program specifications based on relevant characteristics and qualities that were desirable in candidates for their respective residency training programs and aligned to the institutional vision (Table 1).

Different non-cognitive domains were evaluated in each station: professionalism, communication skills, clinical criterion, ethical behavior, tolerance for uncertainty, motivation towards the specialty, feedback and teamwork acceptance, and interpersonal skills.

A theoretical manual was prepared for the MMI interviewers with information about the basis

**Table 1:** MMI blueprint with the non-cognitive domains evaluated in each station and assigned score

| STATION NUMBER         | Motivation toward the specialty | Teamwork/ Interpersonal skills | Reasoning | Moral dilemmas | Feedback acceptance | Communication | Acceptance of professional limits | TOTAL SCORE FOR STATION |
|------------------------|---------------------------------|--------------------------------|-----------|----------------|---------------------|---------------|-----------------------------------|-------------------------|
| 1                      | 70                              |                                |           |                |                     | 30            |                                   | 100                     |
| 2                      |                                 |                                | 50        |                |                     | 50            |                                   | 100                     |
| 3                      |                                 | 30                             | 30        | 40             |                     |               |                                   | 100                     |
| 4                      |                                 | 40                             |           |                | 40                  | 20            |                                   | 100                     |
| 5                      |                                 |                                | 40        |                | 20                  | 40            |                                   | 100                     |
| 6                      |                                 | 50                             |           |                | 30                  | 20            |                                   | 100                     |
| 7                      |                                 | 50                             | 30        |                |                     | 20            |                                   | 100                     |
| 8                      |                                 |                                | 40        | 30             |                     |               | 30                                | 100                     |
| 9                      |                                 | 70                             |           |                |                     | 30            |                                   | 100                     |
| 10                     |                                 |                                |           | 50             |                     | 50            |                                   | 100                     |
| Total score for domain | 70                              | 240                            | 190       | 120            | 90                  | 260           | 30                                | 1000                    |

The Table shows different non-cognitive domains and the score assigned to each station and domain, with the sum of the total score for each station and domain

of the evaluation system, characteristics of the process, operational details and methodological description. Each evaluator previously read the assigned station ensuring confidentiality, so that the applicants would not know the station in advance. Then, the interviewers were trained by professionals with a master's degree in medical education from the Department of Family Medicine and General Medicine of the *Hospital Italiano de Buenos Aires*, with induction seminars and practice during simulated scenarios.

After the interview process, candidates and interviewers were asked to answer the questions of a brief self-administered and anonymous survey about how they perceived the MMI for the residents' selection and thus evaluate the acceptability of the method.

We used a generalizability theory study to analyze the reliability. The generalizability studies were fully crossed designs with the following facets: R/S, R:C/S and R/D (R: facet of differentiation residents; C: facet circuits; R:C: facet of differentiation of residents nested in each of both circuits; and D: facet of generalization domains).

Generalizability theory is an extension of classical reliability theory. It assumes that there is no arbitrary variation: a test is entirely determined by the condition of the 'true' construct being measured and the condition of the 'error' factors which influence the score (24). By analyzing the components of variance, it makes use of all the data to quantify all the sources of error without multiple experiments (25). This improves the statistical power and allows for a direct expression of the degree to which the results reflect all possible measurements of the same construct. By mathematical modeling, it also allows the

assessment strategies to be planned and overcome the main sources of error whilst keeping sampling to a minimum (26).

Generalizability theory can be applied in formative and summative examinations, and its use is recommended to investigate the sources of error and the number of observations required for a given level of reliability (27). Analysis of the sources of error in summative exams is useful as a quality control procedure to ensure reliable inferences from the results. G-theory assumes that the observed score of a person (the object of measurement) consists of a universal score (analogous to the true score in the classical test theory) and one or more sources of variation or facets (28). The strengths of G-theory lie in its ability to identify which facets of the MMI model (stations, domains, or residents) are the greatest sources of measurement error. It also allows the decision maker to determine the number of examination occasions, the test formats, and the examiners needed to obtain reliable scores (29). For all reliability analyses, a relative G coefficient equal to or greater than 0.70 was considered fair. The relative G coefficient indicates the reliability of the ranking generated from the candidates' global scores across all stations (R/S and R:C/S) or domains (R/D).

In addition, a decision study (D- study) was designed to estimate how many stations would be necessary to improve the reliability of the MMI. We used the EduG 6.1e software (Generalizability Study. Working Group - Edometrics - Quality of measurement in education of the Swiss Society for Research in Education).

