

Rater Perceptions of Bias Using the Multiple Mini-interview Format: a Qualitative Study

Richard L. Alweis^{1,2}, Caroline Fitzpatrick³, Anthony A. Donato^{1,2}

¹Reading Hospital and Medical Center, West Reading, Pennsylvania, USA

²Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania, USA

³Alvernia University, Reading, Pennsylvania, USA

Correspondence: Richard Alweis, MD, 6th Avenue and Spruce Street, Department of Internal Medicine, West Reading, Pennsylvania, USA

Received: May 13, 2015 Accepted: June 8, 2015 Online Published: June 26, 2015

doi:10.11114/jets.v3i5.818 URL: <http://dx.doi.org/10.11114/jets.v3i5.818>

Abstract

Introduction

The Multiple Mini-Interview (MMI) format appears to mitigate individual rater biases. However, the format itself may introduce structural systematic bias, favoring extroverted personality types. This study aimed to gain a better understanding of these biases from the perspective of the interviewer.

Methods

A sample of MMI interviewers participated in a series of primary and follow-up one-on-one semi-structured interviews. Interviews pursued subjects of perception of biases (including norming; applicant personality, appearance and behavior; and interviewer personality) associated with the MMI process. Emergent qualitative data analysis was performed using the constant-comparative method.

Results

A number of perceived biases were identified by subjects, sub-grouped into cultural factors, personality factors, perception of prior preparation, concerns with norming, and biases associated with specific applicant characteristics.

Discussion

While the MMI appears to help mitigate individual rater biases, our analysis suggests that raters perceive structural systematic biases may be introduced by the question type and format of the MMI itself. Whether rater awareness of these biases mitigates them, and whether these herald other unconscious biases is unknown.

Keywords: multiple mini-interview, rater perception, structural biases

1. Introduction

The preadmission interview is the primary source of non-cognitive information on applicants in almost all medical schools and residency programs (Axelson & Kreiter, 2009). However, the interview is beset by a number of biases, including content underrepresentation (CU), such as the context specificity of any single interview questions, and construct-irrelevant variance (CIV), including rater biases (leniency, stringency, similar backgrounds, interviewer expectations) that may account for as much as 56% of the biases in rating scores (Harasym, Woloschuk, Mandin, & Brundin-Mather, 1996). Applying the principles of the Objective Structured Clinical Exam, Eva *et al.* created the Multiple Mini-Interview (MMI) (Eva, Rosenfeld, Reiter, & Norman, 2004), a series of brief, semi-structured stations attended by trained raters. Adding more individual rater interviews using structured questions was intended to decrease the effects of context specificity while mitigating the effect of individual rater biases. The MMI was found to be reliable ($r=0.7-0.85$) using individual interviewers in 6-12 stations that ranged from 5.5-10 minutes each (Eva *et al.*, 2004; Harris & Owen, 2007; Hecker *et al.*, 2009; Hofmeister, Lockyer, & Crutcher, 2009; Roberts *et al.*, 2008). In addition, the MMI demonstrated predictive validity for future assessments of non-cognitive performance (Eva *et al.*, 2009; Lemay, Lockyer, Collin, & Brownell, 2007; Reiter, Eva, Rosenfeld, & Norman, 2007; Roberts *et al.*, 2008), was acceptable to applicants and interviewers (Diaz Fraga, Oluwasanjo, Wasser, Donato, & Alweis, 2013; Humphrey, Dowson, Wall,

Diwakar, & Goodyear, 2008; Kumar, Roberts, Rothnie, du Fresne, & Walton, 2009; Razack et al., 2009) and cost-effective (Rosenfeld, Reiter, Trinh, & Eva, 2008), leading to its widespread adoption as an interviewing format in Canada, the United States, and Australia (Eva et al., 2004; B. Griffin & Wilson, 2012; Kreiter, Yin, Solow, & Brennan, 2004).

