

修士論文

医学教育における動画付問題チェックリストの  
開発と問題作成支援

Development of a Video-Based Problem Checklist  
and Problem Creation Support in Medical Education

社会文化科学教育部博士前期課程教授システム学専攻

237-G8817

黄世捷

主指導：喜多 敏博 教授

副指導：戸田 真志 教授

副指導：マジュンダール リトジット 准教授

2026年1月5日

## 目次

要旨（日本語） .....	5
要旨（英語） .....	6
<b>第1章 序論</b> .....	<b>8</b>
1.1 研究の背景：医学教育評価のパラダイムシフトと社会的要請.....	8
1.2 動画付臨床問題作成における「暗黙知」の壁.....	11
1.3 研究の目的：暗黙知の形式知化による作題支援 .....	11
1.4 本論文の構成 .....	12
<b>第2章 先行研究と理論的枠組み</b> .....	<b>13</b>
2.1 医学教育における評価の変遷：コンピテンシーへのシフト .....	13
2.2 動画付臨床問題の役割と要件.....	13
2.3 良質な臨床問題作成の要件（テスト理論の視点） .....	14
2.4 動画活用における認知科学的要件（IDの視点） .....	14
2.5 本研究のアプローチ：暗黙知の形式知化 .....	15
<b>第3章 チェックリスト「ACREF」の設計と開発</b> .....	<b>16</b>
3.1 開発の背景とコンセプト：暗黙知の形式知化.....	16
3.2 ACREFモデルの詳細定義：3つの評価領域.....	17
3.2.1 Group A: 妥当性（Appropriateness）.....	17
3.2.2 Group C: 構成（Construction） .....	17
3.2.3 Group R: 洗練（Refinement） .....	17
3.3 運用マニュアルと重要ルール.....	18
3.3.1 三要素の相互依存性とネタバレ排除 .....	18
3.3.2 活用プロセス .....	18
<b>第4章 形成的評価の実施方法</b> .....	<b>19</b>
4.1 評価の目的.....	19
4.2 調査対象者 .....	19
4.3 使用した動画素材.....	20

4.4 実施手順.....	20
4.5 データ収集と分析方法.....	21
<b>第 5 章 形成的評価の結果 .....</b>	<b>22</b>
5.1 作成された問題の概要.....	22
5.1.1 提示した動画の概要（動画のイメージ図と説明） .....	22
5.1.2 SME-I（脳神経内科医）の作成した動画 1 問題 .....	23
5.1.3 SME-I（脳神経内科医）の作成した動画 2 問題 .....	25
5.1.4 SME-N（代謝内分泌内科医）の作成した動画 1 問題 .....	27
5.1.4 SME-N（代謝内分泌内科医）の作成した動画 2 問題 .....	30
5.2 チェックリスト評価の定量分析.....	33
5.2.1 ACREF チェックリスト（集計表） .....	33
5.2.2 SME-I の作成した動画 1 問題に対する ACREF チェックリスト .....	34
5.2.3 SME-I の作成した動画 2 問題に対する ACREF チェックリスト .....	35
5.2.4 SME-N の作成した動画 1 問題に対する ACREF チェックリスト.....	36
5.2.5 SME-N の作成した動画 2 問題に対する ACREF チェックリスト.....	37
5.3 インタビューに基づく定性分析.....	38
5.3.2 セルフ評価と他者評価（ピア・エキスパート）の齟齬 .....	39
5.3.4 ACREF チェックリストの改善案と運用方法 .....	40
5.4 インストラクショナル・デザイン（ID）専門家によるレビュー .....	41
5.4.1 ACREF チェックリストの構造・記法に関する提言：GROUP R: 洗練の再定義 .....	41
5.4.2 動画の不完全さに対する提言.....	42
5.4.3 修正アクションを特定するためのトリアーゼ機能.....	42
<b>第 6 章 考察 .....</b>	<b>43</b>
6.1 暗黙知の形式知化としての ACREF の有効性.....	43
6.2 動画によって生じる専門家・非専門家の齟齬とタグによる回避.....	43
6.3 評価基準のアンカリング.....	44
6.4 チェックリストの改訂ポイントと運用 .....	44
<b>第 7 章 結論 .....</b>	<b>46</b>

7.1 総括.....	46
7.2 今後の展望.....	46
<b>参考文献および付録.....</b>	<b>47</b>
付録1 動画付き MCQ「ACREF」チェックリスト研究ご協力をお願い.....	49
付録2 動画付き MCQ 作成支援チェックリスト(ACREF)運用マニュアル.....	51
付録3 動画付き MCQ 作成支援チェックリスト (ACREF).....	55
付録4 SME-I に対するインタビュー文字起こし.....	63
付録5 SME-N のインタビューの文字起こし.....	78
付録6 エキスパートレビューの文起こし.....	92

## 要旨（日本語）

日本の医学教育は、令和 5 年の医師法改正による診療参加型臨床実習の合法化を経て、実践的な診療能力の質保証が求められるフェーズにある。しかし、現行の評価手法には、知識を問うテキストベースの多肢選択問題（MCQ）と、実技を評価する OSCE との間に「動的な視覚・聴覚情報の解釈」を問う手法が欠如しているという課題がある。このギャップを埋める「動画付臨床問題」は重要視されているが、その作成には医学的知識に加え、テスト理論や認知負荷を考慮した編集技術といった高度なスキルが求められる。現状、これらのスキルは一部の熟練者の「暗黙知」に留まっており、組織的な作題体制の構築を阻んでいる。

本研究は、SECI モデルを理論的枠組みとし、熟練作題者の暗黙知を言語化・構造化（形式知化）することを目的とした。具体的には、動画問題の品質を担保するための評価指標「ACREF (Appropriateness: 妥当性、Construction: 構成、Refinement: 洗練)」を開発した。本指標は、最低限遵守すべき必須項目（Essential）と質的向上を目指す推奨項目（Further）の 2 段階基準で構成されている。

指標の有効性を検証するため、領域専門家(SME)2名を対象とした形成的評価を実施した。SME による作題、自己評価、他者評価、および筆者によるエキスパート評価の結果を定量的・定性的に分析した。その結果、ACREF は経験の浅い作題者にとっての「道しるべ」や、教員間の「共通言語」として機能し、客観的なフィードバックを可能にすることが示された。一方で、専門家が動画内の微細な臨床的矛盾に拘泥し、学生レベルでは良問とされるものを不適切と判断する「熟達者の盲点 (Expert Blind Spot)」という課題が浮き彫りになった。この齟齬を回避するため、動画を「タグ (メタデータ)」によって抽象化する手法や、評価前に「臨床実習後の医学生」という対象読者を明示する「アンカリング」の重要性が確認された。

本研究が提案する ACREF および運用モデルは、個人の経験則に依存していた作題プロセスを組織的な品質保証へと転換させる一助となる。今後は、対象診療科の拡大や学習管理システム (LMS) への実装、さらには蓄積されたデータセットを活用した生成 AI による作題支援システムへの発展が展望される。

## 要旨（英語）

Japanese medical education is currently undergoing a significant transition toward participatory clinical clerkships following the 2023 amendment of the Medical Practitioners' Act. This transition requires robust quality assurance of learners' practical clinical competencies. However, a critical gap exists between traditional text-based multiple-choice questions (MCQs), which primarily assess declarative knowledge, and Objective Structured Clinical Examinations (OSCEs), which evaluate observable clinical performance. Specifically, there is a lack of assessment methods that evaluate learners' ability to interpret dynamic visual and auditory clinical information.

To bridge this gap, video-based clinical questions have gained importance. Nevertheless, developing high-quality video-based problems requires not only medical accuracy but also expertise in assessment theory, as well as video editing skills grounded in instructional design principles to manage cognitive load. At present, these competencies remain largely as the tacit knowledge of a limited number of experienced educators, impeding the establishment of an organized and sustainable item development process.

This study adopts the SECI model as its theoretical framework and aims to externalize and systematize the tacit knowledge of expert item writers into explicit knowledge. To this end, a quality assurance checklist named ACREF—comprising Appropriateness, Construction, and Refinement—was developed. The ACREF checklist consists of a two-tier structure: Essential criteria representing minimum requirements and Further criteria designed to enhance educational quality.

To examine the effectiveness of ACREF, a formative evaluation was conducted involving two Subject Matter Experts (SMEs). The evaluation process included item development, self-assessment, peer review, and expert review by the author. Quantitative and qualitative analyses demonstrated that ACREF functions as an effective scaffold for novice item writers and serves as a common language facilitating objective feedback among educators.

A notable finding was the identification of an Expert Blind Spot, whereby specialists overemphasized subtle clinical inconsistencies based on advanced intuition, occasionally judging problems appropriate for student-level assessment as inadequate. To mitigate this discrepancy, the study confirmed the importance of abstracting video content through tags (metadata) and implementing anchoring, which explicitly defines the target audience prior to evaluation.

The ACREF checklist and its operational model support a shift in video-based item development from reliance on individual expertise toward organizational quality assurance in medical education. Future directions include expansion to additional clinical disciplines, integration into Learning Management Systems (LMS), and development of AI-assisted item creation support systems using accumulated datasets.

## 第1章 序論

### 1.1 研究の背景：医学教育評価のパラダイムシフトと社会的要請

日本の医学教育は、卒前・卒後を一貫した「シームレスな医師養成」を目指し、かつてない変革期にある。現在、日本には82の医学部が存在し、各大学が自主性を持って教育課程を編成しているが、その実態は出口評価である「医師国家試験」に強く規定されている。医師として診療を行うためには、医学部卒業後に400問のMCQ（多肢選択問題）で構成される医師国家試験に合格し、さらに2年間の臨床研修を修了することが必須である。

この医師国家試験は合格率90%以上（令和6年度新卒者実績で95.0%）[1]で推移する一方で、不合格となれば医師としてのキャリアを開始できない極めてハイスタークな試験である。そのため、大学教育はどうしても国家試験合格を最優先とした「知識偏重」のカリキュラムに傾斜しやすく、本来涵養すべき実践的な診療能力の教育・評価がおろそかになりがちであるという構造的な課題を長らく抱えてきた。

この「知識偏重」から脱却し、実践能力を担保するために導入されたのが「診療参加型臨床実習（Clinical Clerkship）」である。従来、医学生の実習は法的な制約により「見学型」ととどまることが多かったが、これでは研修医として必要なプライマリ・ケア能力を卒前に獲得することが困難であった。

事態を打開すべく、令和5年（2023年）に医師法が改正され、医学生が指導医の監督下で臨床実習として医行為を行うことが合法化された[2]<sup>1</sup>。これにより、実習は「見学」から「実践」へと大きな転換が図られている。

医学生が患者に対して医行為を行う以上、その技能と安全性を実習開始前に厳格に担保しなければならない。そこで、これまで各大学が実施していた共通試験が、法的な裏付けを持つ「公的試験」へと格上げされた（2024年施行）。

---

<sup>1</sup> 医学生がより診療参加型の臨床実習において実践的な実習を行うことを推進し、医師の資質向上を図る観点から、「共用試験」に合格した医学生について、医師法第17条の規定にかかわらず、大学が行う臨床実習において、医師の指導監督の下、医療に関する知識及び技能を修得するために医業を行うことができることとする。

臨床実習前の共用試験は、知識を問う CBT（Computer Based Testing）と、技能・態度を問う実習前 OSCE（Objective Structured Clinical Examination）から構成されており、これらに合格することが臨床実習に参加するための法的要件となった（表 1-1 参照）[3]これにより、医学教育における評価は、大学の自主管理から、公的な質保証のフェーズへと移行したのである。

表 1-1:医学部における CBT・OSCE と臨床実習

学年	試験名	概要
医学部 4 年生	共用試験 CBT <sup>*1</sup> (Computer Based Testing)	知識の評価目的 (公的試験)
	臨床実習前 共用試験 OSCE <sup>*2</sup> [Pre-Clinical Clerkship OSCE]	基本的臨床能力の 実技試験 (公的試験)
医学部 5 年生	臨床実習(Clinical Clerkship)	診療参加型臨床実習
医学部 6 年生	臨床実習後 共用試験 OSCE <sup>*3</sup> [Post-Clinical Clerkship OSCE]	応用的な臨床能力の 実技試験
	医師国家試験 (国家試験)	400 問 MCQ 筆記試験

\*1\*2\*3 いずれも共用試験。

\*1\*2 は臨床実習で、指導医のもとで許された範囲の医療行為をすることを許可してよいかを判定する試験。(令和 5 年から、医師法で定めた公的試験)

\*3 公的試験ではないが、各大学の卒業判定に関わる。

このように、制度面では「実践型」への転換が進んでいるが、それを支える「評価手法」には依然として限界が存在する。臨床実習後に実践的な診療能力を評価する公的手段は、現状では「臨床実習後 OSCE<sup>2</sup>」[3]（公益社団法人 医療系大学間共用試験実施評価機構, 2025）のみである。一方、医師国家試験は依然としてテキストベースの MCQ（冊子形式）で実施されており、実際の臨床現場で不可欠な「動的な視覚・聴覚情報の解釈」や、時間経過を伴う「文脈に応じた臨床推論」までは十分に評価しきれしていない。

OSCE は実施コストが高く頻回な実施が困難であるため、ペーパーテスト（知識）と OSCE（実技）の間にあるこの「評価のギャップ」を埋めるための打開策として、「動画付臨床問題」の導入と促進が必要だと考える。

---

<sup>2</sup>臨床実習後 OSCE は、医学教育モデル・コア・カリキュラムに準拠して行われます。学生が診療参加型臨床実習で十分な臨床能力を修得し、医学部卒業を許容できる臨床能力、換言すれば、卒業後の医師臨床研修を開始するにあたって安全・安心な医療を提供できる能力を修得できているかを全国統一の基準で評価する試験です。一方では、国民・社会に対し、わが国のすべての医学部・医科大学が医師育成機関としての責務を全うしているかを評価されることにもなります。（共用試験実施評価機構共用試験ガイドブック 23 版[3]より引用）

## 1.2 動画付臨床問題作成における「暗黙知」の壁

動画付臨床問題の重要性は、既に政策レベルでは認識されている。厚生労働省の医師国家試験改善検討部会の報告書では、国家試験の CBT 化を見据え、動画や音声などのマルチメディア素材を活用した出題が推奨されている。これを受け、厚生労働科学研究（河北班）による「ICT を利用した医学教育コンテンツの開発と活用に向けた研究」[4]など、質の高い教育コンテンツの開発も進められてきた。

しかし、各大学レベルでの導入状況を見ると、その普及は極めて限定的である。[5]、[6]その背景には、現行の国家試験で動画が出題されていないことによるインセンティブの欠如や、患者由来の動画素材を管理するインフラの問題が存在する。加えて、普及を阻む最大の障壁は、インフラ以上に「良質な問題を作成できる領域専門家（Subject Matter Expert: SME）の不足」にある。

動画問題の作成には、①医学的な正確性、②テストとしての公正・妥当性という従来の要件に加え、③動画独自の情報を診断の鍵として配置する構成力、④認知負荷を考慮した編集技術という、複合的なスキルが求められる。講義資料として動画を用いる教員は多いが、これを「厳密な試験問題」として成立させられる教員は稀である。

現状、こうした高度な作題スキルは一部の熟練者の「暗黙知」として蓄積されるに留まっており、形式知化されていない。その結果、作題者による品質のばらつきや、静止面による「動く紙芝居」のような動画を用いる必然性のない不適切な問題が作成されるリスクが生じている。組織として質の高い問題を安定供給するためには、この暗黙知を共有可能な形式知へと変換する必要がある。

## 1.3 研究の目的：暗黙知の形式知化による作題支援

そこで本研究では、知識創造理論である SECI モデル[7]を理論的枠組みとし、熟練作題者の暗黙知を言語化・構造化（形式知化）することを目的とする。具体的には、動画付臨床問題の品質を担保するための評価指標を策定し、経験の浅い SME でも利用可能なチェックリスト「ACREF（Appropriateness, Construction, Refinement）」を開発する。なお、本チェックリストは導入の容易さを考慮し、必須項目（Essential）と推奨項目（Further）の2段階基準を設定する。

本研究の成果は、個々の教員の作題負担を軽減するだけでなく、教員間での相互評価（ピアレビュー）における共通言語として機能する。これにより、個人の経験則に依存していた作題プロセスを脱却し、医学教育機関全体としての評価の質的向上に寄与することが期待される。

#### 1.4 本論文の構成

本論文は以下の構成をとる。第2章では、医学教育およびID理論の観点から動画問題の要件を整理する。第3章では、チェックリストACREFの設計プロセスを述べる。第4章・第5章では、実際のSMEを対象とした形成的評価の方法と結果を報告し、第6章でその考察、第7章でその有効性と今後の展望について述べる。

## 第2章 先行研究と理論的枠組み

### 2.1 医学教育における評価の変遷：コンピテンシーへのシフト

日本の医学教育は、文部科学省が策定した「医学教育モデル・コア・カリキュラム」[8]に基づき、知識の習得だけでなく「資質・能力（Competency）」の育成へと舵を切っている。特に、患者の症候から病態を推論し判断する「臨床推論能力」の涵養は中核的なコンピテンシーと位置づけられており、これには臓器横断的な知識だけでなく、診断に必要な情報を取得・解釈するための医療面接や身体診察といった実技技能が不可欠である。

医学教育評価の枠組みとして知られる Miller のピラミッド[9]において、従来の筆記試験は「Knows（知識）」から「Knows how（知識の応用）」を、OSCE（客観的臨床能力試験）は「Shows how（実演）」を評価するものであるとされる[10]。しかし、実臨床現場で求められる「視覚的・聴覚的所見の解釈（見て/聞いて判断する能力）」については、既存の評価手法の間にギャップが存在する。テキストや静止画を用いる従来のペーパーテストでは、動的な所見の変化を評価することができない。一方で、OSCE では「Shows how」としてこの能力を評価することは理論上可能だが、模擬患者（SP）を用いる性質上、麻痺や眼振といった「健常者では表現困難な病的所見」を扱うことには限界がある。さらに、OSCE の実施には莫大な人的リソースを要する点も課題である。例えば、公的試験である Pre-CC OSCE では一大学あたり延べ 200 人以上の運営スタッフや SP を要し[11]、Post-CC OSCE においても機構課題と大学独自課題の実施で手一杯であるのが現状である。そのため、評価の網羅性を高めるために OSCE の課題数を増やすことは、運用コストの観点から現実的ではない。

### 2.2 動画付臨床問題の役割と要件

この評価のギャップを埋めるツールとして、動画・音声付問題が位置づけられる。ガニエの学習成果の分類[12]に従えば、動画問題は単なる「言語情報（知識）」の再生ではなく、動的な視覚情報を識別し、概念を適用する「知的技能」を評価する手法である。これを語学試験（英語）に置き換えて類推すると、従来の MCQ（テキスト問

題)が「Reading (読解)」の評価であり、OSCE (対人実技)が「Speaking (会話・発話)」の評価であるならば、動画付臨床問題は「Listening (聴解・観察)」の評価に相当すると言える。すなわち、OSCE ほどの人的コストをかけずに、Reading よりも実践に近い「情報の受容と解釈」の技能を評価できる点に、動画問題の独自の価値がある。

臨床推論能力の評価手法として、Nayer ら (2018) は「Key-Feature Questions (KFQ)」の有効性を提唱している[13]。Key-Feature (KF) とは、「臨床問題の解決における重要なステップ (significant step)」と定義され、診断や管理において学習者が最も誤りを犯しやすい決定的な局面を指す。従来のテストでは KF をテキストや静止画で提示せざるを得なかったが、動画を用いることで、実際の症例でしか再現できない動的な KF (例：一過性の不随意運動や特有の歩行障害) を問題に盛り込むことが可能となる。

### 2.3 良質な臨床問題作成の要件 (テスト理論の視点)

動画付臨床問題の作成にあたっては、まず前提として多肢選択問題 (MCQ) としての品質が担保されなければならない。Haladyna らは、選択肢の粒度を揃えることや、不適切なヒント (Cueing) を排除することなど、31 の作題ルール (Taxonomy of item-writing rules) を提唱している[14]。これらはテストの信頼性と妥当性を保つための基本的な「作法」である。

さらに医学教育特有の要件として、米国医師国家試験 (USMLE) を所管する National Board of Medical Examiners は、単なる「知識の再生」ではなく「知識の適用」を問うための「臨床シナリオ (Clinical Vignette)」の重要性を説いている。特に、選択肢を見なくとも設問文だけで状況が理解でき、正答が推測できる状態を目指す「Cover-the-options rule」は、良質な臨床問題の必須条件とされる[15]。

### 2.4 動画活用における認知科学的要件 (ID の視点)

動画はリッチな情報源である反面、学習者に過度な認知負荷 (Cognitive Load) を与えるリスクがある。Mayer & Moreno (2003) は、マルチメディア学習の効果を高める

原則として、学習目標に関係のない情報を削除する「Weeding（雑草取り）の原理」や、重要な部分に注意を誘導する「Signaling（合図）の原理」、視覚情報と聴覚情報の無駄な重複を避ける「Redundancy（冗長性）の排除」を提唱している。[16]動画問題の作成においては、これらの原理を応用し、診断に関係のないノイズ（外因性認知負荷）を最小化する「編集力」が求められる。

医学教育用ビデオの品質評価としては「IVQC (Instructional Video Quality Checklist)」[17]などが存在するが、これらは主に「教材」としての品質を問うものであり、「評価（テスト）」特有の要件（ネタバレ防止や選択肢との整合性）まではカバーしていない。ここに、評価用動画問題に特化した新たなチェックリスト開発の必要性がある。