An adequate dissociation of any filial information to the candidate that could allow the identification of any subject was made. We

randomly assigned numbers to the candidates, whose assignment was blind to the evaluators and data analysts

#### Ethical considerations

The confidentiality and anonymity of the persons participating in the study were guaranteed. The participants received an oral invitation to participate, mentioning the aims of the survey, the voluntary nature of their participation in the study and the confidentiality of the data. The study was conducted following the law N° 3301 of the city of Buenos Aires about protection of human subjects in health research and the recommendations of the Declaration of Helsinki, and approved by the Committee on Research and Ethics of our institution (Code number approval: 1162018).

#### Results

Nine faculty members were trained and 22 candidates were evaluated. We conducted the MMI in one day, in two consecutive circuits. Each one of them lasted 90 minutes. The candidates rotated through a circuit of 11 stations (10 station for candidate's evaluation and one rest station) in a sequence of 11 candidates per session. The same evaluators were assigned to the same stations in the two sessions.

Candidates were randomly assigned to the time and start of circuit station, and they were previously provided with a description of the process and basic instructions. They were assigned with labels with an identification number to guarantee a blinded evaluation.

A station identifier sign was placed on the door of each office and the vignette of each station describing the scene was hung. The candidate had 1 minute to read it and after a beep it was indicated to enter to the office. Each evaluation lasted 7 minutes; at the end of that time an audible signal

was made and the applicant left the office and rotated the station clockwise. Each interviewer completed an evaluation form with an analog visual scale to rate each non-cognitive attribute for each candidate. A free-text commentary field was available to explain the reasons, elements and characteristics observed in the candidate to rate the global performance (in order to provide a context for the score).

Interviewers were also allowed to note any 'red flag' issues observed during the interview, indicating lack of professionalism that could exclude the candidate during the rating process. Table 2 shows the different relative G coefficients and the variance associated with the facet of differentiation according to the different designs.

The determination of the optimal number of stations using the D-study are shown in Table 3.

The acceptability survey was completed by 100% of the participants. Eighty-two percent of the candidates stated that the time available to present their ideas and strengths in the stations was fair and that the MMI process was free of cultural or gender bias. 91% of the candidates mentioned that they preferred MMI over other types of interviews as a selection method for admission to the residency program, and 72% answered that the day of the interview was not excessively stressful (Figure 1).

Acceptability among the interviewers was also fair, considering that 100% of them had enough time to evaluate the candidates and their strengths as future residents; also, 75% of them stated that the number of interviewers was one of the strengths of this method, and 78% did not consider the activity as stressful, and that the score sheet allowed them to differentiate the candidates' competences (Figure 2).

#### Discussion

This study describes the feasibility of

**Table 2:** Relative G coefficient and the variance associated with the facet of differentiation according to different designs of generalizability. (R: residents; C: circuits; R:C; residents nested in each circuit; D: domains)

| Design | Relative G coefficient | Variance % associated with the facet of differentiation (facet=variance %) |
|--------|------------------------|--|
| R/S    | 0.56                   | R=8.7%   |
| R:C/S  | 0.57                   | R:C=7.9  |
| R/D    | 0.73                   | R=1.3  |

**Table 3:** Decision study: relative G coefficient for each design of generalization (R: residents; C: circuit; R:C; residents nested in each circuit)

| No. of stations | R/S  | R:C/S |
|-----------------|------|-------|
| 10              | 0.56 | 0.57  |
| 12              | 0.60 | 0.61  |
| 14              | 0.64 | 0.65  |
| 16              | 0.67 | 0.68  |