Given that the format of the MMI was based on parallels with the OSCE, similar construct-irrelevant variance validity threats may also be introduced by the adoption of this method (Steven M Downing & Rachel Yudkowsky, 2009). Those CIV threats include structural systematic biases that may favor specific groups of students. Because the structure of the MMI requires multiple individual encounters, each encounter is typically shortened and structured in order to optimize the number of raters producing a judgment. This format, sometimes compared to 'speed dating' (Carpiniello, 2013), has the potential to favor those extroverted personalities who may perform better answering thought-provoking questions in a compressed time period. Indeed, many (B. N. Griffin & Wilson, 2010; B. Griffin & Wilson, 2012; Jerant et al., 2012; Oliver, Hecker, Hausdorf, & Conlon, 2014), but not all (Kulasegaram, Reiter, Wiesner, Hackett, & Norman, 2010) studies have indicated that extroversion and conscientiousness personality factors are positively associated with MMI scores (B. Griffin & Wilson, 2012; Lievens, Ones, & Dilchert, 2009). Qualitative analyses of interviewers have noted that their MMI scores may possibly be influenced by applicant personality (Humphrey et al., 2008) and applicant verbal fluency (Kumar et al., 2009). If the MMI format does favor extroverted personality types, it may be important to determine if the bias exists in the raters, in the process itself, or in both. Therefore, a greater understanding of potential biases in MMI interview formats at the locus of the interviewer is critical, as these biases may have an impact on selection and character of the next generation of physicians (Jerant et al., 2012).

Research Questions:

We addressed the following research questions in this paper:

What do MMI interviewers perceive as their biases when using the MMI format?

How do raters perceive the structure of the MMI itself influences their ratings of candidates?

2. Methods

2.1 Context

This was a qualitative study of MMI interviewers performed at one northeastern United States internal medical residency at an independent academic medical center. The residency program interviews approximately 240 candidates for 12 categorical (3-year internal medicine) and 6 transitional year (one year internal medicine, with transition to other non-internal medicine specialty) positions each year. In May, 2011, the program switched to the Multiple Mini-Interview (MMI) model of interviewing, with a modification to preserve a traditional program director interview. Our MMI structure consists of five, 8-minute MMI stations and a sixth 16-minute program director station (traditional interview) with two-minute breaks between stations, the details of which are reported elsewhere (Diaz Fraga et al., 2013; Humphrey et al., 2008; Kumar et al., 2009; Razack et al., 2009). Questions were developed internally by faculty after discussions regarding ideal candidate qualities, with the questions focused on each individual quality. The qualities identified were professionalism, team player, constructive response to stress, capacity for self-reflection, capacity for empathy, adaptability/tolerance of uncertainty, and the ability to incorporate feedback. Generalizability analysis indicated a high reliability of our MMI process (>0.9). Each station is staffed by one trained interviewer. Each MMI station involves the exploration of an applicant's reasoning on a clinical scenario with emphasis on non-cognitive elements. Scenarios are situational-type questions, probing applicants' responses to handling a challenging future hypothetical quandary in residency. The applicant's rationale for their answers are sought through using a set of scripted follow-up questions from the interviewer that use a behavioral format (i.e., asking how an applicant has handled a similar situation in the past). Raters evaluate applicants on two domains, interpersonal/communication skills and an overall score, both of which are graded on a behaviorally anchored, 7-point Likert scale. Interviewers meet at a debriefing session at the end of each recruitment day to discuss applicant scores and MMI process issues.