## 2.5 本研究のアプローチ：暗黙知の形式知化

上記の「テスト理論（医学的妥当性）」と「ID理論（認知的妥当性）」を統合し、実際に良質な動画問題を作成するには、熟練教員の経験則（暗黙知）が不可欠である。本研究では、SECIモデルを枠組みとして採用し、熟練者の暗黙知を表出化

(Externalization) することで、経験の浅い作題者でも参照可能な「ACREF（形式知）」を構築するアプローチをとる。一般に、暗黙知の表出化には多数の専門家による合意形成（Delphi法など）が用いられることが多い。しかし、動画付臨床問題の作題経験を持つ教員は現状極めて少なく、大規模な調査は困難である。そのため本研究では、ゼロベースでの合意形成を行うのではなく、動画作題のエキスパートである筆者自身の経験と先行研究を基盤としてプロトタイプを構築し、それを実証的に検証するアプローチを採用した。

## 第3章 チェックリスト「ACREF」の設計と開発

### 3.1 開発の背景とコンセプト：暗黙知の形式知化

医学教育において、動画・音声付問題は「実践的な臨床推論能力（Clinical Reasoning）」を評価するために不可欠なツールである。従来の筆記試験では知識の再生を、OSCEでは実技手技を評価できるが、実際の診療で求められる「視覚・聴覚情報からの所見の認知と解釈」を効率的に測定するには、マルチメディアを用いた評価が最適解となる。しかし、その作成には医学知識だけでなく、メディア特性を理解した高度な判断能力が求められる。現状、講義資料として動画を用いることはあるものの、これらを厳密な試験問題として導入する事例は限定的であり、良質な動画付臨床問題を作成するためのノウハウは、一部の熟練した個人の経験則（暗黙知）に留まっている。その結果、作題者間での質のばらつきや、動画が単なる装飾に留まる不適切な問題の発生が生じている。

良質な問題を組織的かつ持続的に蓄積するためには、熟練医が無意識に行っている作題プロセスを言語化・構造化（形式知化）し、問題作代経験の浅い領域専門家

（Subject Matter Expert: SME）でも再現可能な「チェックリスト」として提供する必要がある。本研究におけるチェックリスト開発は、知識創造理論であるSECIモデルにおける、個人の暗黙知を共有可能な概念へと変換する「表出化

（Externalization）」と、それを体系的な知識として組み合わせる「連結化

（Combination）」を支援する試みとして位置づけられる[7]。

本研究では、動画・音声付臨床問題作成の経験が豊富な筆者が、これまでの作題経験において良問・悪問と判断した事例を分析し、それらを弁別可能な指標として言語化した。開発されたチェックリスト「ACREF」は、評価の「妥当性（Appropriateness）」、問題としての「構成・品質（Construction）」、学習効果を高める「洗練・工夫（Refinement）」という3つの評価領域（ACR）によって体系化されている。さらに、多忙な医学教育現場への導入障壁を下げするため、最低限遵守すべき「必須（Essential）」項目と、より高い教育効果を目指すための「推奨（Further）」項目の2水準を設定し、段階的な質向上を促す設計とした。

## 3.2 ACREF モデルの詳細定義：3つの評価領域

### 3.2.1 Group A: 妥当性 (Appropriateness)

Group A は、動画・音声を用いることの「医学的・教育的必然性」を問う、本モデルの根幹となる領域である。ここでは、テキストや静止画では代替できない「動き（時間的変化）」や「音（聴覚情報）」こそが、診断において決定的な判断（Key Feature） [13]となる症例を選定しているかを規定する。これは、臨床推論能力を効果的に評価するための「Key-Feature Questions (KFQ)」の概念に通じるものであり、単に珍しい症例を見せるのではなく、「決定的な判断に動画が必要か？」という着眼点を評価するものである。具体的には、A-1（動画の必須性）、A-2（動きの必須性）、A-3（音声の必須性）といった項目により、メディア利用の根拠を明確化している。

### 3.2.2 Group C: 構成 (Construction)

Group C は、テストとしての「公平性」と「整合性」を担保する領域である。設問文・動画・選択肢の3要素が矛盾なく機能しているか、また選択肢の粒度が揃っているかなど、テスト理論 (Psychometrics) 的な品質を規定する。不適切な選択肢作成や設問の不備は、学生の知識に関わらず正答を導けてしまう「技術的欠陥 (Item Writing Flaws)」 [15]として知られており、これらを排除することは評価の信頼性を保つ上で不可欠である。本領域では、NBME のガイドライン等に基づき、C-3（選択肢の適切性）や C-1（品質の担保）などの項目を設定した。

### 3.2.3 Group R: 洗練 (Refinement)

Group R は、学習者の認知負荷を最適化するための領域である。この領域の項目を満たしていなくても問題としては成立するが、学習効果を高めるためには、不必要な情報の排除や、学習者の能力に応じた足場かけ (Scaffolding) を担う。Mayer & Moreno (2003) は、マルチメディア学習における認知負荷を低減する方法として、重要な情報に合図を送る「Signaling (合図) の原理」や、学習に関係のない興味深い情報を排除する「Weeding (雑草取り) の原理」を提唱している [16]。ACREF ではこれらの原理に基づき、R-1（学習支援の工夫：矢印やズームの活用）や、R-3（設問の公正性：不要なノイズの排除）を定義した。

### 3.3 運用マニュアルと重要ルール

#### 3.3.1 三要素の相互依存性とネタバレ排除

「設問文・動画・選択肢」から構成される動画問題においては、それぞれのコンポーネントが単独で完結するのではなく、相互に依存する。そのため、設問文が動画の内容を説明してしまう「ネタバレ (Cueing)」の排除が不可欠である。NBME のガイドラインでは、選択肢を見なくても設問文だけで正答が推測できる「Cover-the-options rule」が推奨されているが、[15]動画問題においては「動画を見なければ正答が特定できない (設問文だけでは情報が不足する)」状態を作ることが求められる。また、テキストと動画で同一の情報を重複して提示することは、Mayer & Moreno (2003) が指摘する「Redundancy (冗長性) の原理」に抵触し、学習者の認知処理を阻害する恐れがある。[16]したがって、C-4 (設問文の必須性) および A-4 (所見の非開示) により、テキストと動画の適切な役割分担を規定した。

#### 3.3.2 活用プロセス

本チェックリストは、単独で使用するものではなく、作題プロセス全体の中に組み込まれることを想定している。具体的には、作題者自身による「自己点検 (Self-check)」の指針として、また、複数の教員間で行う「ピアレビュー (Peer-review)」における共通言語として機能する。このプロセスを経ることで、個人の暗黙知を組織的な形式知へと昇華させるとともに、作題経験の乏しい教員にとっては、良質な問題を作成するための「足場かけ」として機能することが期待される。

## 第 4 章 形成的評価の実施方法

### 4.1 評価の目的

本評価の目的は、ACREF チェックリストを用いて医学部教員（Subject Matter Expert: SME）が動画付臨床問題を作成・評価するプロセスを観察し、ツールの「実用性（Usability）」と指標の「妥当性（Validity）」を検証することである。

本研究では、ACREF チェックリストと所定の動画素材を提示した上で、複数名の SME が動画付臨床問題を作成するトライアルを実施した。評価においては、作題者自身による自己評価（Self-check）、および自分以外の作題者と筆者による他者評価（Peer/Expert-check）を行い、その評価結果の一致・不一致を観察する。なお、本評価ではあらかじめ撮影・編集済みの動画素材を提供することで、作題者の「動画撮影・編集技能」による影響を排除し、純粋に「動画を用いた作題能力（着眼点や設問設計）」と「ACREF による評価の揺らぎ」を検証することに焦点を当てた。

### 4.2 調査対象者

調査対象は、筆者と同じ大学病院に所属し、医学教育に携わる医師・教員 2 名（SME-I, SME-N）とした。両名とも内科専門医資格および領域別の専門医資格、医学博士号を有しており、高い臨床能力と研究能力を備えている。また、日常的に系統講義や臨床実習指導、試験問題作成に従事している。特筆すべき点として、SME-I は動画付問題の作題経験を有する一方、SME-N は未経験であり、この経験差が ACREF の利用にどのような影響を与えるかも観察の対象とした。

- **SME-I**： 臨床経験 20 年。総合内科専門医・脳神経内科専門医・医学博士。動画問題作成経験あり（教材として約 10 題、試験問題として約 6 題の作成実績）。
- **SME-N**： 臨床経験 10 年以上。総合内科専門医・糖尿病内科専門医・医学博士。動画問題作成経験なし。

### 4.3 使用した動画素材

作題に使用する素材として、以下の2つの動画を選定した。

- **動画1：「振戦（不随意運動）」の症例動画**
- **動画2：「片麻痺（臥位での神経学的評価）」の再現動画**

これらの症例は、静止画では表現困難な「時間的变化」や「動きの質」を伴う所見であり、ACREFのGroup A（妥当性）を検証するのに適した題材である。また、動画1（振戦）はSME-I（脳神経内科）とSME-N（糖尿病内科）の双方の専門領域に関連する所見である一方、動画2（片麻痺）はSME-Iの専門領域であるが、一般的な内科診療でも遭遇しうる所見である。このように、専門性との親和性が異なる素材を用いることで、作題のバリエーションや着眼点の差異を観察することを意図した。

### 4.4 実施手順

研究依頼(付録1)に同意したSME-IおよびSME-Nに対して、動画付きMCQ作成支援チェックリスト運用マニュアル(付録2)、動画付きMCQ作成支援チェックリスト、(付録3)および動画を提示した。評価は以下の4段階のフェーズで実施した。

1. **作題フェーズ：**対象者は提供された動画を用い、ACREFを参照しながら臨床問題を作成する。なお、作題の負荷を軽減し評価の本質に集中させるため、標準的な臨床問題の形式（設問文、選択肢）に準拠したテンプレートを提供した。
2. **セルフチェック：**作成した問題に対し、作題者自身がACREF(付録3)を用いて評価を行う。
3. **ピアチェック：**相手が作成した問題に対し、ACREF(付録3)を用いて相互に評価を行う。
4. **エキスパートチェック：**筆者（およびIDの知見を有する専門家）が、作成されたすべての問題をACREF(付録3)を用いて評価する。

## 4.5 データ収集と分析方法

収集したデータは、定量的分析と定性的分析の両面から評価した。定量的データとして、3者（自己・他者・専門家）間でのチェックリスト評価結果の一致率を算出し、評価基準の明確さを検証する。定性的データとして、実施後のインタビュー(付録 4.5)や記述回答から、ACREF の使用感、判断に迷った項目、および作題意図を抽出する。これらのインタビューおよび結果に関して、IDer であり医学教育の専門家によるレビュー・インタビューからチェックリストの改善（Refinement）に向けた示唆を得る。（付録 6）

## 第5章 形成的評価の結果

### 5.1 作成された問題の概要

#### 5.1.1 提示した動画の概要（動画のイメージ図と説明）

動画1の概要：「振戦（不随意運動）」の症例動画



両手の上に置いた手が震えている様子。震えの方向は左右方向で粗動な動き。

明らかな左右差はなく、挙上するとさらに震えは明瞭になる。

手首の底屈-背屈ではなく左右の震えを認める。

動画2の概要：「片麻痺（臥位での神経学的評価）」の再現動画



臥位の模擬患者。左手の上腕は屈曲し、拳が胸の上にある状況。左足はやや外転（つま先が外側を向いている）医師の指示で右手、左手、右足、左足の順で挙げるよう指示され、右手・右足は正常に動く。左手と左足は最初の段階から全く動かない。

5.1.2 SME-I（脳神経内科医）の作成した動画 1 問題

動画 1 を使用した問題



設問文

60 歳の男性。

現病歴：5 年前にコップを持っているときに手が震えていることに気づいた。3 年前から徐々にひどくなり安静にしているときにも震えを自覚した。心配になり受診した。

既往歴：高血圧症にてアムロジピン内服中。20 年前に虫垂炎にて手術を受けた。

生活歴：喫煙は 20 本/日を 10 年間。飲酒はビール 350ml を週 3 回。職業は事務職。

家族歴：特記すべきことはない。

現症：意識は清明(JCS または GCS)。身長 150cm、体重 70kg。体温 36.5°C。心拍数 60/分、整。血圧 145/90mmHg。呼吸数 15/分。SpO2 98%(room air)。

身体所見：眼瞼結膜に貧血を認めず、眼球結膜に黄染を認めない。頸部リンパ節を触知しない。診察の動画を示す。

最も考えられる疾患の特徴を確認するために追加で聞くべき質問はどれか。

選択肢

- a.「ご出身はどちらですか？」
- b.「動悸がすることはありますか？」
- c.「歩きづらいことはありますか？」
- d.「著しい頭痛が出たことはありますか？」
- e.「ご家族には同じ症状の方はいらっしゃいますか？」

正答:e

#### 解説

5年前からの振戦、徐々に悪化を認めている。鑑別疾患としては疾患頻度の最も多いものとしては本態性振戦、治療可能な病態として甲状腺機能亢進症、パーキンソン病、肝性脳症があげられる。動画では振戦があり、動作時に悪化している。動作時優位の振戦であり、本態性振戦の特徴に一致する。

a.× 遺伝子性疾患を想定した場合有効な質問であるが、本態性振戦では地域性はない。

b.× 動悸がすることはありますか？」

c.× パーキンソン病を鑑別するために重要な質問であるが、本態性振戦の特徴ではない。d.× 本態性振戦では頭痛は出現しない。

e.○ 本態性振戦では家族歴を有することが多い。

### 5.1.3 SME-I（脳神経内科医）の作成した動画 2 問題

#### 動画 2 を使用した問題



#### 設問文

80 歳の男性。

現病歴：5 年前に脳梗塞を発症し麻痺が残存した。3 日前より咳、発熱があり、それまで歩行が可能であったが歩行できなくなり受診した。

既往歴：脳梗塞にてアスピリン内服中。高血圧、糖尿病、脂質異常症あり。6 年前に心筋梗塞にてカテーテル治療を受けた。

生活歴：喫煙は 10 本/日を 20 年間、5 年前より禁煙している。飲酒はビール 350ml を週 1 回。職業は無色。

家族歴：特記すべきことはない。

現症：意識は清明。身長 175cm、体重 75kg。体温 36.2℃。心拍数 90/分、整。血圧 143/76mmHg。呼吸数 16/分。SpO2 98% (room air)。

身体所見：眼瞼結膜に貧血を認めず、眼球結膜に黄染を認めない。頸部リンパ節を触知しない。胸部聴診にて含気良好、右下肺野で湿性ラ音を聴取。心音に雑音を認めない。

診察の動画を示す。

診察所見について正しいのはどれか。

#### 選択肢

- a. 右上肢は MMT4 である。
- b. 右下肢は MMT3 である。
- c. 左上肢は MMT2 である。
- d. 左下肢は MMT1 である。

e.この診察では MMT は評価できない。

正答:d

解説

a.右上肢は MMT4 である。

b.右下肢は MMT3 である。

c.左上肢は MMT2 である。

d.左下肢は MMT1 である。

e.この診察では MMT は評価できない

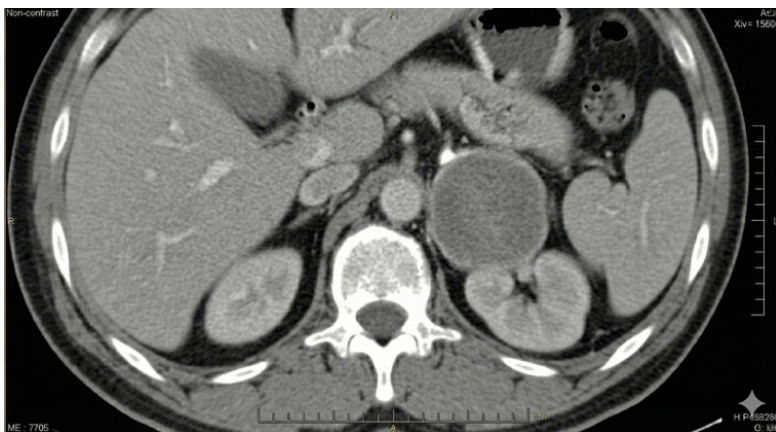
#### 5.1.4 SME-N（代謝内分泌内科医）の作成した動画 1 問題

##### 動画 1 を使用した問題



##### 設問文

**【問題文】** 45 歳の男性。動画に示す症状を主訴に来院した。 **現病歴:** 3 か月前より、発作的に主訴を自覚するようになったため受診した。症状は数十分から 1 時間程度で軽快するが、週に数回繰り返している。頭痛や動悸、発汗を伴うこともある。 **既往歴:** 特記すべきことはない。 **身体所見:** 身長 158 cm、体重 49 kg。体温 36.8℃。脈拍 110/分(整)。血圧 176/104 mmHg。眼球突出や甲状腺腫大を認めない。心音と呼吸音に異常を認めない。腸蠕動音は低下している。四肢の固縮は認めない。 **検査所見:** 尿所見:タンパク(-)、糖(+)。血液生化学所見:空腹時血糖 126 mg/dL、HbA1c 6.2%、Na 140 mEq/L、K 4.2 mEq/L、Cl 102 mEq/L。TSH 2.5  $\mu$ U/mL(基準 0.5~5.0)、FT4 1.2 ng/dL(基準 0.9~1.7)。腹部 CT を別に示す。次のうち、最も考えられるのはどれか。



##### 選択肢

a 褐色細胞腫

- b Basedow 病
- c Parkinson 病
- d Cushing 症候群
- e 原発性アルドステロン症

正答 a 褐色細胞腫

解説

1. 振戦の評価と関連症状:

- 患者は「安静時振戦」を主訴とし、動画(想定)でその様子が示されている。また、この振戦が「発作的に生じ、数十分から1時間程度で軽快し、頭痛、動悸、発汗を伴う」という点から交感神経が過剰に反応している病態を示唆する。
- 選択肢の b) Basedow 病(甲状腺中毒症)や c) Parkinson 病も振戦をきたすが、それぞれ以下の点で異なる。
  - **Basedow 病:** 頻脈、高血圧、振戦のほか、典型的には眼球突出や甲状腺腫大、発汗、体重減少が見られるが、本症例では TSH、FT4 が正常範囲であり(TSH 2.5  $\mu$ U/mL、FT4 1.2 ng/dL)、Basedow 病は否定的である。
  - **Parkinson 病:** 主に安静時振戦、筋強剛(固縮)、無動、姿勢反射障害が特徴。本症例は「四肢の固縮は認めない」と明記されており、Parkinson 病は否定的である。また、発作性の頭痛や動悸、発汗は Parkinson 病の典型的な症状ではない。

2. 身体所見と検査所見からの絞り込み:

- 高血圧: 来院時血圧 176/104 mmHg と高血圧を認める。褐色細胞腫の「一症状である可能性を考える。
- 高血糖: 空腹時血糖 126 mg/dL、HbA1c 6.2%、尿糖(+)と高血糖を認める。これはカテコラミンのインスリン拮抗作用によって引き起こされることがある。
- 体重減少: 体重 49 kg、BMI 19.6 kg/m<sup>2</sup> と比較的低値であり、代謝亢進を示唆する可能性がある。
- 腸蠕動音低下: カテコラミンの作用による消化管運動抑制を示唆する所見である。

- 電解質: K 4.2 mEq/L と正常範囲であり、低 K 血症を特徴とする d) Cushing 症候群や e) 原発性アルドステロン症の可能性を下げる。
- Cushing 症候群: 副腎性 Cushing 症候群の場合、副腎腫瘍が併存することが特徴である。満月様顔貌、中心性肥満、皮膚線条、高血圧、低 K 血症などを特徴とするが、本症例の症状とは合致せず。振戦も典型的ではない。
- 原発性アルドステロン症: 高血圧と低 K 血症を特徴とし、副腎腫瘍も併存することがある。振戦や発作性の頭痛、動悸、発汗は典型的ではない。

### 3. 画像所見(腹部 CT):

- 腹部 CT 画像では、左副腎に腫瘍性病変が描出されている。  
【Gemini に生成させたため、髄質系の腫瘍の特徴を完全には再現できていない。腫瘍の CT 値はもう少し高値であり、内部が不均一であるとより良い】

### 結論

患者の「発作性の頭痛、動悸、発汗、高血圧、振戦、高血糖」という症状と、除外された他の疾患、そして腹部 CT 画像での副腎腫瘍性病変の存在から、褐色細胞腫が最も考えられる診断である。

### 5.1.4 SME-N（代謝内分泌内科医）の作成した動画 2 問題<sup>3</sup>

#### 動画 2 を使用した問題



#### 設問文

**【問題文】** 64 歳の男性。動画に示す症状を主訴に救急車で搬入された。**現病歴**：1 時間前、自宅でテレビを見ていたところ、突然激しい頭痛と嘔気が出現し、主訴を伴ったため家族が救急要請した。**既往歴**：50 歳頃から高血圧を指摘されていたが、治療を中断していた。**身体所見**：意識レベルは JCS I -1。身長 165 cm、体重 72 kg。体温 36.6℃。脈拍 84/分(整)。血圧 192/110mmHg。その他の神経学的所見では、両眼の右側への共同偏視を認める。瞳孔不同はない。

次のうち、最も考えられるのはどれか。

#### 選択肢

- a 右被殻出血
- b 右視床出血
- c 右小脳出血
- d 左橋出血
- e 左頸髄出血

正答 a 右被殻出血

解説

<sup>3</sup> SME-N は代謝・内分泌内科の専門医であり、動画 1、動画 2 の問題作成に当たっては Google Gemini 2.5 による作題補助・CT 画像生成を行っている。