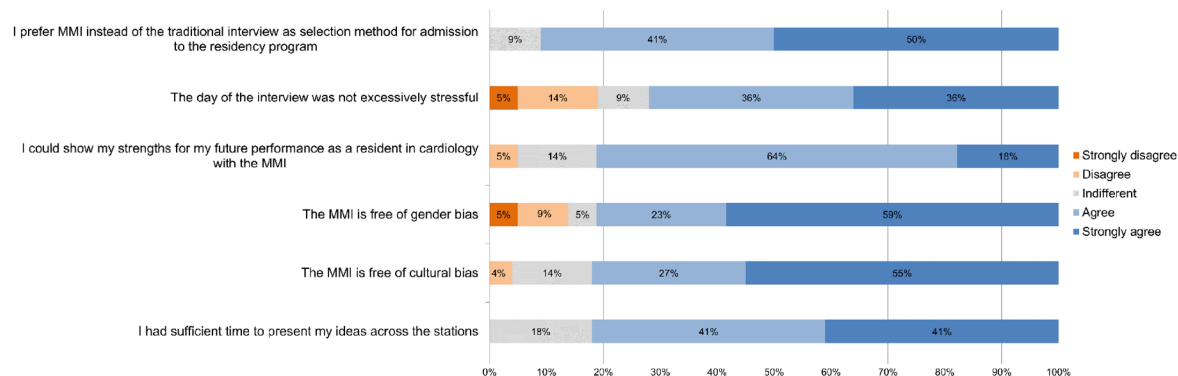


Figure 1: Acceptability of the MMI model among candidates

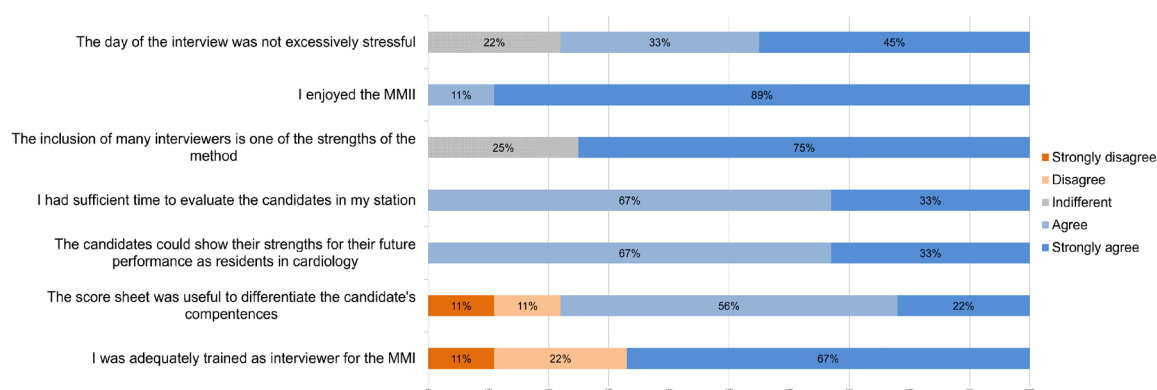


Figure 2: Acceptability of the MMI model among interviewers

implementing the MMI model for the selection of residents in a cardiology residency program. Reliability measured using stations as a facet of generalization was modest ( $<0.6$ ), but improved significantly (0.73) when it was generalized by domains. This could indicate that when residents are evaluated by domains across all the stations, the reliability is higher than when they are evaluated per station. This may be related to the fact that the same domain is evaluated by different observers in different contexts across the stations. Previous studies also demonstrated acceptable reliability. Eva et al. have found reliability coefficients of 0.73, 0.76, and 0.85 using 8, 9 and 12 stations, respectively (6, 20, 23). Roberts et al. described a reliability coefficient of 0.7 on an eight-station MMI study (12). Moreover, Hofmeister et al. reported a reliability value of 0.67 using 12 stations (13), while Fraga et al. found that the reliability of their process was  $>0.9$  with five stations (2).

The D-study demonstrates that a coefficient  $>0.7$  cannot be reached even with 16 stations. A previous investigation using the D-study revealed that the source of error was in line our finding. A number of researchers calculated into their data the hypothetical G coefficients for different numbers of stations and interviewers at each station, and

increasing the number of stations appears to have greater impact on reliability than increasing the number of interviewers at each station. Increasing the number of stations will hypothetically increase the reliability of the assessment, and the reliability of 10 is lower than 15 stations and 20 is the highest ( $G=0.76$ , 0.83 and 0.87, respectively) (30). Using generalisability analysis, Hecker and Violato (31) reported a G coefficient of 0.79 for seven stations with two assessors and a D-study indicated that G 0.81 could be achieved from ten stations with one assessor. Similarly, in Canada, the generalizability coefficient of a seven station MMI for selecting applicants into pediatrics, obstetrics and gynecology, and internal medicine ranged from 0.55 to 0.72, requiring 10 stations to increase reliability to 0.64-0.79 (18). Eva et al. (23) found a reliability of the total score across all nine stations of 0.76, and a D-study suggested that a 12-station MMI in this context would yield reliability of 0.80.