2.2 Sampling Strategy

After three recruitment seasons using MMI interviewing methods, all MMI interviewers still employed at the institution and not directly involved in the study were invited to participate in our study. Thirteen of the fourteen interviewers (93%) agreed to participate and enrolled in the study, all of whom completed it. Study subjects participated in a series of primary and follow-up one-on-one interviews with a trained qualitative researcher that was not involved in the development or application of the MMI to ensure anonymity, triangulation, and data saturation (O'Brien, Harris, Beckman, Reed, & Cook, 2014). To obtain the authentic responses of the participants as well as statements couched in thick description, data was collected using semi-structured, audio-taped interviews (Kvale & Brinkmann, 2008). Interviews pursued subjects of MMI process as well as perception of biases (including norming, applicant personality,

appearance and behavior, and interviewer personality)(Appendix A). They were asked about their best and worst interviews, and characteristics of scenarios they found most useful. Development of the interview tool was performed by the three study authors in conjunction with a literature review of the common biases associated with personal interviews (B. Griffin & Wilson, 2012; Jerant et al., 2012; Kumar et al., 2009). Each subject provided written consent. Research interviews were conducted over twelve months, and one reviewer (CF) de-identified all data prior to review by the two physician authors. This protocol was approved by the IRB of the Reading Health System (TRHMC Protocol Number: 017-013) and was funded by an unrestricted local grant from Reading Health System (\$4400).

2.3 Data Analysis

We used emergent qualitative data analysis (Miles & Huberman, 1994) to examine the interview data within the pre-determined research objectives. Rather than using a priori themes, the emergent analysis approach allowed the researchers to read and review the textual material multiple times to identify themes that “emerged” from the data. There were several stages involved in the analysis. Each of the three researchers independently read the entirety of the interview transcripts, and using the constant-comparative method (Glaser & Strauss, 1999), an initial inductive thematic analysis of the data was conducted to develop a detailed codebook used for identifying preliminary recurring themes. After a second and third review of the entirety of the interview transcripts by all three researchers independently, preliminary themes were refined to five recurrent themes.

3. Results

The thirteen MMI interviewers consisted of six physicians who were senior residents at the time that they served as interviewers, five core faculty of the residency program, one department chair, and one chief academic officer. These thirteen subjects represented 61.9% (13/21) of all trained interviewers; of the eight that were not interviewed, one was unavailable, five were no longer at the institution, and two were the physician authors on this study and thus were excluded. Seven of the thirteen were female (53.8%). The average age of the interviewers was 41.5 years (SD: 10.7). Seven participants described themselves as extroverts, four as introverts, and two as a mixture of both. Of the thirteen subjects, three (23.1%) were international medical graduates.

3.1 Key Themes

Our interviews revealed five major themes regarding interviewer perceptions of the MMI process, sub-grouped into specific applicant characteristics, personality factors, cultural factors, perception of prior preparation, and concerns with norming.

‘Specific applicant characteristics’ was the most commonly coded comment category, and were identified as creating both positive and negative impressions on reviewers. Characteristics identified in MMI sessions that were interpreted as creating a positive impression included enthusiasm, compassion, flexibility, demonstration of empathy, humility, and ability to make good eye contact. Characteristics that made an unfavorable impression included arrogance, inability to reason, inability to see other points of view, inability to articulate positions and maintain eye contact. Most of these were congruent with the characteristics the faculty laid out in creation of the MMI process. Several also noted issues with dress and appearance that made a negative impression, but did not always tie it directly to their ratings. Two illustrative comment regarding dress were

“I mean, some people dress sharper than others, but overall I try not to let that influence me.”- Subject 13

“I think I am a bit harder on women...Sometimes they will wear something to interviews, and I think, ‘I can’t believe you’re wearing that!’ I don’t go looking for these things...but I can’t say I am never affected.” –Subject 1

Subjects noted that they felt that applicants with introverted personalities may have fared less well in the MMI process, which they attributed to the rapid cycling through stations of the format. Many felt that the system itself could be biased toward extroverts who may be able to respond more quickly in situations that require spontaneous answers. Two illustrative quotes were:

“I think personality of both the interviewer and the interviewee can influence the process a lot. I’m an outgoing person. I love to socialize and talk, so I know that I tend to gravitate toward other people like me. I mean, I understand introverts, but it does take longer to get to know them, and in an interview, you don’t always have to time to peel away the shyness of a quiet candidate.” -Subject 6