### 動画と身体所見の解釈:

- **動画の所見:** 動画では、患者は右上下肢を指示通り動かしていますが、**左上下肢は全く動かさない(あるいは重力に抗して保持できない)** 状態である。これは**左完全片麻痺**を示している。左側の麻痺なので、責任病巣は対側である**右側の錐体路(大脳～脳幹)**にあると考えられます。
- **病歴:** 高血圧の既往がある患者における、活動時の突然の頭痛、嘔気、嘔吐、片麻痺の発症は、脳血管障害の中でも**脳出血**を最も強く疑うエピソードである。
- **眼球所見:** 「両眼の**右側**への共同偏視」を認めている。

### 共同偏視(Gaze Preference)の法則:

- **大脳(被殻など)の病変:** 病巣側にある「前頭葉眼球運動野」などが障害されるか、刺激されることで眼球偏位が起こる。出血や梗塞の場合、眼球は病巣側(健側を向く筋肉が麻痺するため、病巣側を向く)に偏位します。
- 本症例では**右**を向いているため、**右大脳半球**の病変が疑われます。

### 各選択肢の検討:

- **a 右被殻出血(正解)**
  - 右被殻の出血は内包後脚へ波及し、対側である**左片麻痺**をきたす。また、右大脳半球の病変であるため、両眼は**右側へ共同偏視**する。
- **b 右視床出血**
  - 右視床出血でも左片麻痺をきたすが、視床出血では眼球が**内下方(鼻先)**を向く「**鼻先凝視**」が特徴的である。縮瞳や対光反射の消失を伴うことが多い。本症例のような水平性の共同偏視は被殻出血の方が典型的である。
- **c 右小脳出血**
  - 小脳出血では、激しい回転性めまい、嘔吐、体幹失調(起立・歩行不能)が主症状である。通常、四肢の運動麻痺(片麻痺)は脳幹を強く圧迫しない限り認めない。
- **d 左橋出血**

- 橋出血では、病巣側(左)の顔面神経麻痺と対側(右)の片麻痺(交叉性麻痺)をきたす可能性がある。本症例は左麻痺なので矛盾する。また、橋出血における共同偏視は、大脳とは逆に「麻痺側(病巣の反対側)を睨む」のが特徴です。もし左橋出血なら、右麻痺に対して眼球は右を向くことになる。さらに、橋出血では高度の縮瞳(pinpoint pupil)が典型的である。

- **e 左頸髄出血**

- 頸髄の障害では病巣側(左)の片麻痺をきたすが、共同偏視は起こらない。また、頭痛や嘔気も典型的ではない。

## 5.2 チェックリスト評価の定量分析

### 5.2.1 ACREF チェックリスト (集計表)

グループ	No.	チェック項目	E (Essential)	F (Further)	注釈
A: Appropriateness  (動画使用の妥当性)	A-1E	【動画の必須性】 設問文と選択肢だけでは解答できず、動画（音声）の視聴が必須である。	<input type="checkbox"/>		(不適切) ・設問文の情報（例：病歴、検査結果）だけで選択肢が絞り込める。 ・設問文の病歴と選択肢の所見名（例：「羽ばたき振戦」）を組み合わせるだけで、正解が推測できる。
	A-2E	【"動き"の必須性】 動画は、「動き」や「時間経過」の確認が必須な所見（静止画では不十分な所見）を含んでいる。	<input type="checkbox"/>		(不適切) 静止画でも伝わる所見（例：皮疹の色調）。 (適切) 「動き」や「変化」が診断の鍵となる所見（例：皮疹の庄退色、呼吸に伴う頸静脈の変動）。
	A-3E	【"音声"の適切性】 (※音声所見が重要な場合) 動画は、解答に必須な"音声"（聴診音や動作指示など）を含んでおり、かつ、学生が解釈すべき所見や解釈そのものを音声で説明していない（例：「腫れています」など）	<input type="checkbox"/>		(不適切) ・解答に必要な情報（例：指示音声、聴診音）がミュートされている。 ・音声（検者）が「腫れています」「熱感があります」「脈が弱いです」のように、学生が解釈すべき所見や解釈そのものを説明している。
	A-4E	【所見の非開示】 設問文は、動画から学生が読み取るべき核心的な所見を、文章で記述していない。	<input type="checkbox"/>		(不適切) 設問文に「...両手を背屈させると、羽ばたきのような振戦が見られた」と、動画で評価すべき所見が文章で記述されている。
	A-5F	【臨床推論の促進】 設問（動画・設問文・選択肢の全体）は、単一所見の認知（所見名当て）でなく、複数の所見を統合し、病態生理まで考慮する臨床推論を要求している。		<input type="checkbox"/>	(不適切) 「片側下肢浮腫」の動画で「下肢エコー」を1つ選ばせる。 (所見の認知のみで回答)  (適切) 「片側下肢浮腫+頸静脈怒張」の動画で「下肢エコー」と「心エコー」の2つを選ばせる。（所見の認知と病態の解釈）
	A-6F	【学習体験の重要性】 頻度は低くとも臨床的に重要な（例：重篤な疾患、見落としやすい所見）事象を扱うことで、学生の知識と実践を結びつける学習体験を提供している。		<input type="checkbox"/>	(不適切) 臨床的重要性の低い動画。  (適切) 知識としては知っていても臨床で未経験な重要所見を示す動画（羽ばたき振戦など）。
C: Construction  (教材の構成・品質)	C-1E	【品質の担保】 動画（音声）は、所見を識別し、臨床的な判断を行うために十分な品質である。	<input type="checkbox"/>		(不適切) 臨床判断に必要な最低限の品質を満たさない ・画質：解像度や照度が低い、ピンボケしている。 ・画角：ズームしすぎ（部位不明瞭）またはパンしすぎ（所見不明瞭）。 ・カメラワーク：手ブレが酷い、所見の途中でカメラが動く。  (注：満たさない場合、出題動画の見直しや、作題補助者への動画の再編集を依頼)
	C-2E	【文脈の明示】 設問文は、動画を解釈する前提となる文脈（例：場面設定）を明確に示している。また、動画が再現（シミュレーション）である場合は、そのことを明記している。	<input type="checkbox"/>		(適切) 「...の再現動画を示す」と明記し、所見の矛盾・不一致を回避する。（筋力低下があるのに萎縮がない、など）  (補足) 文脈（例：患者背景）を記述する際、動画で見ると見えない所見（例：「羽ばたきのような振戦」）を文章で記述しないよう注意する。 (A-4Eの不適切に相当)
	C-3E	【選択肢の適切性】 選択肢は、「粒度」（例：身体所見の種類、検査の種類）が揃っている。 また、所見を問う場合、単なる「見たままの記述」ではなく、「鶏歩」のような医学的概念で構成されている。	<input type="checkbox"/>		(不適切) 眼球運動の動画を示した上で、選択肢がa. 眼振、b. 瞳孔不同 (適切) 眼球運動の動画を示した上で、選択肢がa. 垂直性眼振、b. 水平性眼振  (不適切) 所見がそのまま選択肢になっている。「つま先から歩いている」 (適切) 所見の解釈が選択肢になっている。「鶏歩」
	C-4E	【設問文の必須性】 動画と選択肢だけでは解答できず、設問文の情報（例：患者背景、具体的な問い）が解答に必須である。	<input type="checkbox"/>		(不適切) 設問文の文脈（例：患者背景）が不足しているため、動画所見（例：失調）と選択肢（例：脳梗塞）だけで解答できてしまう。
	C-5F	【"実臨床"の再現性】 動画（音声）は、臨床的に妥当な範囲で、実臨床に近い文脈（例：環境音、診断上除外すべき陰性所見）を意図的に含んでいる。		<input type="checkbox"/>	(適切) ・心雑音の評価時に、鑑別に必要なJaneway斑を否定するための爪の視診（陰性所見）を含めること。
R: Refinement  (学習支援の工夫・洗練)	R-1F	【学習支援の工夫】 動画は、学習効果を高めるための意図的な編集（スロー再生、矢印、健側との比較、ズームなど）が加えられている。		<input type="checkbox"/>	常に必須ではないが、学習者レベル：（臨床実習前か、後かなど）にあわせた調整。 ・スロー再生：持続時間が短いあるいは速い所見に対して。 ・矢印：複数の所見が同時に存在する場合、問いたい所見を明確化するため。 ・健側との比較：筋萎縮や腱反射など、個人差が大きい所見の「左右差」を明示するため。
	R-2F	【真正性（リアリティ）】 提示されたシチュエーション（特に学習者が担う役割・立場）は、学習者が「自分ごと」として捉えられ、かつ「実臨床であり得る」ものである。		<input type="checkbox"/>	(不適切なアンカリング) ・学習者の経験から遠い、または過度に専門的な役割（例：「あなたは指導医です」「あなたは腫瘍内科医です」）。  (適切なアンカリング) ・学習者が近い将来に担う一般的・緊急的な役割（例：「あなたは研修医です」「コードブルーに最初に到着した」）。 (例：不公正な引っかけ)
	R-3F	【設問の公正性】 設問文および動画は、本質的な評価とは無関係な「引っかけ」要素（＝臨床的に非妥当な情報）を意図的に排除している。		<input type="checkbox"/>	・心音聴診の動画で、鑑別と無関係な四肢MMTの動画を意図的に含めること。

セルフチェックを1、ピアチェックを2、エキスパートピアのチェックを4として集計

## 5.2.2 SME-I の作成した動画 1 問題に対する ACREF チェックリスト

使用動画			動画1	動画1	動画1	動画1	動画1	動画1	
問題作成者			SME-I	SME-I	SME-I	SME-I	SME-I	SME-I	
評価者			SME-I	SME-I	SME-N	SME-N	黄	黄	
評価分類			セルフ	セルフ	ピア	ピア	エキスパート ピア	エキスパート ピア	
グループ	No.	チェック項目	E (Essential)	F (Further)	E (Essential)	F (Further)	E (Essential)	F (Further)	
A: Appropriateness (動画使用の妥当性)	A-1E	【動画の必須性】 設問文と選択肢だけでは解答できず、動画 (音声) の視聴が必須である。	1		2		4		7
	A-2E	【“動き”の必須性】 動画は、「動き」や「時間経過」の確認が必 須な所見(静止画では不十分な所見)を含ん でいる。	1		2		4		7
	A-3E	【“音声”の適切性】 (※音声所見が重要な場合) 動画は、解答に必要な“音声”(聴診音や動作指 示など)を含んでおり、かつ、学生が解釈す べき所見や解釈そのものを音声で説明してい ない (例:「腫れています」など)	1		2		4		7
	A-4E	【所見の非開示】 設問文は、動画から学生が読み取るべき核 心的な所見を、文章で記述していない。	1		2		4		7
	A-5F	【臨床推論の促進】 設問(動画・設問文・選択肢の全体)は、単 一所見の認知(所見名当て)でなく、複数の 所見を統合し、病態生理まで考慮する臨床推 論を要求している。		0		2		0	2
	A-6F	【学習体験の重要性】 頻度は低くとも臨床的に重要な(例:重篤な 疾患、見落としやすい所見)事象を扱うこと で、学生の知識と実践を結びつける学習体験 を提供している。		0		2		0	2
C: Construction (教材の構成・品質)	C-1E	【品質の担保】 動画(音声)は、所見を識別し、臨床的な判 断を行うために十分な品質である。	1		2		4		7
	C-2E	【文脈の明示】 設問文は、動画を解釈する前提となる文脈 (例:場面設定)を明確に示している。また、 動画が再現(シミュレーション)である場合 は、そのことを明記している。	1		2		4		7
	C-3E	【選択肢の適切性】 選択肢は、「粒度」(例:身体所見の種類、 検査の種類)が揃っている。 また、所見を問う場合、単なる「見たままの 記述」ではなく、「鶏歩」のような医学的概 念で構成されている。	1		2		0		3
	C-4E	【設問文の必須性】 動画と選択肢だけでは解答できず、設問文の 情報(例:患者背景、具体的な問い)が解答 に必須である。	0		2		4		6
	C-5F	【“実臨床”の再現性】 動画(音声)は、臨床的に妥当な範囲で、実 臨床に近い文脈(例:環境音、診断上除外す べき陰性所見)を意図的に含んでいる。		0		2		0	2
R: Refinement (学習支援の工夫・洗練)	R-1F	【学習支援の工夫】 動画は、学習効果を高めるための意図的な編集(スロー再生、矢印、 疑問との比較、ズームなど)が加えられている。		0		2		4	6
	R-2F	【真正性(リアリティ)】 提示されたシチュエーション(特に学習者が 担う役割・立場)は、学習者が「自分ごと」 として捉えられ、かつ「実臨床であり得る」 ものである。		0		2		0	2
	R-3F	【設問の公正性】 設問文および動画は、本質的な評価とは無関 係な「引っかけ」要素(=臨床的に非妥当な 情報)を意図的に排除している。		0		2		0	2

セルフチェックを1、ピアチェックを2、エキスパートピアのチェックを4として集計

### 5.2.3 SME-I の作成した動画 2 問題に対する ACREF チェックリスト

使用動画			動画2	動画2	動画2	動画2	動画2	動画2	
問題作成者			SME-I	SME-I	SME-I	SME-I	SME-I	SME-I	
評価者			SME-I	SME-I	SME-N	SME-N	黄	黄	
評価分類			セルフ	セルフ	ピア	ピア	エキスパートピア	エキスパートピア	
グループ	No.	チェック項目	E (Essential)	F (Further)	E (Essential)	F (Further)	E (Essential)	F (Further)	
A: Appropriateness (動画使用の妥当性)	A-1E	【動画の必須性】 設問文と選択肢だけでは解答できず、動画(音声)の視聴が必須である。	1		2		4		7
	A-2E	【"動き"の必須性】 動画は、「動き」や「時間経過」の確認が必須な所見(静止画では不十分な所見)を含んでいる。	1		2		4		7
	A-3E	【"音声"の適切性】 (※音声所見が重要な場合) 動画は、解答に必須な"音声"(聴診音や動作指示など)を含んでおり、かつ、学生が解釈すべき所見や解釈そのものを音声で説明していない(例:「腫れています」など)	1		2		4		7
	A-4E	【所見の非開示】 設問文は、動画から学生が読み取るべき核心的な所見を、文章で記述していない。	1		2		4		7
	A-5F	【臨床推論の促進】 設問(動画・設問文・選択肢の全体)は、単一所見の認知(所見名当て)でなく、複数の所見を統合し、病態生理まで考慮する臨床推論を要求している。		1		0		0	1
	A-6F	【学習体験の重要性】 頻度は低くとも臨床的に重要な(例:重篤な疾患、見落としやすい所見)事象を扱うことで、学生の知識と実践を結びつける学習体験を提供している。		1		0		0	1
	C: Construction (教材の構成・品質)	C-1E	【品質の担保】 動画(音声)は、所見を識別し、臨床的な判断を行うために十分な品質である。	1		2		4	
C-2E		【文脈の明示】 設問文は、動画を解釈する前提となる文脈(例:場面設定)を明確に示している。また、動画が再現(シミュレーション)である場合は、そのことを明記している。	1		0		0		1
C-3E		【選択肢の適切性】 選択肢は、「粒度」(例:身体所見の種類、検査の種類)が揃っている。 また、所見を問う場合、単なる「見たままの記述」ではなく、「鶏歩」のような医学的概念で構成されている。	1		2		4		7
C-4E		【設問文の必須性】 動画と選択肢だけでは解答できず、設問文の情報(例:患者背景、具体的な問い)が解答に必須である。	1		0		0		1
C-5F		【"実臨床"の再現性】 動画(音声)は、臨床的に妥当な範囲で、実臨床に近い文脈(例:環境音、診断上除外すべき陰性所見)を意図的に含んでいる。		1		2		0	3
R: Refinement (学習支援の工夫・洗練)		R-1F	【学習支援の工夫】 動画は、学習効果を高めるための意図的な編集(スロー再生、矢印、健側との比較、ズームなど)が加えられている。		0		2		0
	R-2F	【真正性(リアリティ)】 提示されたシチュエーション(特に学習者が担う役割・立場)は、学習者が「自分ごと」として捉えられ、かつ「実臨床であり得る」ものである。		1		2		0	3
	R-3F	【設問の公正性】 設問文および動画は、本質的な評価とは無関係な「引っかけ」要素(=臨床的に非妥当な情報)を意図的に排除している。		1		0		0	1

セルフチェックを1、ピアチェックを2、エキスパートピアのチェックを4として集計

## 5.2.4 SME-N の作成した動画 1 問題に対する ACREF チェックリスト

使用動画			動画1	動画1	動画1	動画1	動画1	動画1	
問題作成者			SME-N	SME-N	SME-N	SME-N	SME-N	SME-N	
評価者			SME-N	SME-N	SME-I	SME-I	黄	黄	
評価分類			セルフ	セルフ	ピア	ピア	エキスパートピア	エキスパートピア	
グループ	No.	チェック項目	E (Essential)	F (Further)	E (Essential)	F (Further)	E (Essential)	F (Further)	
A: Appropriateness (動画使用の妥当性)	A-1E	【動画の必須性】 設問文と選択肢だけでは解答できず、動画(音声)の視聴が必須である。	0		0		4		4
	A-2E	【"動き"の必須性】 動画は、「動き」や「時間経過」の確認が必須な所見(静止画では不十分な所見)を含んでいる。	1		2		4		7
	A-3E	【"音声"の適切性】 (※音声所見が重要な場合) 動画は、解答に必要な"音声"(聴診音や動作指示など)を含んでおり、かつ、学生が解釈すべき所見や解釈そのものを音声で説明していない(例:「腫れています」など)	0		2		4		6
	A-4E	【所見の非開示】 設問文は、動画から学生が読み取るべき核心的な所見を、文章で記述していない。	1		2		4		7
	A-5F	【臨床推論の促進】 設問(動画・設問文・選択肢の全体)は、単一所見の認知(所見名当て)でなく、複数の所見を統合し、病態生理まで考慮する臨床推論を要求している。		1		2		4	7
	A-6F	【学習体験の重要性】 頻度は低くとも臨床的に重要な(例:重篤な疾患、見落としやすい所見)事象を扱うことで、学生の知識と実践を結びつける学習体験を提供している。		1		2		4	7
	C: Construction (教材の構成・品質)	C-1E	【品質の担保】 動画(音声)は、所見を識別し、臨床的な判断を行うために十分な品質である。	0		2		4	
C-2E		【文脈の明示】 設問文は、動画を解釈する前提となる文脈(例:場面設定)を明確に示している。また、動画が再現(シミュレーション)である場合は、そのことを明記している。	1		2		4		7
C-3E		【選択肢の適切性】 選択肢は、「粒度」(例:身体所見の種類、検査の種類)が揃っている。 また、所見を問う場合、単なる「見たままの記述」ではなく、「鶏歩」のような医学的概念で構成されている。	1		2		4		7
C-4E		【設問文の必須性】 動画と選択肢だけでは解答できず、設問文の情報(例:患者背景、具体的な問い)が解答に必須である。	1		0		4		5
C-5F		【"実臨床"の再現性】 動画(音声)は、臨床的に妥当な範囲で、実臨床に近い文脈(例:環境音、診断上除外すべき陰性所見)を意図的に含んでいる。		1		0		0	1
R: Refinement (学習支援の工夫・洗練)		R-1F	【学習支援の工夫】 動画は、学習効果を高めるための意図的な編集(スロー再生、矢印、健側との比較、ズームなど)が加えられている。		1		0		0
	R-2F	【真正性(リアリティ)】 提示されたシチュエーション(特に学習者が担う役割・立場)は、学習者が「自分ごと」として捉えられ、かつ「実臨床であり得る」ものである。		1		0		0	1
	R-3F	【設問の公正性】 設問文および動画は、本質的な評価とは無関係な「引っかけ」要素(=臨床的に非妥当な情報)を意図的に排除している。		1		0		4	5

セルフチェックを1、ピアチェックを2、エキスパートピアのチェックを4として集計

### 5.2.5 SME-N の作成した動画 2 問題に対する ACREF チェックリスト

使用動画			動画2	動画2	動画2	動画2	動画2	動画2	
問題作成者			SME-N	SME-N	SME-N	SME-N	SME-N	SME-N	
評価者			SME-N	SME-N	SME-I	SME-I	黄	黄	
評価分類			セルフ	セルフ	ピア	ピア	エキスパートピア	エキスパートピア	
グループ	No.	チェック項目	E (Essential)	F (Further)	E (Essential)	F (Further)	E (Essential)	F (Further)	
A: Appropriateness (動画使用の妥当性)	A-1E	【動画の必須性】 設問文と選択肢だけでは解答できず、動画（音声）の視聴が必須である。	1		0		4		5
	A-2E	【“動き”の必須性】 動画は、「動き」や「時間経過」の確認が必須な所見（静止画では不十分な所見）を含んでいる。	1		2		4		7
	A-3E	【“音声”の適切性】 (※音声所見が重要な場合) 動画は、解答に必須な“音声”（聴診音や動作指示など）を含んでおり、かつ、学生が解釈すべき所見や解釈そのものを音声で説明していない（例：「腫れています」など）	1		2		4		7
	A-4E	【所見の非開示】 設問文は、動画から学生が読み取るべき核心的な所見を、文章で記述していない。	1		0		4		5
	A-5F	【臨床推論の促進】 設問（動画・設問文・選択肢の全体）は、単一所見の認知（所見名当て）でなく、複数の所見を統合し、病態生理まで考慮する臨床推論を要求している。		1		2		4	7
	A-6F	【学習体験の重要性】 頻度は低くとも臨床的に重要な（例：重篤な疾患、見落としやすい所見）事象を扱うことで、学生の知識と実践を結びつける学習体験を提供している。		1		2		4	7
	C: Construction (教材の構成・品質)	C-1E	【品質の担保】 動画（音声）は、所見を識別し、臨床的な判断を行うために十分な品質である。	1		2		4	
C-2E		【文脈の明示】 設問文は、動画を解釈する前提となる文脈（例：場面設定）を明確に示している。また、動画が再現（シミュレーション）である場合は、そのことを明記している。	1		2		4		7
C-3E		【選択肢の適切性】 選択肢は、「粒度」（例：身体所見の種類、検査の種類）が揃っている。 また、所見を問う場合、単なる「見たままの記述」ではなく、「鶏歩」のような医学的概念で構成されている。	1		2		4		7
C-4E		【設問文の必須性】 動画と選択肢だけでは解答できず、設問文の情報（例：患者背景、具体的な問い）が解答に必須である。	1		2		4		7
C-5F		【“実臨床”の再現性】 動画（音声）は、臨床的に妥当な範囲で、実臨床に近い文脈（例：環境音、診断上除外すべき陰性所見）を意図的に含んでいる。		1		2		0	3
R: Refinement (学習支援の工夫・洗練)		R-1F	【学習支援の工夫】 動画は、学習効果を高めるための意図的な編集（スロー再生、矢印、健側との比較、ズームなど）が加えられている。		1		2		0
	R-2F	【真正性（リアリティ）】 提示されたシチュエーション（特に学習者が担う役割・立場）は、学習者が「自分ごと」として捉えられ、かつ「実臨床であり得る」ものである。		1		0		4	5
	R-3F	【設問の公正性】 設問文および動画は、本質的な評価とは無関係な「引っかけ」要素（＝臨床的に非妥当な情報）を意図的に排除している。		1		0		4	5

