

厚生労働科学研究費補助金
(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）)

総括研究報告書

保健師助産師看護師国家試験の問題作成の支援と効率化に向けた
ICT・AI 技術等の活用策の検討のための研究

研究代表者 聖路加国際大学大学院看護学研究科 林 直子

研究要旨

本研究の目的は、保健師助産師看護師国家試験（以下「看護師等国家試験」）において、ICT・AI 技術等を活用した具体的な作問システムを検討すること(研究 1)、また看護師等国家試験におけるコンピュータを活用した試験（CBT：Computer-based Testing）の実装に向けて、CBT システムの試用版を開発・試行し、試験問題の妥当性について難易度、識別指数、IRT スコア等から評価すると共に、出題形式、問題管理システム、受験者側の受容性に関する調査を行い CBT 導入に向けた課題を明示すること(研究 2)である。本研究は 3 年計画で進行している。

初年度（2022 年度）研究 1 では、過去 10 年間の看護師国家試験必修問題（500 問）を、一定の基準をもとに 5 区分の評価（良問、易しすぎる問題、良問だが難問、あまり適切でない問題、いずれにも該当しない問題）に分類し、問題形式、解答形式の観点で小項目ごとの問題の分析を行い、小項目単位での良問ルールの抽出を開始した。また、既存の大規模事前学習済み言語モデルを試用して作問を行い、今後の作問システム開発への示唆を得た。研究 2 については、CBT システムの試験運用に向けた問題収集（500 問）と動画や音声等のマルチメディアを活用した問題の作成、さらに CBT 実装可能性の検討に向けたインタビュー調査の準備を行った。

2 年目の 2023 年度は、看護師国試過去問題（必修問題）の分析結果を統合した改善提案（ルールブック）の作成と、3 種類の大規模言語モデル（ChatGPT 3.5、ChatGPT 4、JSLM）に基づく 4 タイプの作問システム（gpt4、gpt3.5-few shot learning、gpt3.5-Fine Tuning、JSLM）を用いて試験的作問を行い、専門家による問題評価と各モデルの生成性能を評価するための自動評価指標の作成を試みた。過去問題の分析の結果 258 の改善提案が示され、提案内容の分析の結果、200 の小項目に対して 147 のルールが抽出された。また大規模言語モデルの活用可能性の検討の結果、gpt3.5-FT が他と比べて再現性、精度の評価指標値が良いことが示された（研究 1）。研究 2 では、看護系大学と看護専門学校計 3 校の学生 27 名を対象に、10 問のサンプル問題について CBT システムと筆記試験の双方での解答を依頼し、さらに CBT システム導入の可能性に関するフォーカスグループインタビュー調査を実施した。また看護教員 17 名を対象に、看護師国家試験への CBT 導入の可能性に関するインタビュー調査を行った。その結果、学生、教員ともに看護師国試への CBT 導入の受容性が示されたが、一方で CBT の実施に向けた教員側の意識改革などの課題も明らかとなった。