“I think extroverts talk a little more. I think with more introverted people, <they> follow the script because the sessions are so brief and they don’t feel comfortable speaking off the cuff...they don’t divulge as much as quickly <as extroverts do>.”- Subject 10

Participants expressed concerns that the questions in the MMI appeared to require some knowledge and understanding of national norms concerning family structure, hospital hierarchy, and hospital systems, which we characterized as

“cultural factors”. Seven of the participants, including all three international medical graduates, specifically noted that applicants unfamiliar with national systems and cultural norms may have been at a disadvantage in the interview process. Two illustrative quotations were:

“There is a difference between the internationals who haven’t spent much time in the system compared to others. They are at a disadvantage. The cultural differences are more prominent in some of the scenarios.” - Subject 2

“For example, with medical error, there are specific things taught in <US> medical school, like that you should apologize, prevent further problems, and emphasize that the system will be improved. If you come from a medical school that hasn’t done that, you are at a disadvantage. If you come from a training environment that is hierarchical, there is more of a tendency to do what your senior says, as opposed to, ya know, what is your independent thought.” - Subject 3

When participants perceived that the applicant had been prepared in advance or was giving a rehearsed answer, they felt negatively toward the candidate; they were mixed on the perceived effects on their ratings of that candidate. Two illustrative quotes were:

“I mean, I found that I wasn’t as annoyed by the over-preparation as I was annoyed by their inability to mask it. I tend to give more average scores to those people because I cannot tell whether they are real or not.” - Subject 8

“There were a few people that seemed rehearsed in their answers. It was almost a little monologue where they would hit all of the relevant points...almost like they were reading out of a book. I tried to ask them more follow-up questions than I normally would, but I don’t think it affected my scoring.” - Subject 9

Five subjects voiced concern that scores might be biased by norm-referencing, or comparing the applicants to those coming before or after them on a specific interview day, as evidenced by the following two quotes:

“I think it depends on where they fall in the interview day...if I am interviewing someone really strong first, then I have a higher standard for everyone afterwards, but I might not realize that until the third person or so.” - Subject 1

“The first few people you interview in a day can help set the bar low or high for the rest of the candidates that day. If I interview a strong candidate first, then I have a higher standard for everyone else and vice-versa.” - Subject 4

4. Discussion

Construct-Irrelevant Variance threats to interpretation of scores can come from any structural systematic biases introduced by individual raters, the process, or a mix of rater biases brought on by the process itself. In this study, we attempted to better understand the question of conscious biases of raters in an MMI process. Our raters could consciously identify that their ratings were affected by physical appearance, and evidence of prior preparation, and norm-referencing. They were able to point out that the system itself seemed to be biased against those those who were not aware of cultural norms and systems issues and those with introverted personality types.

Most of the applicant characteristics cited by the subjects were directly congruent with the goal values laid out by the faculty, and are unlikely to constitute biases. However they were also able to identify physical appearance attributes that may have influenced their scoring. Other qualitative reviews of MMI interviewers have noted that subjects felt that the criterion-based MMI format actually decreased the weight of ‘grooming’ and other subjective physical biases (Kumar et al., 2009). While we could not ascertain whether these biases are reduced by the MMI process relative to standard interviews, it might be expected that any individual rater bias would be mitigated by a process that increased the number of interviews as the MMI does.

Our participants noted a bias against those with what appeared to be pre-prepared answers. This is in contrast to work by others (Reiter, Salvatori, Rosenfeld, Trinh, & Eva, 2006), who found that having the MMI questions available did not positively affect the scores of participants as compared to those without access to the questions. In light of our work, it is possible that the benefit of advanced preparation may have been offset by the interviewers’ negative perception of “canned answers” or the lack of authenticity that this perception portrayed to raters.