セルフチェックを1、ピアチェックを2、エキスパートピアのチェックを4として集計

## 5.3 インタビューに基づく定性分析

評価入力後、SME-I（付録4）およびSME-N（付録5）に対し、それぞれ約30分間の対面による半構造化インタビューを実施した。得られた発言データおよびチェックリストへの記述コメントを定性的に分析した結果、ACREFの有用性や専門性による評価の乖離に関して、主に以下の4つの観点が抽出された。（付録4、付録5）。

### 5.3.1 ACREF チェックリストの有用性と課題

ACREFの導入効果について、作題経験の有無にかかわらず、両名のSMEから「作題の足場かけ（Scaffolding）」および「共通言語」としての有用性が認められた。特に作題未経験者であるSME-Nは、ACREFを作題プロセスにおける「道しるべ」と表現し、経験の浅い教員に対するガイドとして機能すると評価した。

*「動画付問題の作成経験のない教員にとって、道しるべとしては、とても良いと思います...問題をレビューする際には文脈的に『この問題は、ここが不適切だね』といったことは言われますけど、こういうチェックリストがあることでちゃんと項目がクリアになっていいなと思いました。」（SME-N）*

一方、経験者であるSME-Iもまた、自身の経験則（暗黙知）がACREFによって適切に言語化されている点を評価した。

*「今までに作った時に陥りがちなピットフォールを、こういう風に説明すればわかるんだな、という...見直すポイントというのが、ものすごく要約されていて、非常に良いと思いました。」（SME-I）*

しかし、UI/UXデザイン上の課題も指摘された。具体的には、項目の重み付け（Essential/Further）の意図が伝わりにくい点や、チェックボックスにチェックを入れる行為が「適合（Positive）」なのか「不適合（Negative）」なのか直感的に判別しづらい点である。

*「そのEssentialかFurther、EまたはFという点は、ちょっと特殊だったので、一瞬間くら*

ったところがあったかなと思います。」(SME-N) 「その項目が”満たしている”でチェックなのか、”満たしていない”でチェックなのがちよっと分かりにくかった」(SME-I)

「除外している」といった否定形の項目において、チェックが入ることがポジティブな評価に繋がる設計が、利用者の認知負荷を高めたことが示唆された。

### 5.3.2 セルフ評価と他者評価（ピア・エキスパート）の齟齬

定量分析において、ピアおよびエキスパートが高い評価（4点または6点）を与えているにもかかわらず、作題者自身のセルフ評価が低い（不適切とする）事例が、特に Group A（妥当性）において観察された。インタビューによる深掘りの結果、この乖離の主因は、作題者が自身の「高い専門性」を基準に評価を行っていたことにあると判明した。SME-I（専門医）は、自身の専門領域の問題に対し、「専門医の視点では動画を見なくても解けてしまう」あるいは「設問文だけで情報が十分である」と判断し、厳格に自己評価を下げていた。

「左頸髄出血しか左の病変がない。加えて、頸髄病変では共同偏視は出ないので、動画を診なくても問題文だけで解ける。...（専門医視点では）不適切な動画という評価をしている。」(SME-I)

他者が「学生レベルでは良問である」と判断しても、作題者本人は「専門家として不完全である」と判断してしまうこの事象は、ACREF 単独では解決できない課題であり、評価基準の前提となる「対象読者 (Target Audience)」の認識合わせが必要であることが明らかになった。

「『問題を作成する教員にとって』なのか『問題を解く学生のレベルで』なのかが書かれていない...チェックリスト利用者のアンカリングができていない」(SME-I)

### 5.3.3 動画問題だからこそ起きる「医学のアート」の多義性

テキストベースの筆記試験では、記述された所見は揺るぎない「事実」として固定される。しかし、動画を用いた場合、そこには解釈の幅が生じ、専門家ほど多くの情報を受け取ってしまうという特性が明らかになった。

SME-N は、非専門領域の動画に対しては「大枠の所見」のみを捉えてスムーズに作題できた一方、SME-I は専門領域の動画に対して「臨床的な不一致や矛盾」に気づいてしまい、作題に難渋した。SME-N はこの現象を「医学のアート (Art)」の部分に起因すると分析している。

*「文字だけであれば、「振戦」という情報に多様性がない一方、動画になると「振戦」の情報量が増える。だから観る人によって所見の取り方が変わる。だから医学のアートの部分（専門性の領域・知識を超えた能力）が、動画の部分だと個人個人によって分かれてしまう」(SME-N)*

専門家ゆえに「正解のない動画」に見えてしまうこのジレンマに対し、解決策として、生成 AI を用いた理想的な動画の作成や、再現動画であることを明示することで「不完全さ」を許容させるアプローチが提案された。

#### 5.3.4 ACREF チェックリストの改善案と運用方法

以上の結果を踏まえ、ACREF を実運用に乗せるための具体的な改善案が提示された。専門家と非専門家では評価の不一致が生じることを前提とし、システムと運用の両面からアプローチする必要性が示唆された。

評価プロセスの改善: Essential 項目を満たしていなくても直ちに却下せず、ピア・エキスパートレビューのサイクルに回すことで、作題者が見落としている視点（良問の可能性）を救い上げる。

*「チェックが付かない項目に対して、IDer である私が、具体的な改善案を示すと、納得が得られると考えます。」(筆者/SME-N との対話)*

アンカリングの明示: チェックリスト使用前に、「医学生レベルで評価する」という前提条件（アンカリング）を明文化する。

セルフ評価の重み付けの見直し: 専門家によるセルフ評価は過度に厳格になる傾向があるため、その重み付けを軽くし、非専門家（ピア・エキスパート）が「適切」と判断した問題を積極的に採用するアルゴリズムを検討する。

「専門家であることが、必ずしも問題を作るのに良いとは限らないでしょう。」(SME-I)

この結果は、動画付臨床問題の作成においては、一人の「完璧な専門家」に依存するのではなく、ACREF を介した「専門家・非専門家・ID 専門家」による多層的なチームアプローチこそが質を担保することを示している。

5.4 インストラクショナル・デザイン (ID) 専門家によるレビュー  
形成的評価として、ID および医療者教育を専門とする浅田義和氏(教授システム学修士・大学医学部医学教育部門准教授・日本医学教育学会認定医学教育専門家)に対し、ACREF の構造的妥当性および SME インタビューで得られた課題に対する解決策についてエキスパートレビューを求めた。(付録 6) 以下に、提示された 3 つの主要な知見を示す。

#### 5.4.1 ACREF チェックリストの構造・記法に関する提言：Group R: 洗練の再定義

ACREF の記法について、将来的な項目の拡張性やデータベース管理を見据え、優先順位がコード自体から判別可能な記法（例：A-E1 ではなく、AE-1）への変更が提案された。また、Group R（洗練）にのみ、Essential 項目が存在しない点について指摘があった。R はいずれも問題としての価値を高めるものとして定義したため、Essential 項目を設けなかったが、A や C がレイヤーモデルにおけるレベル 0(SME 的要件) だとすると、R-3F は満たしていない場合、レベル-1(いらつきのなさ) となるため、本項目は R における Essential 項目とすべきとの提案であった。

この指摘により、Group A・C は「テストの成立要件」を、Group R は「学習体験の質」を担保する領域であるという、ACREF の階層構造が明確化された。

## 5.4.2 動画の不完全さに対する提言

インタビューで見られた「動画の細部と病態の不一致」に対し、浅田氏は「動画という生データ (Raw Data) をそのまま評価対象にすることのリスク」を指摘した。専門家ほど、動画内のノイズに過剰反応してしまう現象に対し、解決策として「タグ (メタデータ) による抽象化」が提案された。

タグを加えることによって、ACREF の評価対象を「動画単体と問題の組み合わせ」から「動画+タグと問題の組み合わせ」への変更を行うことが可能となる。タグにより、動画が不十分であっても、タグによって「この動画は痙性麻痺として扱う」という約束事 (教育的定義) が作題者と評価者で共有されることで、作題者である専門医は、動画の不完全さから解放され、教育的意図に基づいた作題が可能となる。また、評価者も動画自身に改善点があることと問題の質を分離して評価することが可能となる。

またタグ付けに関しても領域専門家以外の専門家が加わることで、過度に専門的すぎるタグ付けを回避できるとの意見があった。

## 5.4.3 修正アクションを特定するためのトリアージ機能

ACREF の 3 領域 (A・C・R) の区分けについて、浅田氏は単なる分類以上に、評価後の「修正アクション (誰が何を直すか)」を特定するための診断機能としての重要性を強調した。

エキスパートレビューでは、A-4E では”【所見の非開示】設問文は、動画から学生が読み取るべき核心的な所見を、文章で記述していない。”としているが、このチェック項目を満たさない場合、「設問文から (動画視聴で本来得られるはずの) 情報を削除」するか、「動画自体を削除」して通常の臨床問題とするかという選択肢がある。

現行の A 領域と C 領域のチェックリストは、これらを満たさなかった場合に、設問文・選択肢の修正を行う必要があるのか、動画の修正の修正を行う必要があるのかが混在している。ACREF を「問題としてに完成度を判定するツール」としてだけでなく、問題作成支援ツールとして、「誰が (作題者か IDer か動画編集者か)、何を (設問文・選択肢・動画) を修正するか」を明確にするために改善する余地があるという指摘であった。

## 第6章 考察

### 6.1 暗黙知の形式知化としての ACREF の有効性

動画・音声付臨床問題作成の経験が豊富な筆者の経験に基づく暗黙知を、共同化・表出化のプロセスを経て形式知化した成果物が ACREF チェックリストである。本研究の主目的である「熟練作題者の暗黙知を言語化・構造化（形式知化）」が達成されたかについて、形成的評価の結果から考察する。

2名の領域専門家（SME）に対するインタビューでは、ACREF に対し「ピットフォールが要約されている」「道しるべになる」との発言が得られた。これは、作題経験の多寡にかかわらず、筆者の暗黙知が ACREF を介して他者に伝達され、「連結化」したことを示唆している。特に、これまで「分かりにくい」といった感覚的な指摘に終始しがちであった作題指導において、ACREF は具体的な項目に基づく客観的なフィードバックを可能にした。

したがって、ACREF は動画付臨床問題作成における専門家間の「共通言語」として機能し、教員間の相互評価の精度や効率を質的に向上させるツールであると結論づけられる。

### 6.2 動画によって生じる専門家・非専門家の齟齬とタグによる回避

2名の領域専門家によるセルフ・ピアチェックと、動画付臨床問題のエキスパートである筆者のチェックで齟齬が生じた。その要因として、動画というリッチメディアが持つ「情報の多義性」に対する専門家と非専門家の情報処理の深さ（医学のアートの領域）の相違が明らかになった。この齟齬に対する ID（インストラクショナル・デザイン）的解決策として、動画単体（Raw Data）を評価対象とするのではなく、その動画が意図した臨床的情報を「タグ（メタデータ）」として付与し、抽象化して扱う手法が妥当である。

タグは、動画が本来持つ多様な情報のうち、問題に影響しない情報（ノイズ）を抑制し、問題を作成する上で重要なコアの情報に集約・抽象化する機能を持つ。

再現動画など、典型的な臨床像と一致しない場合でも、動画の細部から切り離して、教育的意図に基づいた動画付臨床問題の作成に注力することが可能となる。

また評価対象を動画のもつ教育的なコア部分をタグとすることで、作題された問題の不備

が、動画そのものにあるのか（タグの情報と動画が一致しない）、動画以外の設問・選択肢によるものか（タグの情報が設問文に含まれてしまっている、など）を判別することに有用である。

### 6.3 評価基準のアンカリング

3者間の齟齬のもう一つの要因として、評価基準（アンカリング）の不在が確認された。特に SME-I の事例からは、作題者が「専門医レベル」で厳格に評価する一方、他者は「医学生レベル」で許容するという基準のズレが浮き彫りになった。したがって、ACREF の運用にあたっては、チェックリスト使用前に「誰に向けた試験か」という前提情報を共有するプロセスが不可欠である。動画付臨床問題が、臨床実習における実践的能力の涵養と評価を目的としていることを踏まえれば、原則として「臨床実習後の医学生」を基準（アンカー）として明記すべきである。

また、ID エキスパートからは、出題疾患の妥当性を担保するために、医師国家試験出題基準に準拠するといった客観的な指標を併用する提案があった。これにより、「稀すぎる疾患」や「不適切な難問」を排除し、評価の信頼性を高めることが可能となる。

### 6.4 チェックリストの改訂ポイントと運用

チェックリストの改訂は、SME へのインタビューやエキスパートレビューからの指摘に対応して、以下の点に留意して今後取り組む。また ACREF の3つの評価領域（Group A, C, R）は、筆者が動画付臨床問題を作成する上で得た暗黙知を形式化したものであるため、複数人での動画付臨床問題の作成のフローに適用する上で、不備の分類はできているものの、どのステークホルダーが次の行動の責任を担っているのかが明確ではない点が、今後の改訂点として挙げられた。

改訂点のポイントを挙げ、研究の今後の課題として示す。

ACREF を用いた運用モデルとして、作題後の事後チェックだけでなく、作題プロセスにおけるトラブルシューティング・ツールとしての活用を期待する。

1. **評価基準(アンカリング)の明記とタグ:** チェックリスト冒頭に、対象読者(Target Audience)および動画が示すコア情報(タグ情報)を記入する欄を設け、評価の前提を固定した。

2. **項目記法の構造化:** 拡張性と視認性を高めるため、項目番号を「AE-01(領域-レベル-連番)」の形式に変更し、優先順位(Essential)を即座に判別可能とした。
3. **Refinement 領域の Essential レベル追加:** ID エキスパートの指摘に基づき、学習体験の質(User Experience)を担保するため、既存の R-3F を Essential レベルに追加する。
4. **アクションプランに基づく役割分担:** 不備が発見された際、誰が修正責任を担うのか(作題 SME、評価 SME、IDer、クリエイター)が明確になるよう再構成する。

特に、作題 SME は臨床的シチュエーションの設定とタグ付けを担い、評価 SME は他領域の視点から学生レベルとしての妥当性(外的妥当性)を担保するという役割の分担を明確化することで、組織的な作題体制の構築が期待される。

## 第7章 結論

### 7.1 総括

本研究では、医学教育における動画付臨床問題の質的向上と作題支援を目的として、熟練作題者の暗黙知を形式知化した評価指標「ACREF チェックリスト」を開発した。形成的評価の結果、ACREF は教員間の「共通言語」として機能し、具体的な改善案を提示する足場かけとしての有用性が確認された。

一方で、専門家特有の視点や、動画に含まれる情報多義性が作題の制約となるという問題点も明らかになった。これに対し、動画をタグによって抽象化する手法や、医学生レベルへのアンカリングを運用プロセスに組み込むことで、この齟齬を回避できる可能性を示した。本研究で提案した ACREF およびその運用モデルは、個人の経験則に依存していた動画付臨床問題作成プロセスを、組織全体に広めるとともに、質の高い問題作成を促進する一助となるものである。

### 7.2 今後の展望

本研究の限界として、形成的評価を行った SME が2名であり、かつ特定の診療科(内科系)に限定されていた点が挙げられる。外科系やマイナー科など、視覚情報の質が異なる領域においても ACREF が同様に機能するかは、今後の検証課題である。

また、本研究では作題者(教員)の視点に焦点を当てたが、実際に作成された問題が学習者(医学生)の正答率や学習効果にどのような影響を与えるかについては検証されていない。今後は、対象診療科を拡大した追加調査を行うとともに、ACREF を LMS (学習管理システム) に実装し、タグ情報と連携した「作題支援システム」へと発展させることが展望される。さらには、蓄積された「動画+タグ+良問」のデータセットを活用し、生成 AI による動画問題の自動生成や自動評価アシストへの応用も期待される。

## 参考文献および付録

### 参考文献

- [1] 「第119回医師国家試験の合格発表について」, 2025. [Online]. 入手先:  
<https://www.mhlw.go.jp/general/sikaku/successlist/2025/siken01/about.html>. [参照:  
02-1月-2026].
- [2] 厚生労働省第31回地域医療構想に関するワーキンググループ, 「良質かつ適切な医療を効率的に提供する体制の確保を推進するための医療法等の一部を改正する法律案の閣議決定について」. 12-2月-2021.
- [3] 医療系大学間共用試験実施評価機構公益社団法人, 「共用試験ガイドブック第23版(令和7年)\_250905」, 7月-2025. [Online]. 入手先: <https://www.cato.or.jp/e-book/23/index.html>. [参照: 02-1月-2026].
- [4] 河北博文, 「厚生労働科学研究補助金 政策科学総合研究事業(臨床研究等 ICT 基盤構築・人工知能実装研究事業) ICT を利用した医学教育コンテンツの開発と活用に向けた研究 令和4年度 総括・分担・研究協力者研究報」, 5月-2023. [Online]. 入手先:  
[https://mhlw-grants.niph.go.jp/system/files/download\\_pdf/2022/202203011A.pdf](https://mhlw-grants.niph.go.jp/system/files/download_pdf/2022/202203011A.pdf). [参照: 02-1月-2026].
- [5] 松山泰., 岡崎仁昭と浅田義和, 「3. 動画・音声素材を活用した CBT 開発の経験から」, *Igaku kyoiku*, vol. 53, no. 3, pp. 221-227, 6月 2022.
- [6] 黄世捷, 伊佐早健司, 望月篤., 伊野美幸, 明石嘉浩と信岡祐彦, 「医学部卒前教育における ICT を利用した実践型 Problem-Based Learning の開発と導入」, 聖マリアンナ医大誌, vol. 48, no. 4, pp. 197-210, 4月 2021.
- [7] I. Nonaka と H. Takeuchi, *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York, NY: Oxford University Press, 1995.
- [8] モデル・コア・カリキュラム改訂に関する連絡調整委員会, 「医学教育モデル・コア・カリキュラム 令和4年度版」, *医学教育モデル・コア・カリキュラム(令和4年度改訂版)*, 2022. [Online]. 入手先:  
[https://www.mext.go.jp/content/20240220\\_mxt\\_igaku-000028108\\_01.pdf](https://www.mext.go.jp/content/20240220_mxt_igaku-000028108_01.pdf).
- [9] G. E. Miller, 「The assessment of clinical skills/competence/performance」, *Acad.*

- Med.*, vol. 65, no. 9 Suppl, pp. S63-7, 9月 1990.
- [10]田川まさみと西城卓也, 「医学教育における学習者の評価(1) 総論」, *Igaku kyoiku*, vol. 44, no. 5, pp. 345-357, 7月 2013.
- [11]栗原敏., 「厚生労働行政推進調査事業費 地域医療基盤開発推進研究事業 Post-CC OSCE を国家試験化するにあたっての妥当性・信頼性・公平性等の検証に関する研究」. 5月-2024.
- [12]R. M. Gagne, *Conditions of Learning*, 3 番目 版. London, England: Thomson Learning, 1977.
- [13]M. Nayer, S. Glover Takahashi と P. Hrynchak, 「Twelve tips for developing key-feature questions (KFQ) for effective assessment of clinical reasoning」, *Med. Teach.*, vol. 40, no. 11, pp. 1116-1122, 11月 2018.
- [14]T. M. Haladyna と S. M. Downing, 「A taxonomy of multiple-choice item-writing rules」, *Appl. Meas. Educ.*, vol. 2, no. 1, pp. 37-50, 1月 1989.
- [15]National Board of Medical Examiners, 「Constructing written test questions for the basic and clinical sciences」, National Board of Medical Examiners, Philadelphia, PA, 2016.
- [16]R. E. Mayer と R. Moreno, 「Nine ways to reduce cognitive load in multimedia learning」, *Web-Based Learning*, Emerald Publishing Limited, 2003, pp. 23-44.
- [17]L. Niekrenz と C. Spreckelsen, 「How to design effective educational videos for teaching evidence-based medicine to undergraduate learners - systematic review with complementing qualitative research to develop a practicable guide」, *Med. Educ. Online*, vol. 29, no. 1, p. 2339569, 12月 2024.