研究分担者氏名 所属 職位

徳永 健伸 東京工業大学 教授

宇佐美 慧 東京大学 准教授

佐伯 由香 愛媛大学 教授

佐々木 幾美 日本赤十字看護大学 教授

米倉 佑貴 聖路加国際大学 准教授

佐居 由美 聖路加国際大学 教授

研究協力者氏名 所属 職位

乾 健太郎 東北大学大学院 教授

宮本 千津子 東京医療保健大学 教授

鈴木 久美 大坂医科薬科大学 教授

伊藤 圭 大学入試センター 准教授

西崎 祐史 順天堂大学 前任准教授

三浦 友里子 聖路加国際大学 准教授

木村 理加 聖路加国際大学 助教

A. 研究目的

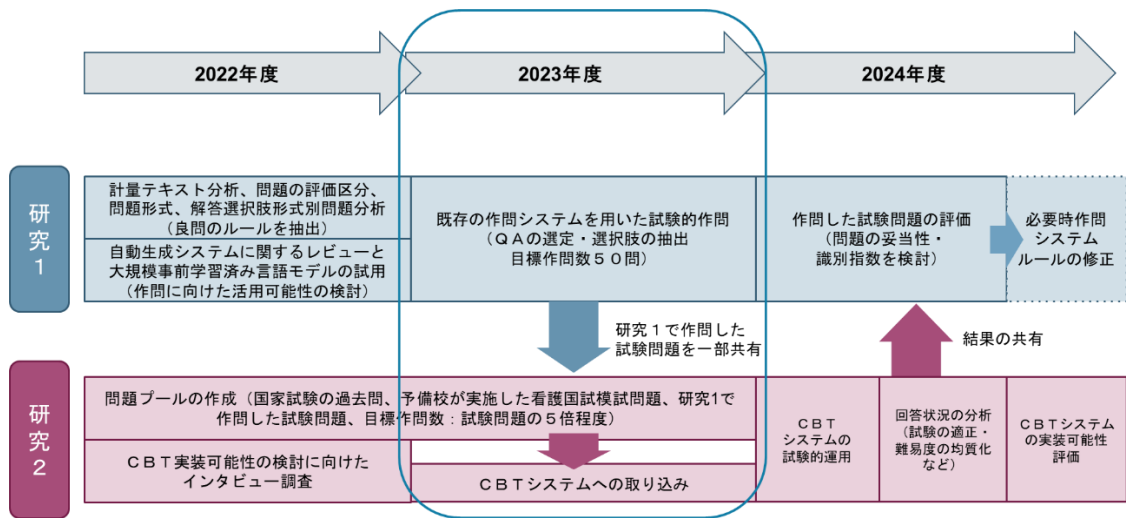
本研究は、保健師助産師看護師国家試験（以下「看護師等国家試験」）において、ICT・AI 技術等を活用した具体的な作問システムを検討すること(研究 1)、また看護師等国家試験におけるコンピュータを活用した試験（CBT：Computer-based Testing）の実装に向けて、CBT システムの試用版を開発・試行し、試験問題の妥当性について難易度、識別指数、IRT スコア等から評価すると共に、出題形式、問題管理システム、受験者側の受容性に関する調査を行い CBT 導入に向けた課題を明示すること(研究 2)を目的とする。

看護師等国家試験については、令和 3 年医道審議会報告書（看護師等国家試験制度改善検討部会）において、災害や感染症等の危機管理の観点からコンピュータ活用の必要性が示され、本研究の目的は正に喫緊の課題であ

る。特に、膨大な問題を作問する人的負荷の軽減と問題の質の担保は、国家試験において重要な課題である。そこで、研究 1 では近年多様な分野での応用が著しい ICT・AI 技術の看護師等国家試験問題作問への活用を検討し、具体的に作問を試みる。さらに、研究 2 では将来の CBT 導入を視野に、CBT システムの実装可能性と課題を探り、基盤となるデータを創出する。なお、看護師国家試験は必修問題、一般問題、状況設定問題の 3 種類で構成されるが、本研究は、基礎段階にあることから、タクソノミー I 型（単純想起問題）の必修問題に焦点をあてる。

本研究は、令和 4 年度から 6 年度までの 3 年計画で進行している。令和 5 年度は、研究 1、研究 2 とともに図 1 に示す 2023 年度の計画を遂行した。

図1 本研究の3年間の全体計画



B. 研究方法

本研究班は、看護師等国家試験委員経験者をはじめ、看護教育学、看護情報学、医学（医師国家試験 CBT 研究有識者）、工学（情報工学、システム設計）、教育学（教育測定学）の専門家が研究分担者・協力者として参与し、研究方法の妥当性と課題を議論し、適宜方法の修正等を図りながら進めた。

1. 【研究1】ICT・AI技術を活用した看護師等国家試験問題作問システムの構築

担当：（研究代表者）林、（研究分担者）徳永、佐伯、佐々木、佐居、（研究協力者）乾、宮本、鈴木、三浦、木村

（1）過去問題の分析と作問提案

①過去問題の内容分析

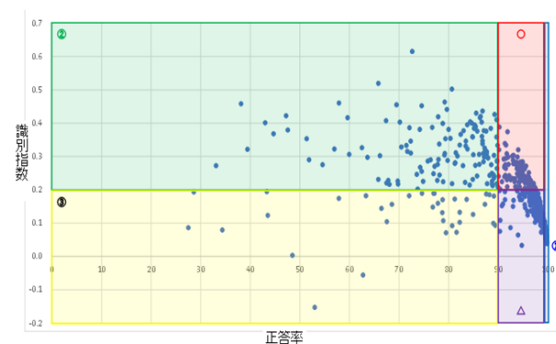
昨年度の研究成果（表1、図1）をもとに、過去10年分の看護師国家試験問題の必修問題500問について、正答率、識別指数から設定した問題の評価基準（良問：評価○、易しすぎた問題：評価①、良問だが難問：評価②、あまり適切ではない問題：評価③、非該当（いずれにも該当しない）：評価△）と、問いの種類、選択肢の種類に基づき小項目を

