

医学生の実能・態度の評価に関する研究

～日本の卒前医学教育における WBA(Workplace-Based Assessment)のあり方について～

研究分担者 錦織 宏 (名古屋大学医学系研究科 教授)
小西 靖彦 (順天堂大学医学教育研究室 特任教授)
伴 信太郎 (愛知医科大学医学教育センター 特命教育教授)
松山 泰 (自治医科大学医学教育センター 教授)
高村 昭輝 (富山大学学術研究部医学系 医学教育学 教授)

研究要旨

我が国の卒前医学教育における臨床実習の質保証と、医師国家試験における技能・態度の評価は、旧来より医学教育の重要なテーマである。本研究は、診療現場における観察評価である Workplace-Based Assessment(WBA)を、我が国の医学教育においてどのように実装すればよいかを検討することを目的とした。

本研究では、国内 82 の医学部の JACME 自己点検評価報告書を精査し、WBA の導入状況を調査した。さらに、WBA を積極的に導入している 5 大学を対象にインタビューを行い、運用上の課題や展望を抽出した。また、米国の USMLE における Step 2 CS の OSCE 廃止と WBA 導入の背景を踏まえ、ECFMG および NBME の関係者に対するインタビューを実施し、海外の制度の動向も分析した。

その結果、国内の 65 大学で何らかの WBA が導入されていたが、その多くは診療科ごとの裁量に委ねられた形成的評価にとどまっていた。一方、米国では Mini-CEX や Global Practice Assessment を通じた診療科単位での総括的評価が進められており、Programmatic Assessment の一部導入も確認された。

以上の知見を踏まえ、本研究は、我が国における WBA の実装に向けたブループリントを提示した。このモデルは、コミュニケーション能力およびプロフェッショナリズムを中心に、全診療科で複数回の WBA を実施し、形成的評価と総括的評価を統合する枠組みで構成されている。本研究は、我が国における臨床実習の評価と医師国家試験の改善に向けた実証的知見と実装モデルを提示するものであり、今後の教育政策に資すると考える。

A. 研究目的・背景

医学教育、特に臨床教育における評価を考える際、「知っていることはできることを保証しない」という Miller のピラミッドに基づいた考え方は広く浸透している。400 問の多選択肢式問題から構成される我が国の医師国家試験が時に批判を受けるのも、主にこの観点からである。「評価法が学習の仕方に大きな影響を与える」という教育学の理論を持ち出すまでもなく、臨床教育における評価は、診療参加型臨床実習の推進を図る上でも非常に

重要である。

では実際、現場の臨床医・指導医はどのようにして、医学生や研修医、若手医師の評価をしているのだろうか？「一緒に働けばわかる」という言葉が現場からはしばしば聞かれるが、その言葉の意味するところは、臨床現場における観察評価の重要性である。そして、臨床現場では能力を細かい要素に分けるのではなく、概略を評価していると言える。

この「臨床現場における観察評価」を、たまたま

やる気のある指導医がゆとりのある時に行うのではなく、系統的に全ての学習者が評価を受けられるようにできないかと考えたのが、本研究の出発点である。忙しい臨床医・指導医がどのようにしたらフィードバックに時間を割けるかという実現可能性の視点も含めて、調査・研究・検討を行った。

特に、米国が近年、国家試験に該当する試験 USMLE(United States Medical Licensing Examination)において、Step 2 CS の OSCE (Objective Structured Clinical Examination:客観的臨床能力試験)を廃止して、臨床現場における観察評価にシフトしてきている。

さらに、外国の医学部卒業生教育委員会(以下、ECFMG)は米国外医学部卒業者に対し、6 症例の Mini-CEX 合格を代替条件の一つとして定めた。

臨床現場における観察評価は、英語では WBA(Workplace-Based Assessment)と呼ばれる。本邦において WBA を実質化することで、臨床実習がより充実したものになることが期待さ

れる。なお、本報告書の最後のパートでは、我が国における WBA のモデルをブループリントの形で提示しているが、今後、パイロット実施も含めた形で実装する中でより良い形を作っていくためのたたき台の案であることを申し添える。

1. 国内における WBA の現状について

我が国の医学部における臨床実習中の WBA 実施の現状を踏まえ、我が国において医師免許を与えるために必要とされる技能(態度領域を含む臨床能力)の総括的評価について、WBA による評価が可能であるのかを検証することを目的とする。

