作成日：令和８年６月 15 日　第１版
〔ホームページ公開用資料〕

オンライン OSTE における人工知能（AI）を用いた評価代替可能性の検証

１．臨床研究について

九州大学医学科は、日本や世界に貢献する医学・医療の担い手を育成することが、重要な使命の一つです。その使命を達成するために、医学教育講座では、どのよう教育が学生の学びにつながっているかという評価研究を行っています。

今回の研究の実施にあたっては、九州大学医系地区部局観察研究倫理審査委員会の審査を経て、研究機関の長より許可を受けています。この研究が許可されている期間は、令和 11 年 3 月 31 日までです。

２．研究の目的や意義について

＜背景＞

客観的臨床技術評価（OSTE：Objective Structured Teaching Examination）は、医学教育における客観的な指導力評価方法です。評価者は構造化されたチェックリストや評価ルーブリックに基づき、学習者のパフォーマンスを採点します。

この評価形式は高い妥当性を持つ反面、次のような課題が存在します：

- 評価者確保の困難さ： 複数名の評価者が必要であり、教員の時間的負担が大きい
- 評価者間信頼性（inter-rater reliability）の問題： 評価者によって採点にばらつきが生じやすい
- フィードバックの遅延： 採点・集計・フィードバック返却に時間がかかる
- スケーラビリティの制約： 受験者数の増加に比例して評価コストが増大する

我々は前回の研究でビデオ会議ツールを用いたオンライン形式での OSTE（オンライン OSTE）を開発しました。このオンライン化により、評価セッションの動画記録が容易に取得可能となりました。動画記録の蓄積は、AI による事後的な評価・再分析を可能にする基盤を提供しています。

1-3. 人工知能（AI）による評価の可能性

近年の生成 AI・大規模言語モデル（LLM）の発展により、音声・映像・テキストを統合したマルチモーダルな評価が現実的な技術水準に達しつつあります。医学教育分野では、OSCE の文脈において AI（特に LLM）を用いた評価の試みが報告されており、人間評価者との高い一致率を示した事例が存在する一方 OSTE については報告がありません。

1-4. 本研究の位置づけ

本研究は、収集したオンライン OSTE で収集済みの動画データおよび人間評価者によるスコアを二次利用し、AI による評価との比較検証を後ろ向きに実施します。既存データを活用することで、新たな参加者負担を生じさせることなく、AI 評価システムの実用可能性に関するエビデンスを創出することができます。

＜方法＞

オンライン OSTE の妥当性検証を行う際に取得した動画データを利用し、研究を実施します。

- AI が評価する際に必要なルーブリックを作成・精緻化する

＜検証すべき課題＞

OSTE における、AI による評価者代替の検証

＜医学的・社会的意義＞

より幅広い被験者を取り入れることが可能になります。

そのため、評価を受けることのできる絶対値を増やすことができ、収集できるデータ量が増え、より深い知見につなげることができます。たくさんの指導医が指導医評価を受け、その指導力の向上に寄与することができれば、若手指導医の指導力は医療の質に影響することが示唆されており、医療の質全体に寄与できる可能性が広がります。　すでに移動の負担や準備の負担などを減らすだけでなく、評価者の評価負担も減らすことが可能となります。

