

主論文の要旨

**Reliability and acceptability of six station multiple
mini-interviews: past-behavioural versus situational
questions in postgraduate medical admission**

〔 6ステーションのマルチプルミニインタビューの信頼性と受容性：
卒後医師採用試験における、過去の行動質問対状況質問 〕

名古屋大学大学院医学系研究科 総合医学専攻
発育・加齢医学講座 総合診療医学分野

(指導：葛谷 雅文 教授)

山田 徹

【緒言】

採用面接の代表的な手法には **Single station personal interview(SSPI)**と **Multiple mini-interview(MMI)**がある。**SSPI**は世間でもっとも多く行われている面接方法であり、1つの面接ステーションで複数の面接官が応募者に対して志望動機、自身の長所など様々な複数の質問をする形式である。しかし **SSPI**は質問内容・質問の数・評価基準の構造化を問わないため、インタビューバイアスや文脈特異性の問題などが指摘されていた。そのため面接官がよほど入念なトレーニングをされていない限り、**SSPI**の信頼性や妥当性を担保するのは困難とされている。一方 **MMI**では1つの面接ステーションで1人の面接官が1項目のみ質問を行い、質問終了後に応募者は別ステーションに移動し、別の評価項目に対する質問を受けるという形式である。面接官交代によるインタビューバイアス対策、評価項目の明示による文脈特異性の問題に対する対策、ルーブリックによる採点基準の明確化などがなされており、**SSPI**より高い信頼性が示されている。その反面、信頼性を保つためには一般的には10ステーション以上を準備する必要があるなど高コストであり、実行可能性面の問題が指摘されている。

欧米では **MMI**は卒前、卒後の医師採用面接に広く使用されている。**MMI**の質問形式としては、過去の実際の行動を問う **Past-Behavioural Question (PBQ)**と、仮定の状況での行動を問う **Situational Question (SQ)**がある。医学部入試や初期研修入職時の **MMI**では、応募者に医師として働いた経験がなく **PBQ**が使いにくいこともあり、**SQ**で行うのが一般的であった。そのため **PBQ**の有用性は不明確である。後期研修入職時の **MMI**における **PBQ**と **SQ**の信頼性を比較した研究では、**PBQ**と **SQ**の信頼性に有意差はなく、受容性は **SQ**の方が高いことが示されたが、質問順序の固定(1問目：**PBQ**、2問目：**SQ**)によるバイアスがあった。

そのため本研究では卒後医師採用試験における **MMI**で、質問順序のバイアスがわからないよう企画し、**PBQ**と **SQ**の信頼性と受容性の検討をおこなった。また構造化の強化により、より少ない面接官の数で信頼性を確保できるかを検討した。

【方法】

東京ベイ・浦安市川医療センターの2016年度と2017年度の後期研修応募者(初期研修医)40名と、面接官の指導医24名を対象とした。**MMI**は6つの面接ステーションで行った。1ステーションに1名ずつ面接官を配置し、各ステーションで **PBQ**と **SQ**を1問ずつ質問した。質問項目は **ACGME (Accreditation Council for Graduate Medical Education)**の定めた、医師に必要な6つのコンピテンシーから作成した。6つのコンピテンシーとは①患者のケア、②診療を基にした学習と改善能力、③コミュニケーション能力、④プロフェッショナルリズム、⑤システムに基づいた診療、⑥医学知識である。医学知識は面接での評価は不適切なため除外し、研修委員会にてもっとも重要だとされたプロフェッショナルリズムについては2ステーション設定し、合計で6ステーションとした。質問順序は順序によるバイアスがわからないよう、ステーション毎に **PBQ**と **SQ**を交互に入れ替えた。また構造化の強化のため、事前講義や面

接シミュレーションによる面接官のトレーニング、STAR (Situation-Task-Action-Result)アプローチによる質問の構造化、質問手順の明文化と遵守、Rating rubric による評価方法の構造化を行った。受容性は表面的妥当性、参加者の反応、公平性、実現可能性の複合で評価されるため、それぞれをポスト MMI サーベイを用いて検討した(table 1)。

データ解析は、信頼性の検討には一般化可能性理論に基づき、参加者の能力・rating rubric・ステーションを因子とし、一般化可能性研究と決定研究を行った。受容性の検討には対応のある t 検定、一標本 t 検定と二項検定を用いた。統計ソフトは信頼性の検討には Mplus ver.5.21 を、受容性の検討には R ver.3.1.3 を用いた。

【結果】

6 ステーションでの PBQ、SQ、PBQ+SQ の generalizability-coefficient はそれぞれ 0.87、0.96、0.80 であった。決定研究では PBQ のみ、または SQ のみでの 4 ステーションの MMI でも十分な信頼性が得られることが示された(table 2)。ポストアンケートでは参加者の 83%が MMI による面接方法に満足し、公正さ・負担感も 9 割以上の参加者が SSPI より優れると回答した。表面的妥当性は応募者では SQ より PBQ が高かったが($p=0.010$)、面接官では有意差は認めなかった($p=0.377$)。応募者は SQ より PBQ を好む傾向が見られたが統計学的な有意差は認めなかった($p=0.081$)。面接官は明らかに PBQ を好んだ($p=0.007$)。参加者の 8 割以上が PBQ と SQ の両方を用いるべきだと回答した(table 3)。

【考察】

本研究では PBQ、SQ とも後期研修希望者を対象とした MMI で十分な信頼性があることが示された。また過去の研究と比較してより高い信頼性が示されたが、その理由として前述の構造化の工夫に加え、約 70%の面接官が複数回の面接官を経験しており、面接スキルが習熟されていた可能性がある。またどちらか片方の質問のみを用いた 4 ステーションの MMI でも十分な信頼性が示された。実行可能性の点ではこの方法は有用だが、予測妥当性は低下する可能性がある。また前述のとおり 8 割以上の参加者が両方の質問を用いるべきと回答している点からも、片方のみの質問を用いた MMI では受容性の低下も懸念される。

MMI の受容性については当初は SSPI より手間が増えるため受け入れられ難いことと懸念したが、全体満足度は 8 割以上であり、公正さ・負担感も 9 割以上が SSPI より優れるという結果であった。これは MMI が実行可能性の点でも SSPI に劣っていないことを示している。それぞれの質問形式に対する受容性では面接官は明らかに PBQ を好み、応募者も PBQ を好む傾向がみられた。これは質問方式を改善した結果、より正確な受容性が評価できたと考えられる。

本研究の limitation は、単施設での研究のため今後多施設での評価が必要であることである。また本来 MMI では一つのステーションで 1 つの質問にすべきところを、

本研究では2つの質問をした。そのため1問目の結果が2問目に影響を与えた可能性は否定できない。しかしステーション毎に質問順序を交互にしてPBQとSQの条件を同等にしたため、両者の比較という点では大きな問題はないと考えられる。

【結語】

6ステーションのMMIではPBQ、SQとも十分な信頼性があり、SQはより高い信頼性を示した。面接官1名で4ステーションでも十分な信頼性が担保されることが示唆された。面接官は明らかにPBQを好み、応募者もPBQを好む傾向がみられた。受容性の面からはPBQ、SQ両方を用いるべきであることが示された。