2. 海外(主に米国)における WBA の現状について

米国の USMLE では Step 2 CS を廃止して、試験制度の中で WBA を導入されたことから、米国の試験制度に関する論文調査およびインタビュー調査を行い、我が国の医師国家試験における WBA の活用方法を探索することを目的とする。

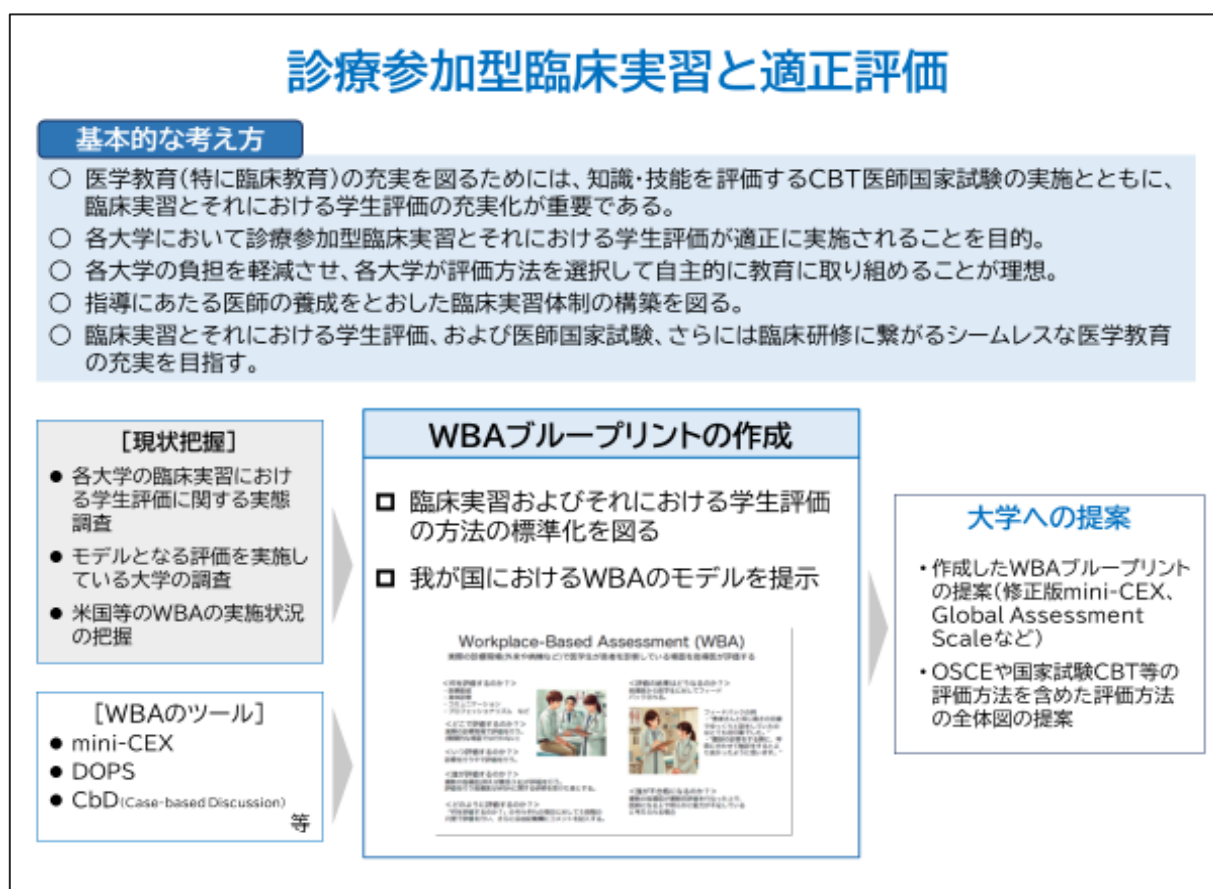


図1. 診療参加型臨床実習と適正評価に関する本研究の全体像

3. 今後の WBA のあり方とブループリントの開発について

臨床実習における医学生の評価の実施状況および海外の WBA の実施状況を参考に、我が国における WBA のブループリントを作成し、臨床実習における医学生の評価の標準化することを目的とする。