世界中どこにいても、どの言語であっても LLM を使用することで言語的・地理的な障壁を取り払うことができます。

３．研究の対象者について

【内科若手指導医】

適格基準：

１．　内科専攻医プログラムに所属している卒後 3 年目から 7 年目の医師

２．　本研究について十分理解し、本人による同意が可能な医師

除外基準：

１．　説明して同意の得られなかった内科専攻医

２．　実施時に卒後年次が 3 年目から 7 年目ではない場合

中止基準：

１．　同意の撤回があった場合

【模擬研修医】

適格基準：

１．　本研究について十分理解し、本人による同意が可能な初期研修医

２．　同意の得られた初期研修医

除外基準：

１．　説明して同意の得られなかった初期研修医

２．　研究実施時点で初期研修を修了している場合

中止基準：

１．同意の撤回があった場合

【評価者】

適格基準：

１．九州大学医学教育学講座認定の若手指導医講習会修了者

除外基準：

１．　説明して同意の得られなかった修了者

中止基準：

１．　同意の撤回があった場合

４．研究の方法について

１．これまでの報告で使用されてきた評価表の項目から、本プログラムで評価可能な評価項目を抜粋し、オンライン OSTE 用の評価表を作成します。

２．評価表の項目や内容をエキスパート数名でチェックを行う。エキスパートとは本研究チームに所属しているものとしました。

３．九州大学医学教育学講座認定の若手指導医講習会修了者に対して評価者の募集をおこないます。

４．参加希望者のうち通年で参加可能な参加者を優先に選定し、本人に別添説明文書に基づき説明を行った上で、本研究への参加について、文書にて同意を得ます。

５．共用試験 OSCE の評価者標準化講習会を参考にロールプレイなどを通し評価者の評価トレーニングを実施します。

６．実際に現場で指導を行っている医師より現実に起こりうる指導場面を基に OSTE に使用する課題を開発します。

７．シナリオ自体の妥当性検証のため、パイロット版を実施するために模擬研修医と模擬指導医を募集する。模擬研修医は科学研究費（基礎研究 B)「マイクロラーニングを活用した若手指導医養成 e ラーニングの開発と効果検証」の研究チームの所属する施設の初期研修医に募集をおこないます。

８．参加者に別添説明文書に基づき説明を行った上で、本研究への参加について、文書にて同意を得ます。

９．共用試験 OSCE の模擬患者標準化講習会を参考にロールプレイなどを通し模擬研修医の標準化を行います。

１０．数名の被験者を対象にパイロットとして OSTE を行います。

１１．パイロットを踏まえてシナリオや評価表を修正します。

１２．内科専攻医プログラムに所属する内科専攻医を研究対象者として募集する。研究対象者（模擬指導医）に関しては、ホームページに公開されている内科専攻医プログラムの責任者に募集をおこないます。

１３．本人に別途説明文書に基づき説明を行った上で、本研究への参加について文書で同意を得ます。

１４．実際に内科専攻医に対して OSTE を実施し、評価表を用い評価します。

１５．上記を踏まえて集められた評価表データを元に一般化可能性理論を活用し、信頼性係数が十分保たれる評価者数および課題数を算出します。

【以下、手順１～１５で収集したデータの二次利用】

１６．手順 15 で得られた評価表データおよびルーブリックをもとに、AI が採点する際の評価基準を整備・精緻化します。整備した評価基準の内容を研究チームのエキスパート数名で確認し、コンセンサスを得ます。

１７．収集済みのオンライン OSTE の動画データ（録画）および音声トランスクリプトを、マルチモーダル大規模言語モデル（Video-LLaVA 等）に入力し、手順 16 で整備した評価ルーブリックに基づく AI スコアを生成します。AI は言語的コミュニケーションに加え、ジェスチャー・表情・アイコンタクトなどの非言語的要素も含めた複合的なパフォーマンスを評価対象とします。

１８．AI による評価スコアと、手順 14 で人間評価者が付けた Likert スケールのスコアを照合し、評価者間一致度を級内相関係数（ICC）および Cohen の重み付き κ 係数（weighted kappa）を用いて統計的

に検証します。

１９．特定の評価項目においてAIが十分な評価精度を示さない場合は、音声認識技術を用いて動画データを文字起こしし、テキストベースの自然言語処理（NLP）モデルを併用する代替戦略を検討・適用します。

２０．Explainable AI（XAI）技術を用いてAIの判断根拠を可視化し、AIがいかなる要素に基づいて評価を行ったかを分析する。研究チームのエキスパートが、臨床指導教育の観点からその妥当性を確認し、必要に応じてAIシステムの改良に反映させます。

２１．上記の結果を論文としてまとめ、国際学術誌への投稿を行い、研究成果を公表します。

５．研究への参加とその撤回について

この研究への参加はあなたの自由な意思で決めてください。同意されなくても、不利益になることは全くありません。

また、いったん同意した場合でも、あなたが不利益を受けることなく、いつでも同意を取り消すことができます。同意を撤回されたい方又は研究対象者のご家族等の代理人の方は、下記の相談窓口までご連絡ください。