## 付録1 動画付き MCQ「ACREF」チェックリスト研究ご協力をお願い

### 動画付きMCQ「ACREF」チェックリスト研究 ご協力をお願い

To: I先生・N先生

#### 🎯 1. 本研究の目的

良質な動画問題（MCQ）を作るための評価ツール「ACREF」を検証し、改善するために、先生方のお力添えをお願いします。

#### 👉 2. ご依頼する作業フロー

作業は大きく分けて4つのステップです。

##### 【Step 1】資料確認 📄

ACREFチェックリストと運用マニュアルに  
目を通してください。

##### 【Step 2】問題作成 📝

提供された動画A・Bを素材に、臨床問題  
(Type A/X2) を作成してください。



##### 【Step 3】自己評価 (Self) 🙋

I先生 → ご自身の問題をACREFで評価  
N先生 → ご自身の問題をACREFで評価



#### 【Step 4】相互評価 (Peer)

(研究者が問題を交換します)

I先生 → N先生の問題をACREFで評価

N先生 → I先生の問題をACREFで評価

##### 【重要】

ステップ4の際、お互いの「自己評価 (Step 3)」の結果は非公開の状態で行います。

### 3. 今後のプロセス

これにより、1つの問題に4つの評価が集まります。

1. **作成者本人** (I先生 / N先生)
2. **ピア** (N先生 / I先生)
3. **研究者** (黄)
4. **医学教育専門家** (Expert)

これらの**一致度・不一致度**を分析し、後日**ヒアリング**をお願いする場合があります。

#### ※ \*\*GOAL\*\*

チェックリストを改善し、今後の動画付き臨床問題の質向上と普及を目指します。

ご協力のほど、何卒よろしくお願い申し上げます。

## 動画付きMCQ作成支援チェックリスト (ACREF) 運用マニュアル

### 1. 目的：なぜこのチェックリストを使うのか？

近年、医師国家試験や医学教育において、動画を用いたMultiple Choice Question (MCQ) の導入が求められています。しかし、動画問題の作成には特有の難しさがあります。

- ！ 「動画がなくても解けてしまう（動画が飾りに過ぎない）」問題
- ！ 「設問文や音声動画の"ネタバレ"をしている」問題
- ！ 「学習者の臨床推論プロセスを評価できていない」問題

このような不適切な問題为避免、**「動画でなければ評価できない、良質な臨床推論問題」**を作成するために、このACREFチェックリストは開発されました。

このマニュアルは、作題者（医学部教員）がご自身の作成した問題を客観的にレビューし、その品質を担保・向上させるためのガイドです。

### 2. ACREFモデルの概要：チェックリストの「読み方」

このチェックリストは、3つのグループ (ACR) と、2つのカテゴリ (E/F) で構成されています。

#### 3つのグループ (ACR)

##### A: Appropriateness (妥当性)

「そもそも、この問題で動画を使うことは妥当か？」を問います。

##### C: Construction (構成・品質)

「教材として論理的・技術的に正しく構成されているか？」を問います。

(例：静止画で十分ではないか？  
設問文だけで解けないか？)

す。(例：動画の品質、設問文の文脈、選択肢の粒度)

### R: Refinement (洗練・工夫)

「学習教材として、より教育効果を高める工夫がされているか？」を問います。  
(例：学習支援、真正性、公正性)

## 2つのカテゴリー (E/F)

これがACREFモデルの最も重要なルールです。

### E E = Essential (必須項目)

これは「最低限の品質担保 (Go / No-Go)」のための項目です。良質な動画問題として絶対に満たすべき必須要件です。

### F F = Further (推奨項目)

これは「品質向上のための改善点」を示す項目です。必須ではありませんが、これらを満たすことで、問題は「成立している」レベルから「教育的に優れている」レベルへと高まります。

## 3. 使用方法：ステップ・バイ・ステップ

このチェックリストは、作成した問題の「自己点検 (セルフレビュー)」に用います。

### ステップ1：問題の作成

Step

まずは「設問文」「動画 (音声)」「選択肢」の3要素からなる問題のドラフトを作成します。

1

### ステップ2：E (Essential) 列のレビュー

Step  
2

この列の項目（A-1E, A-2E, ... C-4E）を一つずつ確認します。  
この列の項目は、すべてチェック [✓] が入る必要があります。

### ステップ3：E項目がチェックできない場合（最重要）

最重  
要

もしE列に一つでもチェックできない項目があった場合、その問題は「動画問題として致命的な欠陥」を含んでいます。必ず「注釈」欄を確認してください。

注釈の「不適切」例に当てはまっていないかを確認し、チェック [✓] が入るまで問題を修正してください。

（例：C-1Eが満たせない場合、注釈（注2）に従い、動画の再編集や差し替えを検討します。）

### ステップ4：F (Further) 列のレビュー

Step  
4

この列がすべてチェックされたら、次にF列の項目を確認し、該当するものにチェック [✓] を入れます。

この列は、空欄 [ ] があっても構いません。F列は、あなたの問題を「より良く」するためのヒント集です。

## 4. 作題のための5つの核となる考え方 (Core Principles)

このチェックリストは、以下の5つの核となる考え方に基づいています。E項目（必須）のレビューで迷った際にご確認ください。

### 原則1：3要素の必須性 (A-1E, C-4E)

原則1

良質な動画問題は、「設問文（文脈）」「動画（所見）」「選択肢」の3つが揃って初めて解答可能になります。どれか1つでも不要な（＝"飾り"になっている）問題は不適切です。

### 原則2："ネタバレ"の排除 (A-3E, A-4E)

原則2

学習者が動画から読み取るべき「所見」や「解釈」を、設問文（文章）や音声（検者の発言）が先に説明してはいけません。

**原則3：「認知」ではなく「解釈」を問う（C-3E, A-5F）**

原則3

不適切な問題は「見たままの所見」や「所見名」を当てさせます（＝認知）。良質な問題は、その所見が「何を意味するか（例：原因病態、鑑別）」を考えさせます（＝解釈・推論）。

**原則4：意図的な「ノイズ」と「支援」（C-5F, R-1F, R-3F）**

原則4

動画内の「支援（矢印など）」や「ノイズ（環境音など）」は、学習者のレベルに応じて意図的に設計されるべきです。ただし、「臨床的に無関係な情報（＝不公正な引っかけ）」は意図的に排除しなければなりません。

**原則5：学習者の視点に立った「真正性」（C-2E, R-2F）**

原則5

問題のシチュエーションは、学習者が「自分ごと」として捉えられる現実的な役割（例：研修医）であるべきです。また、「再現動画」の場合はそのことを明記し、学習者の無用な混乱を避ける必要があります。

## 付録3 動画付き MCQ 作成支援チェックリスト (ACREF)

### 動画付きMCQ作成支援チェックリスト (ACREF)

2025.11/15 237g8817 黄世捷

#### はじめに：評価と提出の手順

##### 提出方法

1. 下のプルダウン3項目を選択し、チェックリスト (E/F) をすべて評価してください。
2. 全ての入力が完了したら、ブラウザの「印刷」機能 (`Ctrl+P` or `Cmd+P`) を開きます。
3. 送信先 (プリンター) を「PDFとして保存」に変更し、PDFファイルとして保存してください。
4. 保存したPDFファイルを、研究者 (黄) にご提出ください。

**注意: このHTMLファイルはブラウザを閉じると入力内容がリセットされます。評価完了後、すぐにPDFとして保存してください。**

##### ファイル名

提出時のファイル名は、以下の命名規則に従ってください。

- 例1 (自己評価) : **Nセルフ1.pdf**  
(N先生が、動画1の問題を自己評価した場合)
- 例2 (ピア評価) : **Iピア2.pdf**  
(I先生が、[N先生の]動画2の問題をピア評価した場合)

##### その他の補足

(ここにその他、必要な追加説明を記載します)

問題作成者:

選択してください ▼

評価者:

選択してください ▼

使用動画:

選択してください ▼

グループ	NO.	チェック項目	E (ESSENTIAL)	F (FURTHER)	注釈
<b>A: Appropriateness</b> (動画使用の妥当性)					
A- 1 E		<b>【動画の必須性】</b> 設問文と選択肢だけでは解答できず、動画(音声)の視聴が必須である。	<input type="checkbox"/>		<b>(不適切)</b> <ul style="list-style-type: none"><li>設問文の情報(例: 病歴、検査結果)だけで選択肢が絞り込める。</li><li>設問文の病歴と選択肢の所見名(例: 「羽ばたき振戦」)を組み合わせるだけで、正解が推測できる。</li></ul>
A- 2 E		<b>【"動き"の必須性】</b> 動画は、「動き」や「時間経過」の確認が必須な所見(静止画では不十分な所見)を含んでいる。	<input type="checkbox"/>		<b>(不適切)</b> 静止画でも伝わる所見(例: 皮疹の色調)。 <b>(適切)</b> 「動き」や「変化」が診断の鍵となる所見(例: 皮疹の圧退色、呼吸に伴う頸静脈の変動)。
A- 3 E		<b>【"音声"の適切性】</b> (※音声所見が重要な場合) 動画は、解答に必須な"音	<input type="checkbox"/>		<b>(不適切)</b> <ul style="list-style-type: none"><li>解答に必要な情報(例: 指示音声、聴診音)がミュートされている。</li><li>音声(検者)が「腫れています」「熱感がありま</li></ul>


グループ	NO.	チェック項目	E (ESSENTIAL)	F (FURTHER)	注釈
		<p>声" (聴診音や動作指示など) を含んでおり、かつ、学生が解釈すべき所見や解釈そのものを音声で説明していない (例: 「腫れています」など)</p>			<p>す」「脈が弱いです」のように、学生が解釈すべき所見や解釈そのものを説明している。</p>
	A- 4 E	<p><b>【所見の非開示】</b> 設問文は、動画から学生が読み取るべき核心的な所見を、文章で記述していない。</p>	□		<p><b>(不適切)</b> 設問文に「...両手を背屈させると、羽ばたくような振戦が見られた」と、動画で評価すべき所見が文章で記述されている。</p>

グループ	NO.	チェック項目	E (ESSENTIAL)	F (FURTHER)	注釈
	A- 5 F	<p><b>【臨床推論の促進】</b></p> <p>設問（動画・設問文・選択肢の全体）は、単一所見の認知（所見名当て）でなく、複数の所見を統合し、病態生理まで考慮する臨床推論を要求している。</p>		<input type="checkbox"/>	<p><b>(不適切)</b> 「片側下肢浮腫」の動画で「下肢エコー」を1つ選ばせる。(所見の認知のみで回答)</p> <p><b>(適切)</b> 「片側下肢浮腫＋頸静脈怒張」の動画で「下肢エコー」と「心エコー」の2つを選ばせる。(所見の認知と病態の解釈)</p>
	A- 6 F	<p><b>【学習体験の重要性】</b></p> <p>頻度は低くとも臨床的に重要な（例：重篤な疾患、見落としやすい所見）事象を扱うことで、学生の知識と実践を結びつける学習体験を提供している。</p>		<input type="checkbox"/>	<p><b>(不適切)</b> 臨床的重要性の低い動画。</p> <p><b>(適切)</b> 知識としては知っていても臨床で未経験な重要所見を示す動画（羽ばたき振戦など）。</p>
<b>C: Construction</b> (教材の構成・品質)					
	C- 1 E	<p><b>【品質の担保】</b></p>		<input type="checkbox"/>	<p><b>(不適切)</b> 臨床判断に必要な最低限の品質を満</p>

グループ	NO.	チェック項目	E (ESSENTIAL)	F (FURTHER)	注釈
		動画（音声）は、所見を識別し、臨床的な判断を行うために十分な品質である。			<p>たささない</p> <ul style="list-style-type: none"> <li>画質: 解像度や照度が低い、ピンボケしている。</li> <li>画角: ズームしすぎ（部位不明瞭）またはパンしすぎ（所見不明瞭）。</li> <li>カメラワーク: 手ブレが酷い、所見の途中でカメラが動く。</li> </ul> <p>（注：満たさない場合、出題動画の見直しや、作題補助者への動画の再編集を依頼）</p>
	C- 2 E	<p><b>【文脈の明示】</b></p> <p>設問文は、動画を解釈する前提となる文脈（例：場面設定）を明確に示している。また、動画が再現（シミュレーション）である場合は、そのことを明記している。</p>	<input type="checkbox"/>		<p><b>（適切）</b></p> <p>「...の再現動画を示す」と明記し、所見の矛盾・不一致を回避する。（筋力低下があるのに萎縮がない、など）</p> <p>（補足）文脈（例：患者背景）を記述する際、動画で見るべき所見（例：「羽ばたくような振戦」）を文章で記述しないよう注意する。（A-4Eの不適切に相当）</p>
	C- 3 E	<p><b>【選択肢の適切性】</b></p>	<input type="checkbox"/>		<p><b>（不適切）</b></p> <p>眼球運動の動画を示した上で、選択肢</p>

グループ	NO.	チェック項目	E (ESSENTIAL)	F (FURTHER)	注釈
		<p>選択肢は、「粒度」(例：身体所見の種類、検査の種類)が揃っている。また、所見を問う場合、単なる「見たままの記述」ではなく、「鶏歩」のような医学的概念で構成されている。</p>			<p>が a. 眼振, b. 瞳孔不同 <b>(適切)</b> 眼球運動の動画を示した上で、選択肢が a. 垂直性眼振, b. 水平性眼振</p> <hr/> <p><b>(不適切)</b> 所見がそのまま選択肢になっている。「つま先から歩いている」 <b>(適切)</b> 所見の解釈が選択肢になっている。「鶏歩」</p>
	C- 4 E	<p><b>【設問文の必須性】</b> 動画と選択肢だけでは解答できず、設問文の情報(例：患者背景、具体的な問い)が解答に必須である。</p>	<input type="checkbox"/>		<p><b>(不適切)</b> 設問文の文脈(例：患者背景)が不足しているため、動画所見(例：失調)と選択肢(例：脳梗塞)だけで解答できてしまう。</p>
	C- 5 F	<p><b>【"実臨床"の再現性】</b> 動画(音声)は、臨床的に妥当な範囲で、実臨床に近い文脈</p>		<input type="checkbox"/>	<p><b>(適切)</b> 心雑音の評価時に、鑑別に必要なJaneway斑を否定するための爪の視診(陰性所見)を含めること。</p>

グループ	NO.	チェック項目	E (ESSENTIAL)	F (FURTHER)	注釈
		(例：環境音、診断上除外すべき陰性所見)を意図的に含んでいる。			
<b>R: Refinement</b> (学習支援の工夫・洗練)					
	R- 1 F	<b>【学習支援の工夫】</b> 動画は、学習効果を高めるための意図的な編集（スロー再生、矢印、健側との比較、ズームなど）が加えられている。		<input type="checkbox"/>	常に必須ではないが、学習者レベル：（臨床実習前か、後かなど）にあわせた調整。 <ul style="list-style-type: none"> <li>スロー再生：持続時間が短いあるいは速い所見に対して。</li> <li>矢印：複数の所見が同時に存在する場合、問いたい所見を明確化するため。</li> <li>健側との比較：筋萎縮や腱反射など、個人差が大きい所見の「左右差」を明示するため。</li> </ul>
	R- 2 F	<b>【真正性（リアリティ）】</b> 提示されたシチュエーション（特に学習者が担う役割・立場）		<input type="checkbox"/>	<b>(不適切)</b> 学習者の経験から遠い、または過度に専門的な役割 (例：「あなたは指導医です」「あなたは腫瘍内科医です」)。 <b>(適切)</b> 学習者が近い将来

グループ	NO.	チェック項目	E (ESSENTIAL)	F (FURTHER)	注釈
		は、学習者が「自分ごと」として捉えられ、かつ「実臨床であり得る」ものである。			に担う一般的・緊急的な役割（例：「あなたは研修医です」「コードブルーに最初に到着した」）。
	R- 3 F	<b>【設問の公正性】</b> 設問文および動画は、本質的な評価とは無関係な「引っかけ」要素（＝臨床的に非妥当な情報）を意図的に排除している。		<input type="checkbox"/>	<b>（不適切）</b> ・心音聴診の動画で、鑑別と無関係な四肢MMTの動画を意図的に含めること。
<p> <b>コメント欄（自由記載）</b></p> <p>評価・入力が困難だった項目、その他、チェックリスト全体に関するご意見をご自由にお書きください。</p> <div data-bbox="391 1422 1284 1624" style="border: 1px solid #ccc; height: 90px; margin-top: 10px;"></div>					

## 付録4 SME-I 対するインタビュー文字起こし

黄

インタビューをさせていただきます。よろしくお願いたします。今日はお忙しい中ありがとうございます。

今日はあくまで、作ってもらった問題がいか悪いかではなくて、このチェックリストがどういう風に、役立ったのかということを中心に聞いていこうと思います。

目的としては、私が動画付臨床問題の作成エキスパートなんですけど、エキスパートの中にある”いい問題”が”悪い問題”かをチェックリストの形で言語化して、その言語化したチェックリストが、I先生のような動画付きの問題を作ったことのある経験者と、経験のないもうN先生が動画付臨床問題を作った時に、どう差が出るかということも含めて、有効性の検討としました。

“使いやすかったどうか”ということではなくて、使ってみて正直どうだったか、このチェックリストをブラッシュアップするためなので、使いにくかったかどうか、くかったという点くらいで、むしろよく、今までに作った時に陥りがちなピットフォールを、こういう風に説明すればわかるんだなっていうような、というリストにすごくなっていたと思います。

その通りいっていただきたいなと思います。

あとは、このチェックリストでチェックをつけたところ、つけなかったところを踏まえ、どうして。そういう風に考えたのかということ伺っていかうかなと思います。

まずは率直にまずチェックリストを使ってみてどうでしたでしょうか。

SME-I

チェックリスト自体は、やっぱり問題作成にあたって、必ず引っかけりそうなポイントがすごく要点が。捉えられていて、これを見ることによって、基本的には作ったことがない人にでも。見直すポイントっていうのが、ものすごく要約されていて、非常に良いと思いました。

チェックリスト自体は、その項目が”満たしている”でチェックなのか、”満たしていない”でチェックなのがちょっと分かりに

黄  
ありがとうございます。例えば、いわゆるポジティブチェックで“つく方が必ずいい”ように作り込んでいました。例えば、インフルエンザの問診票みたいに、問診内容に対する患者さんの回答が”いいえ”の

ほうが、ワクチン接種して良い”という設問-回答関係があるじゃないですか。それだと回答する人にとって違和感が変なので、チェックは、“ついている方が良い”というふうに、ポジティブな意味合いになるようにしたんですけど、具体的にどの項目が”チェックが付く方が良いのか、悪いのか”迷った項目は覚えていますかね。

このチェック項目が、“満たさない”、“つけない”方が”良い”のかどうか悩むところがあった、意味を取りづらい点があったのだと思いますがいかがでしょうか。

これが 1 枚目が先生のセルフチェック結果です。C-4E にチェックが付いていません。”動画と選択肢だけでは回答できず設問文の情報が回答に必須である”。というチェックですが、これは”満たす”方が”良い”、ということが分かりづらいでしょうか。

SME-I

今読んでみると、特にわかりにくいところはないかもしれませんが。C-4E にチェックが付いていないのは別の理由です。

黄

他に R-1F、R-2F、R-3F にもチェックが付いていないです。

R-3F「設問文および動画は、本質的な評価とは無関係な「引っかけ」要素（＝臨床的

に非妥当な情報）を意図的に排除している」。この”排除”という表現がわかりにくいのでしょうか。

SME-I

今ちゃんと読むとわかります。わかりそうな気がしました、この項目だったと思います。

黄

この”排除している”という言葉が、ネガティブな印象があるのかもしれない。

これは、例えば、心音を聴いているのに、徒手筋力テストをする動画をいれたり、意味のないことを入れる、つまり、ただのノイズになるようなものを入れていないか、ちゃんと”そぎ落としていく”という意味合いでこのことばを選んでいました。

普通の 一連の診察の動画を、そのまま出すのではなく、“いらぬものはいらぬって削るべきじゃないか”という意味でこのように表現しました。

黄

研究の手法として本来は、最初に動画付臨床問題を作ってもらった後、このチェックリストを提示して、チェックリストをもとに、自身でもう 1 回、問題をブラッシュアップ自分でしてもらうことを考えていました。

時間的な制約や、問題作成の経験のない教員の負担を考慮して、今回は素材と同時にチェックリストを提示しましたが、実際はチェックリストを見ながら作った、作ってから見たのかというどちらでしょうか。

SME-I

問題を作ってからチェックリストを見ました。

黄

チェックリストをみないで作ってから、チェックリストを評価した。その後、満たしていない点については再度直してみましたか？

SME-I

作り直していません。

黄

わかりました。では、チェックリストのチェックがつかなかった時に、「あ、ここが足りなかったな」というのに納得いくのか、どうかは、いかがでしょうか？

SME-I

納得いくのではないのでしょうか。

ただ、A-1E「【動画の必須性】設問文と選択肢だけでは解答できず、動画（音声）の視聴が必須である。」はチェックを最終的に

に付けましたが、自分では悩みました。

黄

I先生の作った問題1だと設問文に、「手が震えていることに気づいた」とあるので、この「振戦の動画」を見なくても、この問題は解けるのではないか、ということでしょうか。

一応この設問文には「振戦があった」とは書かれていないから、動画を見る価値はあると私は思って、動画付臨床問題作成のエキスパートとして、A-1Eは満たしていると考えましたが、どうでしょうか。この情報は、患者の主訴なので削除するのが難しいと思います。これは「手の震え方の違い」を動画で判断する問題なので、妥当なのではないか。ピルローリングのような細かい震えなのか、粗大な震えなのかは動画を観ないと解けないと判断しました。