Norm-referencing is a threat to the validity of any test, and one that OSCE administrators have addressed by adding strict criterion –referencing (using checklists, determining passing cutpoints) to overcome (Wass, Van der Vleuten, Shatzer, & Jones, 2001). Despite our structured rating form, participants felt that performance of the initial applicants of the day affected scores of those interviewed later in the day. A prior study noted a significant difference in morning as compared to afternoon interview ratings, but this finding disappeared after adding random allocation to interview times (Humphrey et al., 2008). Norm-referencing is common in personal interviews of all formats (Edwards, Johnson, & Molitor, 1990) and in other progress tests in medical education (McHarg et al., 2005; McManus, Mollon, Duke, & Vale, 2005). It is not known whether our addition of behavioral anchors in rating forms mitigated this bias, but the overall beneficial effect on changes in rating forms would be expected to be small (Landy & Farr, 1980). The MMI process is

intended to require no prior knowledge to complete, with the intent of assessing mainly non- medical knowledge attributes. Similar to most MMI scenarios, our ‘situational’ questions (how an applicant might handle a future hypothetical situation) may require knowledge of local rules, norms and culture. Our subjects pointed out that the MMI process may bias our program toward those with previous United States clinical experience. Prior studies have identified similar concerns regarding fairness to applicants with cultural differences (Razack et al., 2009). Our structured follow-up questions asked behavioral-type questions (reflection upon past actions). The combination of both types of questions appears to show better prediction of future job performance in studies (Campion, Campion, & Hudson, 1994; Conway & Peneno, 1999). Behavioral aspects of questions may be better at discriminating between stronger and weaker candidates (Eva & Macala, 2014). However, whether MMI questions without situational components would discriminate as effectively while mitigating potential systematic bias is a matter for future study. If structural systematic bias is introduced by the MMI process, it may be a bias that interviewers are aware of, or it may be an unrecognized bias. We found that both our self-described extrovert and introvert interviewers felt that the MMI process itself (both the question types and format) may have biased their ratings in favor of the quick-responding and more socially-inclined extroverts. Other studies examining structural systematic bias have reached conflicting results. Our findings are similar to others who noted that conscientious and extroverted personality traits were positively associated with MMI scores in medicine (B. N. Griffin & Wilson, 2010; Humphrey et al., 2008; Jerant et al., 2012; Oliver et al., 2014) and in non-medical interviews (Chen, Huang, Huang, & Liu, 2011). However, our results differ from the findings of Kulasegaram *et al* (Kulasegaram et al., 2010), who noted no difference in MMI scores among differing personality traits using the Neo-5 personality score. Whether the identified biases cited by our subjects are attenuated by their awareness or are accompanied by a hidden component that affects the heterogeneity of our applicant pool is unknown. As part of our MMI process, we choose to debrief with all interviewers regarding the applicants following each interview session. Whether our debriefing session may benefit in both exposing previously hidden biases as well as norming our rater outliers, or whether these sessions adversely influence important observations as younger members are influenced by older peers, is a matter for further study.

5. Limitations

This work was intended to be an exploration of bias in the MMI process and raters. An important limitation of any study on bias is that raters may not be conscious of important hidden biases, and they may not have been detected by our study method. Furthermore, the locus of bias (rater or process) is not possible to disentangle, as our raters are inextricably part of the process. More importantly, the relevance and the effect of extroverted personality types on a physician’s performance in practice is unknown, and is a critical question regarding the value of this line of research. In addition, we sought to identify systematic bias in a system at an individual rater level. Whether these individual biases have an impact on the overall system, and whether this is generalizable to other MMI systems, is unknown. Similar to most of the MMI research, the findings of this small study may not be generalizable to other centers with different training, structure, faculty, and question formats, and process (including interview time, and debriefing sessions) (Eva et al., 2004). Given that our MMI interviewers are all volunteers, as was the subset that were participants in the study, their impressions may be skewed in a way that does not reflect those of the interviewer population as a whole.