基準に問題分析を行った。

表1 新たな基準による問題評価結果

評価		正答率%	識別指数	問題数
○	良問	90～99 未満	0.2 以上	100
①	易しすぎた問題	99 以上	—	89
②	良問だが難問	90 未満	0.2 以上	132
③	あまり適切ではない（識別子 0 以下を含む）	90 未満	0.2 未満	36
△	非該当	90～99 未満	0.2 未満	143
合計				500

図2 過去問題500問の正答率・識別指数プロット
* ○、①～③、△の詳細は表1参照



問題分析は研究者1名の初回分析後、研究者3名が分析内容の妥当性を評価し、全員の合

意に至るまで、修正・評価を繰り返し、問題の改善提案を作成した。次に、看護師国家試験委員経験者、あるいは看護教育に従事する研究分担者、研究協力者全員で改善提案を検討し、作問の提案書となるルールブックに掲載する提案内容を検討した。検討の結果、改善提案について可能な限り普遍性ある内容にするべくさらなる内容分析を行い、作問システムのプロンプトとして活用可能なルールを抽出することとした。

②作問ルールブックの作成

①の結果に基づき、263の改善提案について、さらなる内容分析を行なった。方法は改善提案の作成を担当した研究者1名が初回内容分析を行い、研究代表者を含む研究者3名が分析内容の妥当性を評価し、修正と評価を繰り返すことで合意形成を行った。その後、内容分析結果について、分析に携わらなかった研究分担者、研究協力者が分析内容の妥当性を評価し、指摘箇所の最終合意を得てルールブックの完成とした。

(2)既存の大規模事前学習済み言語モデル(ChatGPT)の活用可能性の検討

ChatGPT (gpt3.5-turbo-0301)の潜在的作問能力を評価するために、(1)の過去問題の分析によって作問が困難と判定された4問を選び、その問題のトピック(洗髪、嘔吐、標準予防策、細胞内液)についてChatGPTを使って問題を生成した。ChatGPTへのプロンプトに含める情報として、多肢選択問題の構成要素である(1) 問題文 (Stem)、(2) 正解 (Key)に加え、(3) 看護学の標準的な教科書からトピックに関連する箇所の段落を抽出したテキスト(Text)の3種類の情報を考え、この組合せで8種類 (=2³)のプロンプトを作成し、合計

で32問 (=4トピック×8プロンプト)を作成した。なお、多肢選択問題のもうひとつの構成要素である誤答選択肢(誤答肢)は、自動作問システムの先行研究における重要な研究課題であり、作問者にとっても作問過程において最も労力のかかる課題であることから、誤答肢は事前情報としてプロンプトに含めなかった。また、これら3種類の情報をすべてプロンプトに含めない場合は、トピックのみをプロンプト中に与えた。

プロンプトはあらかじめ与えなかった多肢選択問題の構成要素を段階的に尋ねる単純なものを使い、いわゆるプロンプト・エンジニアリングには注力しなかった。これは実験当時のChatGPTがAPIを通じてブラックボックス的にしか利用できなかったため、プロンプト・エンジニアリングでチューニングしてもシステムが改修された時の動作の再現性が保証されないためである。したがって、今回は前述のようにプロンプト中に含める情報内容を制御することにより性能の変化を分析することとした。

作問した32問は1)過去問題の分析と作問提案の研究チーム(専門家チーム)にフィードバックし、専門家の判断でChatGPTが作問した問題がそのまま試験問題として使えるかどうかを判定することによって評価した。

(3)言語モデルによる誤答選択肢の生成と自動評価指標の作成

問題文と正解をプロンプトに含め、誤答肢をChatGPTに生成させた問題の評価が一番高かったこと、また、多肢選択問題の作成において誤答肢を作る過程が一番労力を要するという作問経験者からの指摘を受け、誤答肢の生成に焦点を当てて研究を進めることとした。

3つの大規模言語モデルを用い、モデルの訓練条件を変えて、誤答肢の生成性能の変化を調査した。性能を分析するためには評価が必要となるが、人手評価は時間とコストがかかるために、まず、既存の自動評価尺度を参考に、生成された誤答肢の良さを自動評価する尺度を考案した。自動評価はモデルに過去問題の問題文と正解を与えた時にモデルがその過去問題の誤答肢をどれくらい正確に再現できるかを計算することによっておこなう。