SME-I

なるほど。でもそれは「知識」の問題だと思います。この設問文にある「コップを持っているときに手が震えている」時点で絶対に動作時振戦であると判断ができます。この病歴が書いてある時点で動画を観なくても選択することができます。

黄

私はこの「コップを持っているときに手が

震えている」という設問だけでそのように考えることができませんでした。それは「エキスパート（領域専門家、SME）」だからです。おそらく、N先生も同じように、そこまでは考えていないのではないのでしょうか。

だから、問題を作った本人だけはA-1Eを満たしていないと思って、ピアレビューのN先生と、動画付臨床問題の作成のエキスパート（SMEではない）である私が満たしているというギャップが生まれてきたのです。非常に参考になるコメントです。この設問を見ただけで、脳神経内科医（SME）であるI先生は動画を観なくても分かるんだけど、僕らは動画を見ないとそこは断別つかないから、専門家だからこそチェックをつけられないってポイントがあったってことかなるほどなるほど面白いな。

SME-I

そうですね、だからちょっと専門領域じゃない方がいいのかもしれない。

黄

今回の動画は、意図的にひとつはI先生の専門領域として動画2「麻痺」を提示しました。しかし、実際は動画1も「振戦」なので、広い意味だと先生の専門領域でしたね。

SME-I

動画2「麻痺」に関しても、ピアレビューしたN先生の問題のように、急性麻痺として問題を作成しようとしてみましたが、あれは絶対急性麻痺じゃないですね。

それって知識があるから、動画を観て余計に考えちゃうから、この問題は適切じゃないってなるのです。少ない問題の情報で代替のことが分かっちゃうのです。

問題の作成者が、専門家であることは、良点でもあるかもしれないけど、つまり問題の”整合性”や”確からしさ”を高める点では、すごくいいと思うのですが、専門家であるがゆえに、本当に少ない情報だけで、どんどんわかるようになっていってしまいます。もしくは情報をぼかすとかしないといけないかもしれません。

黄

I先生の作った動画2の問題は、ACREFのチェック結果を診ると、N先生と私からは、正直、あまり良くないという意見になっています。（自由記載も含め）

しかし、I先生はむしろ、動画2に関してN先生が作った問題は、「医学的に間違っている」と感じていますか。

SME-I

間違っていると思います。急性期の脳卒中で「痙性の麻痺」ができてるのが間違

っています。「弛緩性」でなければいけません。あの動画では「痙攣」や「Tod 麻痺」が疑われます。

だから、ああいう問題しか作れないのです。

黄

なるほどなるほどそういうふうになるのか。

SME-I

最初、N 先生の問題のように作ろうとしましたが、動画を観ると、それでは整合性がつかなくなってしまうのでそのようになりました。

黄

I 先生の動画付臨床問題作成経験の割りに、作った問題に対して N 先生や黄があまりよいチェックをつけられない問題を作った理由としては、その”動画の不適切さ”があまりにも気になって、動画に一致した問題を作れないということがあったということですね。

SME-I

動画に順じて問題を作る必要があると思いました。

黄

集計表をご覧ください。

集計表ではセルフで認めたチェックリス

トは 1、ピアが認めたものは 2、動画付臨床問題作成のエキスパートである黄が認めたものは 4 としてカウントしました。

そうすると各項目において、7 は、セルフ・ピア・エキスパートの全員が一致して認めているという意味です。1 は作題者のみ満たしたと認めている、2 はピアだけ認めている状況で、3 は作題者本人とピアは認めているが、エキスパートの私がダメと言っている。

6 は本人以外、ピアとエキスパートだけが認めているという風に解釈します。

I 先生が作った 2 問目の A-1E A-2E A-3E A-4E はいずれも 7 となり一致しています。C-1E、C-3E も 7 で一致しています。

一方で、R の項目は 7 がないことがわかります。

各項目の結果が「7」ではなく「0」であっても、全員の意見が一致しているといえる。今回の結果で、I 先生の作った問題 1 の C-4E は 6 で、ピアとエキスパートが良いと認めているのに、作題者本人だけが、「これは満たさない」と思っているのがなぜ起きているのかが最初分からなかったが、SME であることが起因していることが分かった。

このように、作題者本人がチェックリストを満たさない、と考える一方で、ピアとエ

キスパートが認めた、つまり 6 の評価になった問題を、「いい問題」として採用するのか、どうすべきなのでしょう。このチェックリストが役立つかどうかという原点にも関わってきます。このチェックリストが「何に役立つのか」という問いです。

#### SME-I

役立つとは思いますが。なぜかという、同じ条件付けとして、それぞれの項目がどうなっているかって、評価項目が明確化しているわけですから。

黄

今回、動画 1 の A-1E のセルフチェックを最終的に満たしたといういみで 1 がつき、全員で一致した 7 になりましたが、もし悩んだとおりに、「満たさなかった」0 にすると、状況が変わっていたと思います。

全員一致で 7 だった場合、周囲から何も突っ込まれないんですけど、もし 0 だったら、「え、なんで、これが満たさない、0なのっていう」質問になって、作題者の言語化につながると思うので、その点では面白かったです。

黄

逆に、エキスパートの黄だけが満たしていないと判断した、つまり 3 になった項目

として、

#### C-3E

【選択肢の適切性】 選択肢は、「粒度」(例：身体所見の種類、検査の種類)が揃っている。また、所見を問う場合、単なる「見たままの記述」ではなく、「鶏歩」のような医学的概念で構成されている。

がある。

動画 2 の選択肢として

- a. 「ご出身はどちらですか？」
- b. 「動悸がすることはありますか？」
- c. 「歩きづらいことはありますか？」
- d. 「著しい頭痛が出たことはありますか？」
- e. 「ご家族には同じ症状の方はいらっしゃいますか？」

とあるが、症状でまとまっているところの中で、ひとつだけ出身地が問われていて、粒度が揃っていないと感じました。

問題の質よりも、「国家試験における問題作問上のルール」に則った指摘だと思うが、その指摘には納得がいくのか？

#### SME-I

そう言われると、粒度が揃っていないという指摘には納得できます。

黄

I先生の問題1の

C-4E

【設問文の必須性】動画と選択肢だけでは解答できず、設問文の情報（例：患者背景、具体的な問い）が解答に必須である。

に関して、ピアのN先生と、黄は良い問題として、この項目を満たしていると評価していますが、同じ理由で、I先生は満たしていないと答えているのですね。

SME-I

A-1Eで悩んだのと同じ理由です。

黄

この辺のギャップがあるから、6（ピアとエキスパートのみ認めている）というものが発生するのですね。

動画1「振戦」の

R-1F

【学習支援の工夫】動画は、学習効果を高めるための意図的な編集（スロー再生、矢印、健側との比較、ズームなど）が加えられている。

を、僕とN先生は、実臨床に則していて、意図的に診断しやすいように出来ていると評価していますが、I先生は、これはあんまりいい診断に使える動画じゃないなと判断されています。

SME-I

この評価は、「意図的な編集をしていない」という点で、満たしていないと判断しました。

黄

なるほど。意図的な編集を加えていなくても、診断に役立つと判断したN先生と黄はチェックを付けているが、編集をしていないからそもそもダメだと捉えられたのですね。このチェックリストは「編集をしていない場合にどう評価するのかがわからない、わかりにくい」という課題ですね。

R-2F

【真正性（リアリティ）】提示されたシチュエーション（特に学習者が担う役割・立場）は、学習者が「自分ごと」として捉えられ、かつ「実臨床であり得る」ものである。

の合計は2でしたが、なぜ御自身では満たしていないと考えましたか？（黄も満たしていないと判断している）

SME-I

リアリティに関する項目ですね。

「自分事として捉えられる」ということが、どういう要素なのかが分かりませんでした。この症例において「自分事として捉え

られるようにする」という意味が分かりませんでした。

黄

了解です。この項目は Further、推奨なので、満たしていなくても動画付臨床問題としては妥当な範囲です。意図としては、すごい珍しい病気のすごい珍しいシチュエーションのような、そんなのめったに見ないよって問題ではないかです。

問題として取り扱うのに妥当な、要はすぐまれなところを見ていないかを評価しています。

SME-I

コモンなものを診ているか、という意味ですね。

黄

そこが満たされると、加点として考えています。そもそも R は一つも満たさなくても問題としてはなり立つ項目として設定しています。

例えば、「あなたは初期研修ですと、5年前にかかっていた患者の〇〇が」「あなたが診察したところ」みたいな設定が良いと考えています。言い換えると、アンカリングができていることを評価したいと考えており、I 先生の動画 1 の問題をこれを満たさないと判断しました。

一方、ピアの N 先生は、この R-2F は満たしていると回答しているのでまた意見を聞きたいと思います。

R-3F については 5、作題者とエキスパートの私は、満たしていないと判断していました。あわせて N 先生がここで満たしているとした理由をまたお伺いしようと思います。

動画 2「麻痺」に関する問題に移ります。N 先生も私も、I 先生の動画 2 の問題は、そもそも病歴背景を含めない、国家試験で言うところの「一般問題」の型式になっていると判断しました。そのため設問にある病歴は全くなくても解ける内容だと評価しています。

その点については、先ほど述べていたような、動画 2 の演技が妥当でないから、という意味であっていますでしょうか？

SME-I

うーん。演技の問題というか。

当初は、もうほんとに急性期の麻痺として診察しましたよ、という流れで問題を作ろうと思いましたが、動画を観ると、やっぱりどうしても、これは慢性の麻痺か、Tod 麻痺になってしまうので、それで作ろうとなりました。

Tod 麻痺の問題って、かなりマニアックで、

臨床はすごい重要ですが、ちょっとマニアックすぎるなと思って、それをテーマにした問題を作る選択肢がなくなりました。そうすると診察の症候しかないかなって言って、かなりこう最初のその動画自体のせいで、診察のこと問うのか、診断を問うのかみたいに悩みました。

動画は診察動画だから、診察の問題か病態をきく問題かを考えたのですが、動画のせいで、かなり病態や診断が狭められているので、結局診察所見のところでは問題を作れないと考えて、苦し紛れにこうなったのです。

黄

専門家からすると、これは再現している動画が下手だという感じですか。（演技の不備か）

例えば私の専門領域である循環器領域で言うと、収縮期の心雑音で、専門外の先生に問題を作題してもらおうと、大動脈弁狭窄症か僧帽弁閉鎖不全症で問題を作成するだろうけど、SMEである私が音声を聴くと、これは大動脈弁狭窄症ではなく、僧帽弁閉鎖不全症の音だと区別できてしまい、大動脈弁狭窄症として問題を作成できない状況に近いのか。

分かっているからこそ、問題の作成に「広がり」を持たせられないという感じに近いと理解しました。

本来は、動画と問題を作る、つまり動画を提供する人と問題を作る人は可能な限り同一人物という前提で作ってるけど、いずれは動画だけどこかにストックしておいて、このストックされた動画で新たに問題を作る、という運用が見込まれます。

そのため今回はだから、動画をあらかじめ指定、提示して問題作成を依頼しましたが、提供された動画によっては、専門家だからこそ問題を作れないというということが起きるかもしれません。

SME-I

動画が何種類かあって、ストックの中から選ぶだったらいいと思うのですが、指定された動画で問題を作れ、という制限があると、動画を観た時点で、問題として問うことができる病態や治療がぐっと狭まってしまう。

その制約の上で、かつ問題を解く学生のレベルにおいて、絶対必要な知識という制約も出てくるため、動画 2 のような苦し紛れの問題ができたのです。

黄

N 先生の動画 2「麻痺」の問題は、所見を個別に評価せず、「片麻痺」として捉えさせるのか、うまいなと思って非常に高く評価しました。

しかし、脳神経内科の I 先生からみると、あの問題は間違っている、となりますか？

SME-I

そうですね。あのようにはなりません。

臨床に則して考えると、脳卒中ではあり得ません。

黄

今回は再現動画でした。もし臨床的に正しい動画を I 先生に作って下さい、といわれたら、再現可能でしょうか。

SME-I

できるのではないのでしょうか。弛緩性麻痺ですから。

グイッと力が入っていましたので、慢性期になるとあのような所見になります。

もしくはてんかんとかで、過剰交奮している状態。陽性徴候として、麻痺が出ているのでしょうか。

もし無理矢理作るとしたら、もともと麻痺がある人が急に歩けなくなりました、何故でしょうか、感染とか脱水のせいでもともとある麻痺が余計広くなったという設定にすればできます。そうでないと痙性麻痺ということは基本的にはありません。

黄

ちなみに、循環器内科医の僕が聞いていて

も、なるほど納得できないレベルだと感じています。

学生には当然高いレベルと思うが、学生にそのレベルは求めていますか？

SME-I

求めています。

黄

このチェックリストが役に立っていないのか、どうかという視点で言うと、そもそも動画が適切じゃないというところに課題があると、チェックリストが適切に運用出来ない可能性が考えられます。

その対策として、「動画が問題作成に適切ではない」という項目も含めて、チェックリストに入れるべきなのでしょうか。今回は 領域専門家だから自然と良い問題を作れるわけではないところが少し見えてきました。

エキスパートだからこそ、こだわってしまうと現実には、その動画が妥当じゃないってなって、問題を作れないなとことが起こります。

今回私は問題を作成していませんが、N 先生が脳神経の専門家じゃないからこそ、あ、じゃあこれでいいでしょう、みたいな感じで問題を作れたのかなと、今想像しています。

このことは、動画付臨床問題に限らず、卒業試験などの問題を、各専門領域の教員が作ると、あまりにもこだわりすぎて、専門領域以外の医師でも、全然解けないような問題ができあがるのと似ています。

動画付臨床問題に限らず、専門領域の医師が問題を作成するうえでの課題として同じ設問文でも、「僕はこれでは動画見えて答えられない」と思ってるけど、専門家は「これを見たら明らかだよ」というふうなことがおきます。

同じ文章中で得られる情報の「受け止め度」が違います。

#### SME-I

動画を用いることの利点は、「見たら分かる」点です。しかし、文字として理解している学生と、「専門家が動画を見て分かった」というレベルに差があります。

動画にすることで学生と専門家の能力の差が、さらに余計開いちゃうと思います。

黄

もうちょっと詳しく教えてください。

#### SME-I

文字で書いてある場合、その文面だけで捉えるでしょう。でも動画だった場合には、そこから収集するデータ量が知識が多ければ多いほど増えますよね。

結果、今みたいな、専門家がこだわってるせいで、わけわからなくなるということになります。

黄

僕はあくまで、学生が動画を見て麻痺だっということがわかればいいやと思っていました。僕が見ても「麻痺だね。右手と右足だね」という程度の捉えていました。

しかし専門家はそれよりも高次元な情報まで受け取ってしまいます。

この専門家レベルの情報にまで合わせた問題を作ろうと思うと、苦しくなっちゃうっていうということですね。

#### SME-I

なるほど、片麻痺だとわかればいい、とするのですね。片麻痺だから、病変はどこかというレベルで良いのでしょうか。

黄

専門家が動画に対して高次元すぎるレベルで情報を受け止めないためには、「これは再現動画」であるってことをちゃんと明示しなきゃいけないかなと思いました。

あとはN先生のようにAIで作成したCTを使って見せた時に、正常ではありえないものが生成された画像内にあることを個人的には懸念しています。

AI で作成された動画・画像で診断できる

ということをゴールにして良いのでしょうか。

例えば、皮疹を勉強するときに、これは AI で作って皮疹だと提示した上で、専門家がそれらしいと言っても、実際の症例とは異なる点があるのかが懸念点です。

ただ、現在聴診の学習で用いている”聴くぞう”や音源の”iPax”は全部再現音声です。あれは、収録した音そのものではないが、こういった方法でしか提示できないものがあることも認識しています。自分でもどうあるべきかがまとまっていません。

#### SME-I

そう言われても思ったことは、動画はあくまで超臨床(臨床現場の中でも極限的な状況)の現場で診断ができることではなく、片麻痺だと診断できれば良いのだと思います。

片麻痺があるということは、脳疾患だよねと言う知識があれば良い、その程度でいいのだと思いました。ある意味動画はデフォルメしてても全然いいのではないのでしょうか。

黄

動画を用いることで、文章右手と右足の MMT が 3 でありうという情報よりも、「まし」な情報提供ができていていると思います。

じゃあ動画があったら、専門家ほどいい問題を作るわけじゃないってことは、なんとなく今見えてきて、それをこの。チェックリストがカバーする内容なのかどうかっていうのはあえて言うと、その過度に専門的すぎないみたいな項目を考えるべきかもしれません。

#### SME-I

学生にとって必須とする知識で解けるかということですね。

黄

僕は今、河北班という研究班で、動画付臨床問題を含む教材を作っています。

教材内に動画を含めているが動画は、動画は単なる参考資料ではなく、問題として提示しています。その方が、学生の学習行動を変化させるのではないかと考えています。

動画を使った教科書を作って学生に見せても、なるほどとは思いますが、そこからさらに学ぼうという気持ちにならないのではないかと感じています。

次に、N 先生の作った問題を今見てもらおうと思います。

N 先生の動画 1 の問題は、N 先生自身も I 先生も A-1E は満たしていないと回答しています。N 先生はコメントの中で、「動画

の意図としては羽ばたき振戦なのではと推察するが、それが伝わりにくく、ただの振戦として問題を作成した。」とあるので、このあたりが関係するのかもしれませんが。あの動画は実際の患者さんで、甲状腺機能亢進症が原因ですが、お二人には伝えていません。N先生は「羽ばたき振戦として伝わりにくい」と考えて、満たさないとしたのかもしれませんが。I先生はなぜ、N先生が作ったこの動画1の問題はA-1Eを満たさないか肝型のでしょうか。

#### SME-I

この問題はCTがついていて、このCTと問題文があれば、動画がなくても解けるのではないかと思いました。

黄

CTがあることが、かえって動画の必要性を失わせているということですね。私は、動画を観て振戦があることが分かっても、甲状腺機能亢進症か褐色細胞腫か鑑別が着かないため、CTで最終的に診断できる、という、問題文、静止画、動画、選択肢がうまく連動した、よい問題だと考えました。でもCTをみて、副腎に異常があると分かって、選択肢を見れば、動画を観なくても解けるという解釈ですね。

確かに、チェックリストの言葉尻をとるとそのように判断することもできますね。

A-3EについてもN先生だけは、満たして

いないと自己判断していますね。

ここについては、N先生にお伺いしてみます。

Rの評価はいずれも0にしています。ただ、同じ動画1を使ったI先生の問題に対しては私自身がR-1Fを満たしていると判断しているのに、N先生の作った問題ではR-1Fを満たしていないと評価しています。その理由について、改めて自分でも考えてみようと思います。

続いて、N先生の動画2を使った問題についてです。N先生自身は、すべての項目を満たしているという評価が得られました。

一方、I先生はN先生が作った動画2の問題に対して、A-1E「【動画の必須性】設問文と選択肢だけでは解答できず、動画（音声）の視聴が必須である。」を満たしていないと判断しています。どのように考えたのでしょうか。

私はこの問題をみたときに、個別の神経所見ではなく、統合して片麻痺として症候をまとめたのが非常にうまいと思いました。また選択肢の粒度も揃っていると感じました。

#### SME-I

両眼の右への共同偏視があるという記載

がありました。これも知識の問題だとは思いますが、これだけで左の病変だということが分かります。そうすると d 左橋出血、e 左頸髄出血しか左の病変がありません。加えて、頸髄病変では共同偏視は出ませんので、動画を診なくても問題文だけで解けます。

黄

それは専門医レベルですね。

このことから分かるこのチェックリストの欠点・改善点は、「問題を作成する教員にとって」なのか「問題を解く学生のレベルで」なのかが書かれていない点ですね。I 先生の評価の仕方が悪いのではなく、このチェックリスト利用者のアンカリングができていないですね。

では、その上で、この問題が不適切な問題なのでしょうか。この症候の専門家ではない N 先生と私は良い問題だと評価しています。

SME-I

これは確かにチェックリストにアンカリングが必要だと思います。

黄

では次の問題として、専門家が適切に「学生が知っておくべき医学知識」を適切に線引きできるのかですね。例えば、循環器内

科の医師が、医師国家試験の範囲は関係なく、最新の心不全治療薬について問いたい、などと言い始めるようなことを心配しています。

改めてチェックリストについて見直してみると、セルフ・ピア・エキスパートチェックが一致しない問題についてはそれぞれの評価者に理由を聞き言語化して貰う必要があると考えました。

つまりピアとエキスパートが「この項目は満たしている」と思っている、SME である作題者本人がセルフチェックで「この項目を満たしていない」と思った理由は、言語化して貰う必要があります。

チェックリストの運用として、項目の一つ一つにコメント欄を付けても記入が煩雑になると考えて、最後に一つのコメント欄で代用しましたが、「チェックリストがつかなかった理由」は個別に必要かもしれません。

SME-I

コメント欄がそれぞれにあった方が、精査はしやすいと思います。専門家であることが、必ずしも問題を作るのに良いとは限らないのでしょうか。

黄

動画音声付の臨床問題を作題するだけで

なく、複数の領域専門家同士でそれぞれの問題の善し悪しを評価し、良い問題と悪い問題を蓄積した後、その差異からチェックリストを作成することを当初考えました。しかし、動画付臨床問題の作題経験のある教員が少なく、私自身がこの作題経験が豊富であることを前提に、基礎となるチェックリストを作成しました。