B. 研究方法

1. 国内における WBA の現状について

Web 上に公開されている日本医学教育評価機構(以下、JACME)による医学教育分野別評価の全国 82 医学部の自己点検評価報告書を詳細に調査した。その中から臨床実習の評価に特に WBA を用いていると思われる大学医学部(国公立大学 3 校、私立大学 2 校)の評価担当者からインタビュー調査を実施した。

2. 海外(主に米国)における WBA の現状について

ECFMG の運営団体 Intealth の CEO である Eric Holmboe 氏が推奨した 15 本の論文を精読し、WBA に関する情報を抽出して、これをもとにインタビューフォームを作成した。同フォームを用い、2024 年 7、8 月に、Holmboe 氏、USMLE 管理団体 NBME の CEO である Peter Katsufakis 氏、および評価副責任者 Christopher Feddock 氏にオンライン・インタビューを実施した。

3. 今後の WBA のあり方とブループリントの開発について

日本における今後の WBA のあり方を検討するにあたり、以下の8つの問いをもとに本研究で討議を行い、我が国の医学教育で活用可能な WBA のブループリントの開発を行った。

< 検討テーマ >

1. WBA をどのように総括的評価に用いるのか？(用いないのか？)
2. 学生へのフィードバックの動線をどのように設計するか？
3. どのような学生が不合格となるのか？
4. EPA(Entrustable Professional Activities)をどのように用いるのか？(用いないのか？)
5. 評価によって、どのように診療参加を促進させるのか？
6. どの診療科で National WBA(仮称)を実施するのか？
7. アンプロフェッショナルな行動・態度の評価をどのように組み込むのか？(組み込まないのか？)
8. OSCE との役割分担をどのように設計するのか？

C. 研究結果

1. 国内における WBA の現状について

公表されている JACME のデータ(2024 年 6 月時点)からは全国 82 の医学部のうち、65 の医学部で何らかの WBA による評価が実施されていた。内訳としては Mini-CEX が最も多く、続いて MSF (Multi-Source Feedback)、360 度評価、観察記録等が多かった。総括的評価として WBA を活用している大学は現時点ではなく、全ての大学で形成的評価として用いられていた。また、臨床実習でローテーションする診療科全てに WBA を課している大学は存在せず、明確に決まった診療科で実施されている医学部が 2 か所、臨床実習後半や地域医療実習で義務として行われている医学部が 3 か所であり、残りの 60 の医学部では各診療科の裁量で実施されていた。インタビュー調査では、WBA の運用に関する現状、問題点、総括的評価として用いることの可能性に焦点を絞って実施した。新型コロナウイルスのパンデミックにより、臨床実習が中断したことで以前に比較して WBA の頻度や回数、質などが下がってしまったと解答している大学が多く、今後、再び元に戻すように努力する、もしくはもっと積極的に WBA を

導入していこうとしている姿勢が見られた。問題点としては、現場で評価する教員の WBA に対する理解、実施に対する姿勢、評価能力に温度差があることで、Faculty Development(以下、FD)の重要性が語られていた。WBA の充実により、各大学における総括評価的が国家試験に求められる水準に耐えうるかどうかについては可能であると答えた大学と、各大学で医師 1 年目に必要な手技的なもの、例えば、BLS(Basic Life Support)や ACLS(Advanced Cardiac Life Support)などは、OSCE を併用実施したほうがよいと答えた大学が存在した。

2. 海外(主に米国)における WBA の現状について

1) ECFMG における Mini-CEX

2022 年の導入後、米国外医学部卒業者の約 15% が Step 2 CS の代替条件として Mini-CEX を選択した。選択者は発展途上国もしくは紛争地域にあり、十分な臨床実習や資格試験が行われにくい環境にあった。6 症例全てで 9 点中 4 点以上を取得する必要があるが、3 点以下の割合は 0.3% にとどまった。Mini-CEX 受験者の研修医プログラム以後の臨床実践力の検証が進められ、その結果から評価体制の見直し(例:評価尺度を EPA 基盤型にする)が議論されている。

2)米国医学部における WBA の位置付け

Step 2 CS 廃止後、各医学部は在学生の最低基準を確認するため、WBA を発展させている。WBA による総括的評価は診療科ごとの Global Practice Assessment(担当教員による 1~2 週間の観察に基づいた総合評価)が主流で、Mini-CEX は形成的評価の1つのツールと位置付けられている。Programmatic Assessment(単一の試験結果に依存する従来の評価方法と異なり、継続的な観察、ポートフォリオ、臨床現場でのパフォーマンス評価など、さまざまな評価手法を組み合わせる評価法)の導入は一部の医学部に限られた。

3. 今後の WBA のあり方について

1) WBA をどのように総括的評価に用いるのか?(用いないのか?)