6. Conclusions

While the MMI appears to help mitigate individual rater biases, our qualitative analysis of raters suggests that systematic biases may be introduced by the question type and format of the MMI itself. The impact of these biases on physician selection, and the downstream impact on patient care are matters for future study.

References

- Axelson, R. D., & Kreiter, C. D. (2009). Rater and occasion impacts on the reliability of pre-admission assessments. *Medical Education*, 43(12), 1198–1202. <http://dx.doi.org/10.1111/j.1365-2923.2009.03537.x>
- Campion, M. A., Campion, J. E., & Hudson, J. P. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, 79(6), 998–1002. <http://dx.doi.org/10.1037/0021-9010.79.6.998>
- Carpiniello, M. (2013). The Multiple Mini-Interview (MMI): Interviewing the Speed-Dating Way Penn & Beyond. *University Life*. <https://ulife.vpupenn.edu/careerservices/blog/2013/11/20/the-multiple-mini-interview-mmi-interviewing-the-speed-dating-way/>
- Chen, C., Huang, Y., Huang, T., & Liu, I. (2011). The mediating processes linking applicant personality traits and interviewer evaluation. *International Journal of Selection and Assessment*, 19(3), 287–300. <http://dx.doi.org/10.1111/j.1468-2389.2011.00558.x>

- Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology*, 13(4), 485–506. <http://dx.doi.org/10.1023/A:1022914803347>
- Diaz Fraga, J., Oluwasanjo, A., Wasser, T., Donato, A., & Alweis, R. (2013). Reliability and acceptability of a five-station multiple mini-interview model for residency program recruitment. *Journal of Community Hospital Internal Medicine Perspectives*, 3(3-4). <http://dx.doi.org/10.3402/jchimp.v3i3-4.21362>
- Edwards, J. C., Johnson, E. K., & Molidor, J. B. (1990). The interview in the admission process. *Academic Medicine: Journal of the Association of American Medical Colleges*, 65(3), 167–177. <http://dx.doi.org/10.1097/00001888-199003000-00008>
- Eva, K. W., & Macala, C. (2014). Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Medical Education*, 48(6), 604–613. <http://dx.doi.org/10.1111/medu.12402>
- 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(8), 767–775. <http://dx.doi.org/10.1111/j.1365-2923.2009.03407.x>
- Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education*, 38(3), 314–326. <http://dx.doi.org/10.1046/j.1365-2923.2004.01776.x>
- Glaser, B., & Strauss, A. (1999). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction.
- Griffin, B. N., & Wilson, I. G. (2010). Interviewer bias in medical student selection. *The Medical Journal of Australia*, 193(6), 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(3), 377–388. <http://dx.doi.org/10.1007/s10459-011-9316-1>
- Harasym, P. H., Woloschuk, W., Mandin, H., & Brundin-Mather, R. (1996). Reliability and validity of interviewers' judgments of medical school candidates. *Academic Medicine: Journal of the Association of American Medical Colleges*, 71(1 Suppl), S40–42. <http://dx.doi.org/10.1097/00001888-199601000-00038>
- 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(3), 234–241. <http://dx.doi.org/10.1111/j.1365-2929.2007.02682.x>
- 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(2), 166–173. <http://dx.doi.org/10.3138/jvme.36.2.166>
- Hofmeister, M., Lockyer, J., & Crutcher, R. (2009). The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Medical Education*, 43(6), 573–579. <http://dx.doi.org/10.1111/j.1365-2923.2009.03380.x>
- Humphrey, S., Dowson, S., Wall, D., Diwakar, V., & Goodyear, H. M. (2008). Multiple mini-interviews: opinions of candidates and interviewers. *Medical Education*, 42(2), 207–213. <http://dx.doi.org/10.1111/j.1365-2923.2007.02972.x>
- Jerant, A., Griffin, E., Rainwater, J., Henderson, M., Sousa, F., Bertakis, K. D., ... Franks, P. (2012). Does applicant personality influence multiple mini-interview performance and medical school acceptance offers? *Academic Medicine: Journal of the Association of American Medical Colleges*, 87(9), 1250–1259. <http://dx.doi.org/10.1097/ACM.0b013e31826102ad>
- 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: Theory and Practice*, 9(2), 147–159. <http://dx.doi.org/10.1023/B:AHSE.0000027464.22411.0f>
- Kulasegaram, 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: Theory and Practice*, 15(3), 415–423. <http://dx.doi.org/10.1007/s10459-009-9209-8>
- Kumar, K., Roberts, C., Rothnie, I., du Fresne, C., & Walton, M. (2009). Experiences of the multiple mini-interview: a qualitative analysis. *Medical Education*, 43(4), 360–367. <http://dx.doi.org/10.1111/j.1365-2923.2009.03291.x>