しかし私自身は臨床領域としては「循環器内科」のSMEであり、それ以外の領域の動画付臨床問題に関しては「学生レベルでみて、良い動画付臨床問題である」だと私が判断しても、その領域の専門家からすると「良い動画付臨床問題ではない」というギャップが生じているようです。

チェックリストの運用のうえで「領域専門家としては動画なしで解ける」となっても、「学生レベルでは動画が必要」となれば、チェックリスト項目「問題を解くのに動画に視聴が必須」は満たしていると評価してもらえればよいのでしょうか。

では、チェックリストを満たしていない問題を提出されたときの対応について考えます。このチェックリストの Essential の項目が埋まるまでは出題してはいけないのでしょうか。

#### SME-I

Essential は満たしてから出題した方が良いと思います。

黄

満たしていない場合、IDer が満たすように問題を改変するサポートをするべきかが課題ですが、まずは「専門家だからチェックが付いていない」という状態ではないか確認する必要があると思います。

では、次の質問です。このチェックリストがなかった場合に、もし SME-N の問題を見たらフィードバックができそうでしょうか。

#### SME-I

それは、フィードバックしにくいと思います。どこの点がダメなのかがうまくフィードバックできません。専門領域の違いで項目の判断が異なる部分もありますが、基本的なフィードバックの基準として使えます。

## 付録5 SME-N のインタビューの文字起こし

黄

はい、よろしくお願いいたします。

改めまして研究へのご協力ありがとうございます。  
ございます。

この研究は、動画付き臨床問題を今後増やしていく上で、基準となるチェックリストがあることが、役立つと考えています。チェックリストに対する忌憚のない意見を頂けると、今後に繋がるので、何とぞよろしくお願いいたします。

N 先生は今回は I 先生と比べると動画付臨床問題の作成経験は多いわけですが、内科医としては 10 年以上の経験と、教育担当、医学博士を持っているというところで、一定以上の問題作成能力がある方だとしてお招きしています。

このチェックリストに関して使ってみてどうだったか、どのような事を感じたかなど教えてください。

SME-N

まず Yes/No ではなく、チェックをするかしないかという評価方法だったので、最初は迷ってしまいました。

その Essential か Further、E または F と

いう点は、ちょっと特殊だったので、一瞬面くらったところがあったかなと思います。

意味を理解すれば、問題なく使えましたし、確かにこういう風なチェックをして作問するという事は、動画付臨床問題に限らず、通常も作問でもあんまりやってこなかったです。

問題をレビューする際には文脈的に「この問題は、ここが不適切だよね」といったことは言われますけど、こういうチェックリストがあることでちゃんと項目がクリアになっていいなと思いました。

黄

ありがとうございます。最初におっしゃってた、「チェックをつけたら良い評価」なのか「チェックを付けない方が良い評価」なのかという点は、もう 1 人の SME、I 先生も似たことを言っていました。これは、「いい方がチェックがつくのか悪い方でもチェックがつくのがわからなかった」とおっしゃってるところがあって、そういうところはありませんか。

例えば、インフルエンザの問診表でも、回答欄の左がすべて「はい」、右が「いいえ」

だと、問診結果を確認する医師が、一つ一つの項目がワクチン接種可能な条件に合致しているか考える必要があります。

一方で、ワクチン接種が可能な選択肢が、回答欄の右側に揃うように回答の選択肢の左右を都度入れ替えてある問診表の場合、記入する方には多少認知負荷がかかるものの、「ワクチン接種が可能か」は一目瞭然です。

その発想で、このチェックリストは、チェック項目はすべてポジティブな評価の時に着くように設計したつもりです。

改めて、これらのチェック項目で、チェックを付ける方が良い評価か悪い評価かわかりづらい項目はありますか？

SME-N

「右側に全部チェックが付けばOK」という風にしたのであれば、Essential の列と Further の列は分けずに一列の方がわかりやすいかもしれません。ただそのデメリットとしては Essential か Further かを意識せず、機械的にチェックを付けてしまい易くなるかもしれません。

Essential と Further があることで、Further は無理にチェックが付かなくても良い、ということが分かったら、あんまり迷わなか

ったと思います。

黄

I 先生とのインタビューでは、

R-3F 【設問の公正性】

設問文および動画は、本質的な評価とは無関係な「引っかけ」要素（＝臨床的に非妥当な情報）を意図的に排除している。

この「排除」という言葉には、悪いニュアンスを感じて、チェックが付く方が良いのか、悪いのか悩んだ、というコメントが見られました。

N 先生はどうですか？

SME-N

確かに日本語的には「排除」という単語は、ちょっと強めなのかもしれませんが、私の場合は、先に「チェックがつくのが全部ポジティブな評価という認識」ができていたので、逆にあんまり違和感はなかったですね。

黄

了解です。

チェックリストの表現と関連して、略語についての感想を伺います。

医学教育分野別認証：JACME などでは、評価項目で基本的水準を Basic、質的向上

のための水準を Quality と略しています。そのため、JACME では領域 1 の 1 つめの評価項目における基本的水準のことを、B1.1.1 などと表現します。同様に、質的向上のための水準は Q1.1.1 などと表現されます。口頭で読み合わせを行う上で、この Q が数字の 9 と混同されやすいことから、Basic と Quailty という表現を避け、Essential と Further としました。

また医学的には Further よりも Advanced の方が使い慣れていると思いますが、E と A で A の方が優れているという表現は、アルファベット順と異なること、A はすでに評価項目の Appropriateness に使っていることから、避けました。

この点に関してご意見を下さい。

SME-N

Further よりも Advanced だと確かにわかりやすいです。

そもそも略語や英語を使う必要はあるのですか？ E と F も基本と推奨という風に日本語にするのはどうでしょうか？

黄

ID における ARCS モデルのように、略語の頭文字をとってチェックリストを表現したいと考えました。その際に読みやすさやわかりやすさを考えると、同じ単語を 2 度使うのは避けてこの ACREF にしてい

ます。

先ほどおっしゃっていましたが、この E と F の意味、E が基本で F がより推奨されるっていうニュアンスは、どの辺で意味がわかりましたか。

チェックリストと一緒に送った説明を読んだらわかったのでしょうか。

当初この研究では、チェックリスト無しで SME の先生に動画付臨床問題を作ってもらい、その後このチェックリストを提示して、チェックリストの不足したポイントを改めて自分でブラッシュアップして貰うことを予定していました。しかし、御自身で用意した動画でなかったり、チェックリストを提示していない場合に問題作成に時間がかかることを予想して、その部分は省略しています。

N 先生は、チェックリストをみてから見ながら問題を作ったのか、問題を作ってからチェックリストを見たのか、どちらの手順で問題を作りましたか？

SME-N

さきに問題を作りました。

問題作って、そのあとチェックリストの項目を見て、そうかそうかと思いながら問題を見直しました。でも、そんなに大きく再

編集（改題）はしませんでした。

その理由としては、このチェックリストほど明確ではなかったにせよ、自分自身が動画付臨床問題を作る上で、大事だろうなど思ったポイントがあって、それと一致していたのだと思います。

黄

では今回提出した問題は、チェックリストを見る前に作題した状態から、チェックリストを経て追加編集した点はない、ということでしょうか。

SME-N

そうです。

黄

わかりました。このチェックリスト自体でわかりにくかった点はありませんか？

SME-N

そうですね。結構使い慣れてない単語が多いなと思いましたね。

また提示された動画で、音声があるものがないものがあつたため、音声についての評価ができなかった。

後は、注釈の文章が長くて、すぐに理解ができなかった。

学習支援の工夫とかの項目も、今回は、我々自身が動画を編集してない点でも評価しづらかったです。

黄

ありがとうございます。

このチェックリストは、本来、動画付臨床問題作成のエキスパートが複数に集まって、みんなで問題を作って、エキスパート同士でどの問題が良いか悪いかを話し合いながら言語化して作成することを念頭においていました。

しかし動画付臨床問題作成のエキスパートが少ない現状を考慮し、まずは動画付臨床問題作成のエキスパートである私自身が、良い問題・悪い問題を言語化してチェックリスト化しています。

その中で、学習者が一生経験しないような希少な疾患を問題化してもダメで、身近なテーマであることや経験し得るという点を、リアリティとして Essential な項目に含めています。

SME-N

逆に希少な疾患を動画付臨床問題として出題することにも、価値があるのではないのでしょうか。

黄  
確かに、現状の医師国家試験でも、我々が  
国家試験でしか見たことのない疾患とい  
うものがあります。これらの疾患は頻度は  
高くないが、見逃さないように、医師の基  
本的知識として出す意義はある。同様に、  
動画付臨床問題では、頻度の高い疾患だけ  
でなく、その動画付臨床問題を解いたこと  
によって、学習者が実際にその疾患に出会  
ったときにどのように行動して良いかが  
分かる、という意義を見いだすことができ  
るかもしれない。評価としては相反してい  
るかもしれません。

SME-N

相反していても良いのではないでしょ  
うか。頻度の高い疾患であるコトがEssential  
であり、希少な疾患だが意義の高い疾患を  
出題することを Further とすることもで  
きます。Essential も Further もどちらも満  
たさないのが最悪な状況だと考えれば良  
いです。

この問題が、何を意図して作ってるという  
ポイントが明確にあれば、どちらの視点で  
作っても良いと思います。

黄

チェックリストの評価を見ていこうと思  
います。

例えば今回は I 先生の作った問題を評価

して貰いましたが、このチェックリストが  
あることで、他人の作った問題に対する批  
判というか、評価がしやすくなりましたか。

SME-N

そう感じます。

このチェックリストで、チェックが付く方  
が良い、ということが明確に確立してしま  
えば、良いか悪いか、というのはクリアに  
言えますよね。

「この項目はやっぱ満たしてないで、ここ  
この修正をお願いします」ということが言い  
やすいので、存在価値があると思います。

黄

チェックリストの評価を見ていきます。

N 先生の作った問題は、N 先生自身のセ  
ルフ評価の方が、ピアの I 先生の評価、エ  
キスパートの黄の評価よりも低い、という  
傾向がありました。

例えば、N 先生の動画 1 の問題に対する  
セルフ評価では

A-1E "【動画の必須性】

設問文と選択肢だけでは解答できず、動  
画（音声）の視聴が必須である。"

がチェックを満たさない、となっていまし  
た。そのように自己評価した理由を教えて  
ください。

SME-N

AI で作成した CT を付けた問題ですね。これは「内分泌科医であれば、動画無しでも解けるな」といった意味ですね。CT と選択肢だけで選択肢から絞って答えを出すことができます。

そこでチェックリストにチェックを付けるか悩みました。動画を観ることを必須にするために、設問文の中で「動画に示す主訴で来院した」という風にしました。

黄

その設問文の表現は、うまい工夫だと思いました。

これによって、動画の視聴が必須になると思ったので、私はエキスパートとしてこのチェックリストの項目 A-1E は満たしていると評価しました。

ちなみに、I 先生も御自身の問題で A-1E を満たさないとしている問題がありました。インタビューでは、例えばこの「コップを持っている時に手が震えた」という情報を設問に加えたことで、動画の視聴が必須ではなくなったと自己評価しています。この表現だけで、主訴が動作時振戦であるということが分かるから解答が絞れてしまうということでした。

今回の 動画 1 と 2 のいずれも、私自身の専門領域ではないため、I 先生の「コップ

を持っているときに手が震えた」という情報や、N 先生の CT と選択肢だけでは私は解けませんでした。そのため、私は N 先生の動画 1 の問題は「動画をいないと解答できない」と判断したため、A-1E は満たしていると判断しました。

専門領域の動画を用いて動画付臨床問題を作題すると、専門家自身が自分に対してだけ「厳しい評価」、チェックをしてしまうことが分かりました。

筆記試験ではこれと逆の現象が起きていると感じています。

筆記試験では、専門家が自分自身の高い専門性を考慮せず、学生に専門性の高い解答を求める傾向があると感じます。(補足：MCQ と異なり、一意に答えを絞らなくて良い、また正答らしい類似選択肢を提示しなくて良いから)

SME-N

私自身は、I 先生の動画 1 の問題は、全く解けませんでした。

黄

チェックリストとしては、セルフ・ピア・エキスパートの 3 名の評価が一致する、つまりあるチェックリストの項目は満たさない全員が評価した、あるいは満たし

たと全員が評価することを期待していました。

しかし実際には、セルフチェックだけ、満たさないといった現象が生じており、領域の専門家だからこそ動画無しで解けてしまう、あるいは動画から得られる情報がより多く、それに一致した問題を作ることが困難であった事が分かりました。

では専門家によるセルフチェックだけ満たさない、という結果がこのチェックリストで得られた場合、このチェックリストをどう判定、活用すれば良いのでしょうか。

#### SME-N

専門家は問題のチェックリスト評価をしない、ということが考えられるのでしょうか。

黄

作題に当たって、チェックリストの Essential 項目だけは、満たして提出して貰うことを念頭に置いていました。しかし、「専門家だからこそチェックが付けられない現象」があるため、Essential をすべて埋めた問題を作れず、専門家は動画付臨床問題を提出できない可能性が出てきます。Essential 項目を満たしていない問題の提出を許可して、いったん問題のプールに集め、専門家以外がチェックリストで評価して、ピアとしてはチェックリストを満たし

ていることを作題者に伝えられ場良いでしょうか。

#### SME-N

もしかしたら、Essential 項目は、作題者以外が評価、Further 項目は本人が評価するという風な使い分けができないでしょうか？あるいは評価の重みを変えて、作題者本人の点数を小さくするというのはいかがでしょうか。

黄

なるほど。

例えば、Further は問題作成の根幹であり、作題者本人にしか改善の余地がありませんが、Essential な部分は、設問文などの部分的な改変で対応出来るかもしれません。選択肢の粒度など作題者にしか代えられない Essential な部分もありますが、Essential と Further の評価者を分けること、領域専門家である作題者の配点を抑えるなどのアイデアをいただきました。

ちなみ、N 先生自身は、自身の作った動画 1 の問題に対して、

#### C-1E " 【品質の担保】

動画（音声）は、所見を識別し、臨床的な判断を行うために十分な品質である。"

を満たしていない、と評価しています。

この動画自身は先生が作ったものではないのですが、そのように判断した理由を教えてください。

SME-N

あの動画では、どのような振戦かという事を評価するのに十分な動画ではないと思いました。

黄

どのような点があれば、十分な動画だと言えるようになりますか？

あの動画自体は、N先生の専門領域である「甲状腺中毒症の患者の振戦動画」でした。

SME-N

なるほど。あの動画では甲状腺機能亢進症の振戦としても矛盾しないものの、他の振戦と区別することも難しかったので、羽ばたき振戦を提示したかったのかと考えました。

そう考えると、あの動画では羽ばたき振戦か評価するには不十分な動画だと考えました。

黄

私はこの動画1に「振戦」という情報しか引き出しておらず、それで問題を作るとを念頭に考えていました。一方で、脳神経内科のI先生は「動作時振戦」という情報

を加え、代謝内分泌内科のN先生は「甲状腺機能亢進症の振戦と肝疾患に伴う振戦の鑑別」を問題に織り込んでいます。

これはI先生とのインタビューでもありましたが、例えば収縮期雑音を素材にした問題を作った際に、専門家以外であれば、大動脈弁狭窄症でも僧帽弁閉鎖不全症のどちらもでも問題を作れますが、循環器の専門家である私であれば、その収縮期雑音をさらに聞き分けて、大動脈弁狭窄症としてしか問題を作れない、という現象に似ていると思いました。

SME-N

振戦は静止画では伝わらない所見で、動画で示す意味があります。

しかし、設問文に「振戦がある」としてしまうと、動画の価値がなくなってしまうので、単に「振戦がある」というレベルではなく、「振戦の鑑別を行う設問」や、「振戦の鑑別ができるクォリティの動画」が必要だと思いました。

黄

先ほどの事例ですが、心尖部に収縮期雑音があるという情報だけで、僧帽弁閉鎖不全症と診断する問題を見た場合、多くの非専門家は妥当な問題だと評価すると思いますが、循環器の専門家である私は、大動

脈弁狭窄症の放散音かもしれませんので胸骨右縁第 2 肋間の心雑音を聴いていない動画は不適切だと判断してしまうのと同じですね。

さて、引き続き動画 1 をつけた N 先生の問題ですが、I 先生は

C-4E

"【設問文の必須性】

動画と選択肢だけでは解答できず、設問文の情報(例:患者背景、具体的な問い)が解答に必須である。"

を満たしていないと評価しています。

ここはインタビューの中で、この問題の選択肢と CT だけみれば、設問文を読まなくてもとけるからだと言っていました。このあたりは、やはり「脳神経内科の自分であれば」という風に解釈しているようです。このあたりは「学生のレベルで」という注釈やアンカリングは必要そうです。

SME-N

脳神経内科の専門医であれば、「振戦」というキーワードだけで様々な鑑別疾患が挙がって、そのうえで、この CT を見れば、動画を観なくて良いということになるのでしょうか。

黄

厳密には N 先生の問題には「振戦」とい

うキーワードは意図的に出ないようにしてありますが、すでに動画 1 で問題を作っていたため、この問題が「振戦」に関する問題だということが分かっていたからだと考えられます。

続いて動画 2 の問題に移ります。

3 者の評価の一致度は、C の項目がすべて一致しています。全体的に動画 1 の問題よりもセルフチェック・ピアチェック・エキスパートチェックが一致しています。

N 先生の動画 2 の問題は、動画付臨床問題の作題経験に比して、I 先生の作題よりも良い問題だと、評価時に思っていました。I 先生に伺ったところ、I 先生の問題にも深い意図があったのですがそのあたりはあとでお伝えします。

動画 2 は普通に作題すると、右手の MMT(徒手筋力テスト)がいくつか、といった工夫のない問題になりがちです。N 先生の作題では、複数の MMT 所見を組み合わせ、右側の片麻痺であると捉えた点は、私は非常に面白いと感じました。そのため、N 先生のセルフチェックと私のエキスパートチェックは比較的一致しています。

つまり、ピアレビューの I 先生の評価だけ、厳しいという結果になっています。

I先生はN先生の作った動画1の問題に対して

**A-1E "【動画の必須性】**

設問文と選択肢だけでは解答できず、動画（音声）の視聴が必須である。"

を満たしていないと評価しています。

インタビューでその理由を聞いたところ、動画2の麻痺は「弛緩性麻痺」ではないから、急性期の麻痺所見としては妥当ではないとのことでした。つまりあの動画から得られる所見を脳神経内科医として厳密に評価すると、N先生の作った動画2の問題は適切ではないという意見でした。

また

**A-4E "【所見の非開示】**

設問文は、動画から学生が読み取るべき核心的な所見を、文章で記述していない。"

も満たしていないと評価しています。その理由として、右の共同偏視がある時点で選択肢が絞られるという意見でした。つまりこの部分も、「脳神経の専門家ではない」N先生と黄ではA-4Eは満たしていると判断し、「脳神経の専門家である」I先生は満たしていないという、専門性によるギャップでした。

このチェックリストの有効性は、セルフ・

ピア・エキスパートが一致することで示せると思っていましたが、そうなりませんでした。

領域専門家と私だけが評価すると、単なる不一致になってこのことに気づけなかったと思います。

SME-N

つまり、動画という教材が教科書と異なるということだと思います。

文字であれば、「振戦」という情報に多様性がない一方、動画になると「振戦」の情報量が増える。だから観る人によって所見の取り方が変わる。だから医学のアートの部分（専門性の領域・知識を超えた能力）が、動画の部分だと個人個人によって分かれてしまうのでしょう。

黄

確かに文字で「振戦」と書かれていれば、作題者・解答者の両方にとって、「振戦」だし、「弛緩性麻痺」や「収縮期雑音」と書けばそこに疑念を挟む余地が出ません。しかし、動画や音声を用いると、その解釈にそれぞれの専門家のより専門的な技能や医師としてのartな要素が出てきてしまうのですね。

SME-N

そうすると専門家を含めて、誰が見ても、所見として納得出来る動画を作らないといけないという事です。

黄

現実的には難しいですね。

SME-N

難しいのであれば、今回の問題でCTをAIで生成したように、動画すらAIで作るしかありません。

黄

他の「逃げ方」として「これは再現動画です」と明示する方法もあります。そもそも筋力低下を認めているのに筋肉の萎縮がないといった臨床的な不一致や再現できない所見があります。再現であることを示すことによって、専門家からの追及を免れる、許してもらうという方法です。

動画を使えば問題として質を挙げられる、より高い学習効果が得られるという発想からこの動画付臨床問題の普及に努めています。動画を使うからこそ起きる課題があることが分かりました。

SME-N

最終的に、動画を観なくても解ける専門家がいたり、専門家からすると一部動画に不

適切な部分があったとしても、専門家以外の教員がその問題に取り組んで、納得して解ける問題になっていけば良いのではないのでしょうか。

黄

私の今回の動画1・2の症候の専門家ではありませんが、動画付臨床問題作成ならびに学生教育のエキスパートです。その私の視点から、チェックリストを満たした、良い問題だと判断しているのに、作題者、つまりその領域の専門家だけがチェックリストを満たしていない、という現象が起きたことを非常に面白いと思っています。

このチェックリストの使い方として、Essential 項目のセルフチェックが満たされるまで作題者だけでブラッシュアップするという方法を考えていましたが、その使用法は妥当ではないと気づきました。また作題者のレベルではなく、「学生のレベル」というアンカリングが必要ですが、教員自身が適切に「学生のレベル」を把握できているかは別の課題として存在します。

SME-N

それは臨床に浸かれば浸かるほどそうなると思います。

黄

逆に、学生が学ぶレベルが、臨床の現場の10年前の常識で止まっていることがあります。(心不全治療薬にジギタリスを挙げる)