- ・合否判定を伴わない評価が形骸化することは広く知られており、最終的に合否判定を行う形をめざすことが望ましい(現在の国家試験の状況に鑑みても、総括的評価のインパクトは明確)。
- ・単回の国家試験に WBA を導入するというイメージではなく、臨床実習中に診療科単位で複数回の WBA を行って、それらの結果をまとめて総括的評価を行うイメージ。

2)学生へのフィードバックの動線をどのように設計するか?

- ・診療科単位で WBA を行い、それぞれの診療科でフィードバックを行う。
- ・世界的には、WBA は、総括的評価から形成的評価へ、そして量的評価から質的評価へという潮流がある。
- ・現在、一人の学生あたり、何人の教員がどのくらいの時間をかけてフィードバックしているのかについて調査をしてみるのもよい。

3)どのような学生が不合格となるのか?

- ・よほどひどい場合でないと不合格にはならないだろう。
- ・不合格となる場合はコミュニケーション能力とプロフェッショナリズムに集約されるのではないか。
- ・複数回の観察でも合格に達しない場合は不合格とするのはどうか。
- ・再試験の設計についてはさらに検討が必要である。

4)EPA(Entrustable Professional Activities)をどのように用いるのか?(用いないのか?)

- ・診療参加型臨床実習を推進するという観点からは、EPA の考え方を取り入れた評価を設計するとよい。
- ・診療科横断的な EPA と、診療科に特異的な EPA とに分けて作成するとよい。

- ・EPA が複雑な概念で混乱を生んでいることはよく知られている。海外の先行事例を調べてみる必要もある。

5) 評価によって、どのように診療参加を促進させるのか？

- ・EPA の考え方をを用いることで診療参加を促進させることができると考える(少なくとも学生が担う役割は明確になるだろう)。
- ・医師法の改正に伴って、医学生が医行為を法的に実施できるようになっていることも考慮する必要がある。

6) どの診療科で National WBA (仮称) を実施するのか？

- ・医学教育モデル・コア・カリキュラムや JACME で定める主要な診療科(内科、外科、産婦人科、小児科、総合診療科、救急医療、地域医療)を想定する。
- ・評価が学習をドライブするという考え方からは全ての診療科で行う方がよい。

7) アンプロフェッショナルな行動・態度の評価をどのように組み込むのか？(組み込まないのか？)

- ・アンプロフェッショナルな行動・態度の評価は、観察評価でないとできないこと、また医療現場に与える影響に鑑みても、必須であろう。
- ・研修医になってから矯正することが難しいこともあるので、卒前教育において、アンプロの総括評価を設計するとよい。
- ・複数の教員・指導医による複数回の評価が必要となる。
- ・卒後研修において、アンプロフェッショナルな行動・態度のために研修が中断となった事例を調査してみるのもよい。

8) OSCE との役割分担をどのように設計するのか？

- ・OSCE でないと評価できない技能(BLS など)は OSCE で評価するのがよい。
- ・コミュニケーションスキルは WBA で行えるので、OSCE で評価する必要はあるか。

- ・OSCE はリソースを多大に必要とする試験なので、必要以上に大規模にしないことが重要である。

D. 考察

本研究は、我が国の卒前医学教育における技能・態度の評価方法として、WBA の導入可能性とそのあり方を明らかにすることを目的とした。調査の結果、国内の多くの医学部ではすでに何らかの形で WBA が導入されているものの、その多くは形成的評価にとどまっており、全診療科を通じた系統的な実施や、総括的評価としての活用には至っていないことが明らかとなった。