- Kvale, S., & Brinkmann, S. (2008). *InterViews: Learning the Craft of Qualitative Research Interviewing* (2nd edition.). Los Angeles: SAGE Publications, Inc.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87(1), 72–107. <http://dx.doi.org/10.1037/0033-2909.87.1.72>
- 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(6), 573–579. <http://dx.doi.org/10.1111/j.1365-2923.2007.02767.x>
- Lievens, F., Ones, D. S., & Dilchert, S. (2009). Personality scale validities increase throughout medical school. *The Journal of Applied Psychology*, 94(6), 1514–1535. <http://dx.doi.org/10.1037/a0016137>
- McHarg, J., Bradley, P., Chamberlain, S., Ricketts, C., Searle, J., & McLachlan, J. C. (2005). Assessment of progress tests. *Medical Education*, 39(2), 221–227. <http://dx.doi.org/10.1111/j.1365-2929.2004.02060.x>
- McManus, I. C., Mollon, J., Duke, O. L., & Vale, J. A. (2005). Changes in standard of candidates taking the MRCP(UK) Part 1 examination, 1985 to 2002: analysis of marker questions. *BMC Medicine*, 3, 13. <http://dx.doi.org/10.1186/1741-7015-3-13>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE.
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine: Journal of the Association of American Medical Colleges*, 89(9), 1245–1251. <http://dx.doi.org/10.1097/ACM.0000000000000388>
- Oliver, T., Hecker, K., Hausdorf, P. A., & Conlon, P. (2014). Validating MMI scores: are we measuring multiple attributes? *Advances in Health Sciences Education: Theory and Practice*, 19(3), 379–392. <http://dx.doi.org/10.1007/s10459-013-9480-6>
- 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(10), 993–1000. <http://dx.doi.org/10.1111/j.1365-2923.2009.03447.x>
- Reiter, H. I., Eva, K. W., Rosenfeld, J., & Norman, G. R. (2007). Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education*, 41(4), 378–384. <http://dx.doi.org/10.1111/j.1365-2929.2007.02709.x>
- 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(1), 36–42. <http://dx.doi.org/10.1111/j.1365-2929.2005.02348.x>
- Roberts, C., Walton, M., Rothnie, I., Crossley, J., Lyon, P., Kumar, K., & Tiller, D. (2008). Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Medical Education*, 42(4), 396–404. <http://dx.doi.org/10.1111/j.1365-2923.2008.03018.x>
- Rosenfeld, J. M., Reiter, H. I., Trinh, K., & Eva, K. W. (2008). A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Advances in Health Sciences Education: Theory and Practice*, 13(1), 43–58. <http://dx.doi.org/10.1007/s10459-006-9029-z>
- Steven, M. D., & Rachel, Y. (2009). *Assessment in Health Professions Education* (first.). Taylor and Francis.
- Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *Lancet*, 357(9260), 945–949. [http://dx.doi.org/10.1016/S0140-6736\(00\)04221-5](http://dx.doi.org/10.1016/S0140-6736(00)04221-5)