Also, the length of time at each station could all potentially impact the reliability. Researchers have examined the impact on the reliability of reducing time at each station, Cameron et al. (32) found that for five stations of eight minute reliability was 0.54; for five stations of six minute stations reliability was 0.66, which is similar to

the results found by Dodson (33).

Reliability also depends on the type of test and variance of the population on whom the test is applied and in the context (23). Probably, as these candidates were previously selected with an exigent test, they might not be very different from each other and thus this MMI is not capable of discriminating the little difference they present in the domains evaluated. It could be considered an MMI with more stations with the same duration and adding more evaluators, taking into account logistic aspects and resources used.

We also demonstrated that the MMI process was acceptable. We found great acceptability of the candidates and interviewers, as reported in studies evaluating undergraduate and postgraduate candidates (2, 6, 12-14, 18, 22, 34, 35). Interviewees did not find the process more stressful; they enjoyed the experience, felt that they had sufficient time, and the candidates could show their strengths. They also found the MMI model as a fair assessment and scoring sheet which allowed them to differentiate between the candidates.

The vast majority of candidates preferred this model to standard interviews; they did not find the interview stressful and found the model free of gender and cultural bias, as found in other MMI experiences (2, 14-26, 34-36). Dore et al. (18), in a group medical graduates to three residency programs in Canada, reported that 88% of candidates believed they could accurately portray themselves during the MMI, and 74% of the interviewers believed that the MMI outperformed the traditional interview.

Recently, a systematic review a systematic review reported the validity evidence of MMI in various educational settings (37). They found evidence to support its validity, and the findings revealed that MMI was flexible for assessing various important attributes of the candidates, such as professionalism, communication skills, ethics and morals, and critical thinking and problem solving, as the domains we evaluated in our MMI model. The MMI was generally acceptable to both candidates and interviewers across 11 countries and was consistently reliable and stable with acceptable Cronbach's alpha across educational settings; as the results of our study showed. And as we previously mention, MMI was reported as a bias-free admission tool for most factors such as culture and personal background. As the systematic review showed, MMI studies commonly included seven to twelve stations per circuit, with each station requiring seven to ten minutes.

These key findings provide evidence to

support the validity of MMI as an admission tool in a cardiology medical residency program.

The limitations of the study are the small sample size of the participants and, as it was conducted in a private center in the city of Buenos Aires, the acceptability of the MMIs may be different elsewhere. Another relevant aspect to mention includes the lack of multiple evaluators at each station which would increase reliability.

The present study analyzed the feasibility and acceptability of the MMI process for the selection of residents in non-English speaking countries. To the best of our knowledge, this report is the first implementation of the MMI for selection of residents into a cardiology residency program in Argentina.

## Conclusion

Our study provides evidence for the feasibility of implementing a 10-station MMI model in applicants to a cardiology residency program in Argentina. The model is well accepted by the candidates and interviewers, with an acceptable level of reliability and could be recommended as a method for residents' selection. Future research with a greater number of stations and adding multiple evaluators to the MMI model analyzing logistic and resource factors are necessary.

## Acknowledgments

We would like to acknowledge the faculty members and administrative staff who voluntarily participated in the MMI.

**Conflict of Interests:** None declared.

## References

1. Edwards JC, Johnson EK, Molidor JB. The interview in the admission process. *Acad Med.* 1990;65:167-77.
2. Fraga JD, Oluwasanjo A, Wasser T, Donato A, Alweis R. Reliability and acceptability of a five-station multiple mini-interview model for residency program recruitment. *J Community Hosp Intern Med Perspect.* 2013;3:3-4.
3. Norman G. The morality of medical school admissions. *Adv Health Sci Educ Theor Pract.* 2004;9:79-82.
4. Mann WC. Interviewer scoring differences in student selection interviews. *Am J of Occup Ther.* 1979;33:235-9.
5. Kreiter CD, Yin P, Solow C, Brennan R. Investigating the reliability of the medical school admissions interview. *Adv Health Sci Educ Theor Pract.* 2004;9:147-59.
6. Eva K, Rosenfeld J, Reiter H, Norma G. An admission OSCE: the multiple mini-interview. *Med Educ.* 2004;38:314-26.
7. Powis DA. How to do it – select medical students. *BMJ.* 1998;317:1149-50.
8. Turnbull J, Danoff D, Norman G. Content specificity