また、WBA の運用に関しては、指導医の理解や評価能力にばらつきがあること、教員側の負担感、評価ツールの設計や記録の煩雑さなど、制度的・運用的な課題が浮き彫りとなった。これらを克服するためには、「不合格」の基準に関する共通理解など、評価者トレーニングの設計が求められる。

一方、海外、特に米国における動向からは、WBA が既に卒前・卒後教育の双方において評価の中心となっており、診療科単位での総括的評価の導入や、Programmatic Assessment による継続的な能力評価が進められている実態が明らかとなった。米国外医学部卒業者に対する Mini-CEX の導入事例や、評価尺度の見直しに向けた議論は、我が国が WBA を国家試験の一部として制度化する際の有益な示唆を与えるものである。

E. 結論

1. 国内における WBA の現状について

今後、教員の十分な FD を前提に各大学での WBA の充実により、臨床能力の総括的評価ができる可能性が示唆された。その場合に各大学の実習と評価の質の担保が一層重要になり、第 3 者による各大学の臨床実習の評価の重要性も示されたとと言える。

2. 海外(主に米国)における WBA の現状について

米国では、各医学部の診療科単位で WBA による総括的評価が進められていた。米国外医学部卒

業者向けの Mini-CEX は、環境が整わない申請者への配慮としての側面が強く、評価尺度を含め評価体制の改善が必要と認識されていた。いずれの設定においても、卒後の臨床実践力を追跡し、評価の妥当性を検証する体制が構築されている点は大いに参考になると言える。

3. WBA のブループリント(案)

1) WBA ブループリントの目的、基本的な考え方

医師法第9条に「医師国家試験は、臨床上必要な医学および公衆衛生に関して、医師として具有すべき知識および技能について、これを行う」とある。

しかしながら、現在、実施されている医師国家試験は、多選択肢式問題によって構成され、かつ紙ベースで行われているために、技能についての評価が十分に行われていない。また、2023 年の共用試験の公的化を経て、標準化した OSCE の全国実施には多大な労力や資源が求められることがわかってきた。さらに、医学生が医行為を実施することが公的に認められることになったことを受け、診療参加型臨床実習のさらなる推進が求められている。これらの状況に鑑みると、診療現場において、医学生の行動を観察して評価する WBA を実質化することが、医学生の卒業時、すなわち国家試験受験時の能力を評価する上でも、非常に重要である。ここでは、各大学で現在散発的に行われている WBA を有効に機能させることができるよう、一定標準化した WBA のモデルをブループリントとして提示する。

2) 総括的評価の要素を取り入れるのか？

各診療科で行われる WBA について、形成的評価として、評価結果を適切に医学生にフィードバックする。また、それらを集約する形で総括的評価を実施する。基準に達していないと思われる学生は不合格とする。不合格になった場合の再試験の設計についても検討する。

3) どういった能力を評価するのか？

医学教育モデル・コア・カリキュラム(令和4年度改訂版)の「医師として求められる基本的な資質・

能力」のうち、主にコミュニケーション能力およびプロフェッショナリズムを評価する。

4) どのようなツールを用いて評価するのか？

評価に手間がかからない概略評価を基本とし、不合格とする場合には十分な観察記録が記載できるようなツールを開発する。必要に応じて、OSCE を廃止して WBA に舵を切った米国の Global Practice Assessment を参考にする。

5) どの診療科で評価を行うのか？

全ての診療科で評価を行う。

6) どのくらいの時間をかけて何回程度評価するのか？

全ての診療科で評価を行うため、少なくともローテートする診療科の数が、評価の回数になる。時間については、今後開発するツールの内容による。

7) 誰が評価するのか？

臨床実習で担当する指導医が評価する。さらに、他職種や患者からの評価、同級生による評価などについても検討する(360度評価)。

8) 評価者トレーニングはどうするのか？

不合格基準のイメージを持ってもらうことを目的とした FD(評価者トレーニング)を行う。コミュニケーション能力とプロフェッショナリズムに関して、不合格となるような例を動画で作成して見てもらう。

9) 合否判定基準をどのようにするか？(どのような学生が不合格になるのか？)

コミュニケーション能力については、共用試験 OSCE の合否判定基準も参考にしながら、合否判定基準を明確化する。プロフェッショナリズムについては、全国の複数の大学で実施されているアンプロフェッショナルな行動・態度の評価の経験を参考に、合否判定基準を明確化する。