このチェックリストの使い方をまとめると、作題者自身が Essential 項目を満たす努力はしますが、満たせなかった場合でも問題プールにいったん提出して貰うことが妥当だと思います。その上で、ピアがその essential な項目をどう満たしていくのか「寄せていく」ような作業を行っていくのだと思います。

#### SME-N

そうするとやっぱり動画の準備をどうするかが課題です。紛れもない自分の実際の患者さんの動画を提示すれば、嘘いつわりありません。だけど実際の患者さんの動画でも典型ではない所見である場合をどうするか。

あるいは先ほど言ったように、動画を AI で生成するか。撮影した経緯や状況が分からない動画を二次利用することが一番危ないと思います(適切な動画か判断できない)。

黄

元々、本研究は、動画付臨床問題に利用できる動画を集めるサーバーを立てること

もスコープに入っています。そのため、将来的な運用として、自分自身が撮影・入手した動画でなく、「他人が撮影・入手した動画」で問題を作ることが想定されます。そのため今回も I 先生、N 先生自身には動画を用意して貰わず、こちらから提示しました。自分が用意していない動画でも問題を作ると言うことが、運用上求められると予想しています。

#### SME-N

そうであれば、せめて提示した動画が「何の動画であるか」、動画の背景情報は提示すべきだと考えます。

黄

今回提示して動画1は、甲状腺機能亢進症の動画です。あえて情報を開示しませんが、その情報があることで作題は変わりましたか。

#### SME-N

疾患名や撮影の状況がわかれば、それに応じた問題を作成出来ました。提示された動画は甲状腺機能亢進症としてはやや動きがお大きいが矛盾しない動画でした。ただ、動画の情報がなく、羽ばたき振戦の鑑別をさせたいと考えたため、不適切な動画という評価をしています。

黄  
最後に、このチェックリストがあることで、動画付臨床問題を作ったことのない教員が作れそうでしょうか。

SME-N

動画付問題の作成経験のない教員にとって、道しるべとしては、とても良いと思います。

使い方としては、問題作成に入る前に、最初にこのチェックリストを提示して、読み込んでもらい、このチェックリストに合致する問題を作って下さい、と伝えておくべきだなと思いました。

このチェックリストを用いることで、作題される問題が、一定のクオリティになる可能性は、高くなると思います。一方で、問題を作って提出するだけに比べると、問題を作る側のハードルは上がる、エフォートを割かれてしまうかもしれません。ただ 1 回経験すれば、次は作り方が分かっているのです、作りやすくなるのではないのでしょうか。

黄  
今後の使い方として、ある教員に作成してもらった上で、チェックが付いていない項目を示すだけでは不十分で、教員としてはどう改善して良いか困ると思います。チェックが付かない項目に対して、IDer

である私が、具体的な改善案を示すと、納得が得られると考えます。

これまで、自分以外の教員に作題して貰った動画付臨床問題を私が改善すると、オリジナルの作題者も、オリジナルより良くなったとフィードバックしてくれていましたが、具体的にどの点がどう変わったかという点を伝えることが難しかったです。その改善点を言語化できないために、次のその教員が単独で問題を作るときにどのような工夫をすれば良いかが伝えられていませんでした。

今回チェックリストができることで、改善すべきギャップがどこにあるのかということ伝える、言語化する役には立ちそうだと自分で考えています。

SME-N

確かにそれはとても良いです。作る人とチェックする人との会話をスムーズにする、共通言語という事ですね。

黄

今後どのように教育を担う医師教員の負担を軽減するかが課題と感じています。

教育に特化した私のような教員以外に、デザイナーやプログラマーが手伝える部分はあるが、医学の基本的な知識がないと、

このチェックリストを評価はできないと  
わかりました。

しょう。

本日はありがとうございました。

例えば、教員の作った問題に対して、

**C-3E "【選択肢の適切性】**

選択肢は、「粒度」(例：身体所見の種類、  
検査の種類) が揃っている。

また、所見を問う場合、単なる「見たまま  
の記述」ではなく、「鶏歩」のような医学  
的概念で構成されている。"

を達成していない、と判断するには、選択  
肢が治療なのか検査なのかが見分けられ  
る程度の知識は必要で、医師以外だと看護  
師などが候補に挙げられると思う。

動画付臨床問題に限らず、従来の紙ベース  
の問題ですら、我々教員が使用する問題作  
成チェックリストというものが存在しま  
せん。動画付臨床問題では、動画の持つア  
ートの部分での差がさらに出るため、明確  
な作成基準となるチェックリストがより  
必要になると感じました。

今後、医師国家試験の CBT 化、動画付臨  
床問題が出題されると思われますが、現在  
導入されている OSCE(客観的臨床能力試  
験) ですら、標準化された評価基準に対  
して全員が納得している訳ではない。動画付  
臨床問題でも同様に、専門家によっては妥  
当でないという評価のある問題も出るで

## 付録6 エキスパートレビューの文起こし

黄

今回、2人の教員に作ってもらいました。1人は神経内科、もう1人は内分泌内科で、私の同期と後輩です。医師としての実力は十分ある方々で、神経内科の先生は以前一緒に作問経験があり、内分泌内科の先生はほぼ初めて作る状況でした。動画だけ渡して、手の震えの動画と、神経診察で右手と右足は拳がるが左側が動かない動画を渡し、それぞれ専門に近い動画を渡しました。今回は動画編集能力までは求めず、将来的にレポジトリの動画を使って、自分が撮っていない動画でも作問できる想定で、あえてそうしました。

共有した Google ドキュメントが2人の教員が作った問題で、チェックの集計も出しています。ACRF チェックリストもリンクで送りました。点数は、セルフレビューがOKで1点、ピアレビューがOKで2点、僕(エキスパート)がOKで4点として、1・2・4の合計で見える化しました。7は全員一致、0も全員一致です。意外と6(本人はダメと思っているがピアと僕はいいと思う)があって、これはなんなんだとなりインタビューしました。6になった理由は把握しています。

例えば麻痺の動画では、内分泌内科の先生は「片麻痺だから」とサクサク作って、僕もいい問題だと思ったんですが、神経内科の先生は「これは痙性麻痺に分類され急性期に起こり得ない」と深読みして、痙性麻痺で合わせようとする疾患が限られて作問できないと言い出しました。専門家ほど厳しくなって作りづらい現象が起きていました。これは循環器でも同じで、文章で「収縮期雑音がある」と書いてあれば多くの人は大動脈弁狭窄でも僧帽弁閉鎖不全でも問題を作れますが、音声で提供されると、循環器の専門家は「これは大動脈弁狭窄じゃなくて僧帽弁閉鎖不全の音だから作れない」と言い始める、というように、得られた素材からの情報精度が高い分、幅が狭まって作りづらいことがあると感じました。点数で「いい問題」を決める話でもないなと思い、チェックリストの言葉を変えることも含めて、運用や価値観、ID的に意味があるのかを相談したいと思っていました。

浅田

今の話を聞くと、ピアで組んだ人がそれぞれ専門に近い動画で作っていましたよね。ピアレビューする時に、専門が入るか入らないかを前提条件にするかどうか、今の説

ではどっちですか。

黄

大学でも卒試問題を作るときは、その領域の先生が見るのを必須にしています。そうしないと他の先生だと分からないので、近接領域の専門家が見るべきだと思います。動画問題も近接専門家が入れたらいいと思いますが、その結果、専門家の深読みは今後も起こり得ます。

浅田

それは、動画や画像を AI に喰わせる時に、結局説明がないと AI が理解できないのと同じです。今まで文字で「この雑音」と書いていたのを音に変えるときも、「これはこういう音」「これはこういう麻痺」という解説が文字で入っていれば、動画や音が合っているかどうかは別として、「この症状ならこういう問題が作れる」というところは繋がります。今はその真ん中が抜けているので、「この動画とこの問題が合わないならダメ」になってしまう。動画＝症状、症状＝問題という形で、どちらかが残るようしておく、最初の悩みが少し解決すると思います。

黄

動画そのものだけで一発勝負ではなく、動画に一言タグをつける、例えば「痙性麻痺」

「収縮期雑音」などをつければ、タグだけで問題を作ることは成り立ちますね。動画の質や内容に影響されすぎずに SME は作問できる。

浅田

もちろん動画を見せた方が、「ここ見えてないから作れない」などが分かるので必要ですが、「この麻痺前提で作る」「この雑音あり前提で作る」という統一性は測れます。作問としては学習者の能力を測るわけではないので、聴診結果を音で示すか「聴取した」と書くかは作問者側にとってはどちらでもよくて、「ある／ない」が分かればいい。学生が理解できるかどうかの話になると、文字より動画で判断させようになります。

黄

最終的には、教員が集めた動画をレポジトリに入れて、診察動画、収縮雑音などタグをたくさんつけて置く。そこから拾って作問してもらう。動画を一回見た後はタグを見て作って、もしタグは合ってるけど動画がダメなら似た動画を探し直せばいい。必要ならその時に撮ればいい、という理解で良いですね。

浅田

今は動画を作った人に文章で「こういう状

況「CTが映ってる」などを書いてもらっていますが、同じことです。要は「こういう状態がある／ない」が分かればいい。

黄

そうすると、専門家だけの“深読みのジレンマ”みたいな問題はある程度解決できるかもしれない。運用の工夫なのか、チェックリスト自体の問題なのか。

浅田

作問者の思考過程で必要な部分と、学習者に求められる思考過程で必要な部分がずれることがあります。動画の話はそれです。国家試験で「収縮期雑音がある」と書けば事実として共有されるけど、動画になった瞬間に、学生・一般医・専門家で受け取れる情報量が違ってギャップが生まれる。これは起こり得る話です。

黄

インタビューでも、動画になった瞬間にギャップが出るという話がありました。例えば手の震えの問題で、僕と内分泌内科の先生は「動画を見ないと振戦の種類が分からない」と思ったけれど、神経内科の先生は「”コップを持った手が震えている”時点で動作時振戦だから絞れる」と言いました。専門家の知識量の差だと思いました。今日はエキスパートレビューとして、この

チェックリストが適切か、あるいは有効かどうかを相談したいです。僕としては、PICO や ARCS のように構造化してシステム化し、経験差や領域差があってもルーブリックのように使えることを想定していました。理想は作問者がチェックリストを見て自分で足りないところを考えることですが、どう直せばいいかの答えはここにはないので、他者が批判する時に「何が足りないか」を明確に言語化して示し、改善策まで含めて返すことで足場かけになるようにしたい、という意図です。

浅田

そうなる、セルフとピアの属性は分かりましたが、先生が言う「エキスパート」はどの意味で置いていますか。

黄

医学領域の専門家ではなく、動画付臨床問題を作る上でのエキスパートという意味です。

浅田

そうなる、 $2 \times 2 \times 2$  でパターンが出てきます。7点ならフィードバック不要。3点なら臨床的には OK だけど問題的には NG、5点ならピアだけ NG で専門領域の穴があるかもしれない、など。4点のケースはどういう時ですか。

黄

ありました。手が震えている動画で、45歳男性の症例に「動画に示す」とし、AIで作ったCTも合わせた問題です。僕は「動画を見ないと分からない」構成で、統合情報も使っていて良いと思いましたが、作った先生は「CTと選択肢だけで褐色細胞腫と分かってしまう」と自己評価を下げました。神経内科の先生も「CTと選択肢で分かるから動画は不要」と言いました。僕は学生レベルにアンカリングして、動画で手の震えを見せて選択肢を絞り、CTで確定する段階を想定していたのですが、一般内科の視点では不要に見えた、ということです。

浅田

その場合は、CTが見えすぎて分かっちゃう、という素材側の問題でもあります。動画とCTの必要性は入ってくる。ピアレビューまで行ったときに、「この動画だと分かりやすすぎる／分かりにくすぎる」など、動画自体に問題があるという評価はありますか。

黄

Rの領域に、R1F（スロー再生、比較、ズームなどの工夫）、R3F（意図的に排除している、など）を入れています。Rは推奨項

目で、満たさなくても問題としては成り立つ。ダメだった場合はSMEに聞いて僕が直すつもりです。長すぎる動画はダメだし、診断に不要な部分をだらだら見せるのは意味がないので、削ぎ落としは必要です。

浅田

例えば、テキストで「この雑音が聞こえる」として作った問題は質が良いが、紐づく音声が変わった場合、直すべきは音声の方で、設問本文をひっくり返す必要はない。そういう切り分けがR3Fの話に近いです。

黄

そうすると、直す対象が「設問」か「動画」かで分けた方がよさそうです。設問はSMEが直せるが、動画はエンジニア側が直す、という分担になる。1つの項目に2つの情報を入れられない方がよいが、項目が増えすぎても困る、という悩みがあります。

黄

それとR3Fの表現がつけづらいと言われました。「意図的に排除している」はネガティブに聞こえるので、「排除できている」などの方がいいのか悩んでいます。ただ、神経診察で異常部位だけ映すと答えが分かりすぎるので、右手も見せるなど、必要な“文脈”もある。どこまでが自然な流れで、どこからが冗長かの境界が難しいです。

浅田

冗長性かどうかは尺の問題が大きいです。背中呼吸を聞いてる場面など、目的と無関係なら削ぎ落とす。一方で、頸動脈を触れるならズームで見やすくする、呼吸を1回で済ますか複数回にするか、といった見やすさは R1F の話に近い。エキスパートなら分かるが、一般には分かりにくい場合にどうするかは、R1F で扱えると思います。

黄

なるほど。R2F (提示されたシチュエーションが自分ごととして捉えられるか)も分かりづらいと言われました。研修医として夜間救急で…のようなアンカリングは妥当だと思う一方で、珍しい疾患も国試には出るので、レアなものを見せるべきという考えもあり、矛盾しています。

浅田

それはチェックリストとは別軸で、「国試の範囲に入ってるか」など既存の軸で整理してもいい。COVID-19 を分からせたいのにインフルが混ざって分かりづらい、合併しないものを混ぜたらアウト、などの観点もあります。

黄

集計を見ると A や C は一致が多い一方、

R は 7 がほぼなく、できていない項目が多いです。設問と動画が混ざっている、E であるべきものが F に入っている、といった構造上の問題があるのかもしれませんが。A・C・R それぞれに E と F がある 2 段構造自体はどう思われますか。

浅田

問題はないと思いますが、今の並びだと隙間があって若干複雑に見える。色で分けているから列を統合できそうですが、縦集計をするなら残してもいい。システムで回すなら、末尾が E か F かでカウント条件を付ければ詰められる。

黄

E だけ見せた後に F も見せる、という運用は意味がありますか。

浅田

意味はあります。まず Essential をしっかりやって、その上で Further にするのが ID 的に自然です。見せ方は ICT なら折りたたみにして、最初は F を閉じておいてボタンで開くようにすれば良い。(HTML 形式のファイルによる ACREF チェックリストの運用)

黄

ありがとうございます。修士論文として、

チェックリストの項目だけでなく、こういう見せ方や使いやすさの工夫まで示すかは迷っていますが、使う意味ではツールも提示した方がよさそうですね。

黄  
逆転表現の項目（A4E、A3F、R3F など）が分かりづらいと言われました。全部ポジティブチェックだと示したつもりでしたが、チェックが付くべき方向が分かりづらい表現があります。

浅田  
学生に解かせる問題ではないので、ガイドとして「チェックが付いた方がいい」と明確に伝える方がいい。表現が肯定系でも否定系でも、チェックが付いた方がいいという意図を太字にするなどで示すのはあります。

黄  
修士論文のレベルでは、先生からの指摘を踏まえた Version 2 を作り、否定表現が混ざらないようにし、R の F 項目に入っていたものから Essential を作って、設問と動画の混在を分けて Essential に下ろした、という形でまとめられそうです。

浅田  
研究テーマとしてはチェックリストを作

る、問題作成支援も入ってますね。チェックが付かなかった時にどうすれば付くようになるか、という「作成支援」の部分も重要です。ダメな例だけでなく、満たすための工夫やアドバイスが右側にあると支援と言える。全部に当てはまらなくても、できる範囲で入れると良い。

浅田  
例えば A1E の文言は、前半の「設問と選択肢だけでは回答できない」と後半の「動画音声の視聴が必須」が混ざっている。前半には、設問の誤りや選択肢に妥当な回答が2つある等、別要因も絡み得るので、チェックリストに「筆記問題としての不適切さ」が混ざってしまう可能性がある。動画を使った作問の妥当性を見るなら A・C で見るが、問題評価としては C が先に来る可能性もある。

黄  
想定としては、国家試験にある既存問題に動画を織り込む形が多いので、C が問題になることは少ないかもしれません。僕は「動画を見なくていい問題は動画にする必要がない」と考えていて、A を満たさないなら動画付きにしない、という順序感で作っています。例えば尿の所見なら動画より写真でいいので、A を満たさないなら動画にしない。

浅田

その観点は OK です。ただ A4E (所見の記述) みたいな項目は C に持っていても良い。動画の妥当性は十分あって、問題文が言い過ぎて動画がいらなくなっているだけなら、直す対象は問題文です。動画に本来必要な情報が含まれていないなら A 側、問題文が言い過ぎなら C 側、というように「手を加えるべき対象」で整理すると分かりやすい。

黄

なるほど。上の 3 つは「動画がないと解けないから動画を入れる」。A4E は「動画がなくても解ける問題に動画を付けてしまった」ので、直し方としては「動画を外す」か「本文から余計な情報を削って動画を生かす」か、という分岐になる。動画を残す前提なら、本文の余計な情報を削る方向になりますね。僕は動画を削る発想が最後までなかったもので、整理できました。

浅田

動画とか画像の解説、あるいはタグが付いていると評価しやすい、という話を最初にしましたよね。その解説文やタグに書いてある内容が本文に入ってしまった場合は、「消したほうがいい」という判断になります。これが A のポイントだと思います。

ます。

黄

なるほど。タグにある情報が本文に入っているといけない、ということですね。タグというのは、「読み取ってほしい情報」そのものですよね。

浅田

そうです。動画レポジトリの質が上がってくると、「この動画を入れるなら、この文章は本文に入れちゃまずいな」「意味がなくなるな」という判断が自然にできるようになります。

黄

確かに。タグは動画のコア要素という位置づけになりますね。

浅田

動画の妥当性、つまり医療的に内容が合っているかどうかのチェックにも使えますし、問題文と情報が重複していないかのチェックにも使えると思います。

黄

ただ、その要素やタグを誰が付けるのか、という問題はありますね。専門家に任せると、例えば僕だったら「麻痺」だけがいいところを「弛緩性麻痺」と書いてしまいそ

うです。

浅田

そこは確かにありますね。ただ、今は動画を提供した本人、つまり作問者に解説を付けてもらう形にしていますよね。細かくなりすぎる可能性はありますが、幸い動画の数が何千本もあるわけではない。一度エキスパートレビューをかけて、「細かすぎる」「粗すぎる」を調整すれば、適切な粒度にそろえられると思います。

黄

今後は、動画を撮った人が Nextcloud にアップロードする段階で、最低限のタグを付けてもらう、という運用が必要になりそうですね。管理者的な役割の人にも必要になると思います。例えば神経診察の動画なら、「最低限、何をやっているかが分かるレベル」のタグを付ける、というイメージです。

浅田

それで十分だと思います。最初から完璧を目指すよりも、「作問に使える最低限」を共有できることの方が重要です。

黄

ありがとうございます。ACREF チェックリスト自体の修正だけでなく、動画レポートの運用やタグ付けまで含めて考える

必要がある、という整理ができました。

浅田

ええ。チェックリスト単体ではなく、「どう回すか」まで含めて設計できているのが、このスタディの強みだと思います。

浅田

チェックリストを作ること自体も大事ですが、それ以上に「チェックが付かなかったときに、どう直せばいいか」が見えるのが重要です。評価で終わらず、改善につながる構造になっているかどうかですね。

黄

そうですね。今回やってみて、点数そのものよりも、「なぜチェックが付かなかったのか」「どこを直せば良くなるのか」を言語化できることの方が大事だと感じました。

浅田

ACREF は、その意味では「良い・悪いを決める道具」というより、「直す方向を示す道具」として機能していると思います。動画が悪いのか、設問が悪いのか、見せ方が悪いのか。その切り分けができています。

黄

はい。例えば A で引っかかれば動画側、C

で引っかければ設問側、Rで引っかければ提示方法、というふうに、修正対象が見えるのは大きいです。

浅田

そこが明確になると、誰が直すかも分かりますよね。設問なら SME、動画なら撮影・編集側、運用なら管理者、というふうに。

黄

確かに。今までは全部を作問者が何とかしようとしていた気がします。

浅田

それは無理があります。役割分担が見える設計になっているのは、インストラクショナルデザイン的にはとても良いです。

黄

ACREF Version 2 では、設問と動画を混在させない、ネガティブに聞こえる表現を整理する、RのF項目から Essential を切り出す、このあたりを修正しようと思います。

浅田

それでいいと思います。特に Essential と Further を明確に分けて、「まずここを満たせばいい」という足場を作るのは大事ですね。

黄

ICT 的には、最初は Essential だけ見せて、必要な人が Further を開く、という見せ方もできそうです。

浅田

ええ、それはとても自然です。学習支援ツールとしても使いやすくなります。

黄

今回のスタディでは、実際にシステムを作るところまでは行きませんが、チェックリスト、運用方法、フィードバックの考え方まで含めて提示できれば、十分意義はあるかなと。

浅田

十分あります。「動画付き臨床問題は難しい」という経験則を、ここまで構造化して説明できているのは大きいです。

黄

ありがとうございます。今日の話を踏まえて、Version 2 を作って整理し直します。

浅田

ぜひ。それを見て、また一段階ブラッシュアップしましょう。