- and oral certification exams. *Med Educ.* 1996;30:56-9.
9. Kim KJ, Kwon BS. Does the sequence of rotations in Multiple Mini Interview stations influence the candidates' performance? *Med Educ Online.* 2018;23:1485433.
  10. Reiter HI, Eva K, Rosenfeld J, Norman G. Multiple mini interviews predict clerkship and licensing examination performance. *Med Educ.* 2007;41:378-84.
  11. Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Adv Health Sci Educ Theory Pract.* 2001;6:159-75.
  12. Roberts C, Walton M, Rothnie I, Crossley J, Lyon P, Kumar K, et al. Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Med Educ.* 2008;42:396-404.
  13. Hofmeister M, Lockyer J, Crutcher R. The acceptability of the multiple mini interview for resident selection. *Fam Med.* 2008;40:734-40.
  14. Lemay JF, Lockyer JM, Colin VT, Brownell AK. Assessment of non-cognitive traits through the admissions multiple mini-interview. *Med Educ.* 2007;41:573-9.
  15. Kumar K, Roberts C, Rothnie I, duFresne C, Walton M. Experiences of the multiple mini-interview: a qualitative analysis. *Med Educ.* 2009;43:360-7.
  16. 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;90:330-5.
  17. Humphrey S, Dowson S, Wall D, Diwakar V, Goodyear HM. Multiple mini interviews: opinions of candidates and interviewers. *Med Educ.* 2008;42:207-13.
  18. Dore KL, Kreuger S, Ladhani M, Rolfson D, Kurtz D, Kulasegaram K, et al. The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Acad Med.* 2010;85:60-3.
  19. Durante E, Arceo MD, Ganiele MN. Las múltiples minientrevistas en la selección de residentes de Medicina Familiar. *Inv Ed Med.* 2017;6:137.
  20. Eva KW, Reiter HI, Rosenfeld J, Norman GR. The relationship between interviewers' characteristics and ratings assigned during a multiple mini-interview. *Acad Med.* 2004;79:602-9.
  21. Rees E, Hawarden A, Dent G, Hays R, Bates J, Hassell A. Evidence regarding the utility of multiple mini-interview (MME) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 37. *Med Teach.* 2016;38:443-55.
  22. Razack S, Faremo S, Drolet F, Snell L, Wiseman J, Pickering J. Multiple mini-interviews versus traditional interviews: Stakeholder acceptability comparison. *Med Educ.* 2009;43:993-1000.
  23. 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;43:767-75.
  24. Crossley J, Russell J, Jolly B, Ricketts C, Roberts C, Schuwirth L. 'I'm pickin' up good regressions': the governance of generalisability analyses. *Med Educ.* 2007;41(10):926-34.
  25. Streiner D, Norman G. Generalizability. In: *Health Measurement Scales: a Practical Guide to their Development and Use*, 2nd edn. New York: Oxford University Press; 1995:128-43.
  26. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: a key to unlock professional assessment. *Med Educ.* 2002;36(10):972-8.
  27. Cronbach L, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas.* 2004;64:391-8.
  28. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach.* 2012;34:960-92.
  29. Shavelson J, Webb N. *MMSS generalizability theory. A primer.* Thousand Oaks, CA: Sage; 1991.
  30. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ.* 2011;45: 1181-9.
  31. Hecker K, Violato C. A generalizability analysis of a veterinary school multiple mini interview: Effect of number of interviewers, type of interviewers, and number of stations. *Teach Learn Med.* 2011;23:331-6.
  32. Cameron A, MacKeigan L. Development and pilot testing of a multiple mini interview for admission to a pharmacy degree program. *Am J Pharm Educ.* 2012;76:10.
  33. Dodson M, Crotty B, Prideaux D, Carne R, Ward A, de Leeuw E. The multiple mini-interview: how long is long enough? *Med Educ.* 2009;43:168-74.
  34. Brownell K, Lockyer J, Collin T, Lemay JF. Introduction of the multiple mini-interview into the admissions process at the University of Calgary: acceptability and feasibility. *Med Teach.* 2007;29:394-6.
  35. 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;79:40-2.
  36. Uijtdehaage S, Parker N. Enhancing the reliability of the multiple mini-interview for selecting prospective health care leaders. *Acad Med.* 2011;86:1032-9.
  37. Yusoff MSB. Multiple Mini Interview as an admission tool in higher education: Insights from a systematic review. *J Taibah Univ Med Sci.* 2019;14:203-40.