10) 学生へのフィードバックをどのように行うのか？

診療科単位で、WBA の結果を学生に返却する

(形成的評価)。

11) OSCE との役割分担はどのような形になるのか？

Post-CC OSCEにおいてもコミュニケーション能力を評価しているため、共用試験との役割分担について、共用試験実施評価機構と対話を重ねる。プロフェッショナリズムについては、Post-CC OSCE では評価していないため、WBA で評価する。

また、標準化した OSCE を大規模に全国展開することによって、学生が臨床実習の現場から遠ざかってしまうことも明らかになっているため、全国統一の形で行う OSCE は最小限として、各大学の各診療科で OSCE より積極的に行う形を推奨する。

12) 医師国家試験との役割分担はどのような形になるのか？

各大学における WBA の合格を医師国家試験の受験要件とする。なお、各大学における WBA の質担保のため、その設計・実施について外部評価を実施する。

F. 健康危険情報

特になし

G. 研究発表

1. 論文発表

なし

2. 学会発表

なし

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

I. 研究協力者一覧 (50 音順・敬称略)

2025年3月31日時点

木村 武司 (名古屋大学医学系研究科)

近藤 諭 (富山大学学術研究部医学系 医学教育学)

【参考文献】

1. Mladenovic J, et al. Evolution of Educational Commission for Foreign Medical Graduates Certification in the Absence of the USMLE Step 2 Clinical Skills Examination. Acad Med. 2023;98:444-447.
2. Barrett A, et al. A BEME (Best Evidence in Medical Education) review of the use of workplace-based assessment in identifying and remediating underperformance among postgraduate medical trainees: BEME Guide No. 43. Med Teach. 2016; 38(12):1188-1198.
3. Anderson H, et al. Implementation and Use of Workplace-Based Assessment in Clinical Learning Environments: A Scoping Review. Acad Med. 2021;96(11S):S164-S174.
4. Ryan MS. et al. How well do workplace-based assessments support summative entrustment decisions? A multi-institutional generalisability study. MedEduc.2024;1-13.
5. Hasen R, et al. Implementing a Real-time Workplace-based Assessment Data Collection System Across an Entire Medical School's Clinical Learning Environment. MedEdPublish 2021, 10:22 Last updated: 12 SEP 2023
6. Lockyer J, et al. Core principles of assessment in competencybased medical education. Med Teach. 2017;39(6):609-616.
7. Holmboe E, et al. Using Longitudinal Milestones Data and Learning

Analytics to Facilitate the Professional Development of Residents: Early Lessons From Three Specialties. *Acad Med.* 2020 Jan;95(1):97-103.

8. Kogan JR, et al. Guidelines: The do's, don'ts and don't knows of direct observation of clinical skills in medical education. *Perspect Med Educ.* 2017 Oct;6(5):286-305.
9. Kogan JR, et al. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med.* 2014 May;89(5):721-7.
10. Kogan JR, et al. Can Rater Training Improve the Quality and Accuracy of Workplace-Based Assessment Narrative Comments and Entrustment Ratings? A Randomized Controlled Trial. *Acad Med.* 2023 Feb 1;98(2):237-247.
11. Katharine Boursicot. Performance and Workplace Assessment. *Practical Guide for Medical Teachers.*
12. David R. Brown, MD, et al. Entrustment Decision Making in the Core Entrustable Professional Activities: Results of a Multi-Institutional Study. *Academic Medicine*, Vol. 97, No. 4 / April 2022
13. The Association for the Study of Medical Education(ASME). Next steps for workplace-based assessments of entrustable professional activities. *Clin Teach.* 2024;e13739. <https://doi.org/10.1111/tct.13739>
14. Claudio Violato, PhD, et al. Validity Evidence for Assessing Entrustable Professional Activities During Undergraduate Medical Education *Acad Med.* 2021;96:S70-S75.doi:10.1097/ACM.0000000000004090.
15. Dana Dunne, MD, MHS et al. WBAs in

UME—How Many Are Needed? A Reliability Analysis of 5 AAMC Core EPAs Implemented in the Internal Medicine Clerkship. *J Gen Intern Med* 37(11):2684-90 DOI: 10.1007/s11606-021-07151-3.