その場合は、収集された情報は廃棄され、取得した情報もそれ以降はこの研究目的で用いられることはありません。ただし、同意を取り消した時にすでに研究結果が論文などで公表されていた場合には、完全に廃棄できないことがあります。

６．個人情報の取扱いについて

研究対象者の情報をこの研究に使用する際には、研究対象者のお名前の代わりに研究用の番号を付けて取り扱います。研究対象者と研究用の番号を結びつける対応表のファイルにはクラウドでパスワードを設定し保存します。

また、この研究の成果を発表したり、それを元に特許等の申請をしたりする場合にも、研究対象者が特定できる情報を使用することはありません。

この研究によって取得した情報は、九州大学大学院医学研究院医学教育学講座・教授・新納宏昭の責任の下、厳重な管理を行います。

ご本人等からの求めに応じて、保有する個人情報を開示します。情報の開示を希望される方は、ご連絡ください。

７．試料や情報の保管等について

〔試料について〕

この研究において得られた研究対象者の情報は原則としてこの研究のために使用し、研究終了後は、九州大学大学院医学研究院医学教育学講座において同分野教授・新納宏昭和の責任の下、１０年間保存した後、研究用の番号等を消去し、廃棄します。

しかしながら、この研究で得られた研究対象者の情報は、将来計画・実施される別の医学研究にとっても大変貴重なものとなる可能性があります。そこで、前述の期間を超えて保管し、将来新たに計画・

実施される医学研究にも使用させていただきたいと考えています。その研究を行う場合には、改めてその研究計画を倫理審査委員会において審査し、承認された後に行います。

８．この研究の費用について

この研究に関する必要な費用は、部局等運営経費・講座寄附金でまかなわれます。

９．利益相反について

本研究に関する必要な経費は部局等運営経費・講座寄附金でまかなわれており、研究遂行にあたって特別な利益相反状態にはありません。

利益相反についてもっと詳しくお知りになりたい方は、下記の窓口へお問い合わせください。

利益相反マネジメント委員会

（窓口：九州大学病院 ARO 次世代医療センター　電話：０９２－６４２－５０８２）

１０．研究に関する情報の公開につい

この研究に参加してくださった方々の個人情報の保護や、この研究の独創性の確保に支障がない範囲で、この研究の研究計画書や研究の方法に関する資料をご覧いただくことができます。資料の閲覧を希望される方は、ご連絡ください。

また、この研究では、学会等への発表や論文の投稿により、研究成果の公表を行う予定です。

１１．特許権等について

この研究の結果として、特許権等が生じる可能性はありません。

１２．研究を中止する場合について

研究責任者の判断により、研究を中止しなければならない何らかの事情が発生した場合には、この研究を中止する場合があります。なお、研究中止後もこの研究に関するお問い合わせ等には誠意をもって対応します。

１３．研究の実施体制について

この研究は以下の体制で実施します。

研究実施場所	九州大学大学院医学研究院医学教育学講座	
研究責任者	九州大学大学院医学研究院医学教育学講座　准教授　菊川誠	
研究分担者	九州大学大学院医学研究院医学教育学講座　大学院生　小杉俊介	
共同研究機関等	機関名　／　研究責任者の職・氏名・(機関の長名)	役割
	①　大阪医科薬科大学　総合診療科　特定助教　橋本忠幸	総括、解析、論文作成
	②　東北大学大学院医学系研究科公衆衛生学分野 講師 香田将英	総括、解析、論文作成

１４．相談窓口について

この研究に関してご質問や相談等ある場合は、下記担当者までご連絡ください。

事務局 （相談窓口）	担当者：九州大学大学院医学研究院医学教育学講座　准教授　菊川　誠 連絡先：〔TEL〕092-642-6186 〔FAX〕092-642-6188

【留意事項】

本研究は九州大学医系地区部局観察研究倫理審査委員会において審査・承認後、以下の研究機関の長（試料・情報の管理について責任を有する者）の許可のもと、実施するものです。

九州大学大学院医学研究院長　須藤　信行