最初に情報検索の分野で利用されている再現率 (Recall) と精度 (Precision) を使うことを考えた。再現率は過去問題の誤答肢をモデルが生成できた割合を表わし、精度はモデルが生成した誤答肢が過去問題の誤答肢と一致する割合を表わす。過去問題 q_i の誤答肢の集合を G_i 、 q_i の問題文と正解を与えられてモデルが生成した誤答肢の集合を S_i とすると q_i に対する再現率 (R_i) と精度 (P_i) は以下の式で定義される。

$$R_i = \frac{|S_i \cap G_i|}{|G_i|}, \quad P_i = \frac{|S_i \cap G_i|}{|S_i|}$$

再現率、精度はモデルの出力と過去問題の誤答肢との表層的な一致に基づいて計算される。選択肢が名詞句のような短い表現であれば一致する可能性が高いが、選択肢が文のような長い表現の場合、たとえ意味的には同じであっても表層的な一致の可能性は著しく低くなる。つまり、再現率、精度ではモデルの出力を過少評価してしまう可能性がでてくる。この問題を改善するために、我々は一致・不一致の二値判定に基づく再現率、精度を拡張し、表現の類似度に基づく再現率、精度を提案した。まず、最近の自然言語処理の分野で標準的に用いられている埋め込み技術を使って、比較する誤答肢をそれぞれ多次元ベクトルに変換し、それらのベクトルが張る角度によって2つの誤答肢の意味的な類似度

を計算する。モデルが生成した誤答肢と過去問題の誤答肢のすべての組合せについて類似度を計算し、最大の値をとる対を一致したと見なし、その値を再現率、精度の計算に利用する。2つの誤答肢が完全に一致する場合は類似度が1となるように類似度を正規化しておけば、完全一致の場合は通常の再現率、精度を計算するのと同じになる。再現率、精度の定義における分子、つまり、モデルの出力と過去問題の誤答肢の積集合は以下のように書き直せる。

$$|S_i \cap G_i| = \frac{\sum_{j=1}^{|G_i|} \mathbb{1}(g_j \in S_i)}{|G_i|} = \frac{\sum_{k=1}^{|S_i|} \mathbb{1}(s_k \in G_i)}{|S_i|}$$

この式を使うと q_i に対する類似度に基づく再現率 (R_{si})、精度 (P_{si}) は通常の再現率、精度の拡張として以下のように定義できる。ただし $\text{sim}(\cdot, \cdot)$ は2つのベクトルの類似度を返す関数である。

$$R_{si} = \frac{\sum_{j=1}^{|G_i|} \text{sim}(\arg \max_{s \in S_i} \text{sim}(s, g_j), g_j)}{|G_i|}$$

$$P_{si} = \frac{\sum_{k=1}^{|S_i|} \text{sim}(s_k, \arg \max_{g \in G_i} \text{sim}(s_k, g))}{|S_i|}$$

R_s 、 P_s は完全一致の条件を緩和しているとはいえ、モデル出力と過去問題の誤答肢の一対比較の最大値に基づいて類似度を計算している。しかし、これらの尺度では、両方の集合の要素の対応関係について考慮していないため、システム出力の各誤答肢がそれぞれ異なる過去問題の誤答肢に対応付けられる状況とシステム出力の誤答肢がすべて同じひとつの過去問題の誤答肢に対応付けられる場合を区別できない。この問題を改善するために R_s 、 P_s をさらに拡張し、各組の組合せを考慮した類似度に基づく再現率 (R_{cs})、精度 (P_{cs}) を提案した。基本的な考え方は、重複を許さずに作ったモデル出力と過去問題の誤答肢の組の集合の類似度の総和が最大になるような

組合せを見つけ、この類似度の総和を再現率、精度の計算における分子として使うというものである。Si と Gi を節点集合、それらの節点の組を作ったときの対応関係を弧、類似度を弧の重みと考えると、この問題はグラフ理論における重み付き完全二部グラフの最大重みマッチング問題として定式化することができる。この問題を解く効率のよいアルゴリズムが知られている。節点集合 V1、V2 からなる重み付き完全二部グラフの最大重みマッチングを返す関数を MVM(V1, V2) とすると、Rcs、Pcs は以下のように定義できる。

$$R_{csi} = \frac{MWM(S_i, G_i)}{G_i}, \quad P_{csi} = \frac{MWM(S_i, G_i)}{S_i}$$

ChatGPT によって生成した問題の不適切な点として、同じような誤答肢が含まれているという指摘があったため、生成された誤答肢同士がどれくらい似ているかを表す尺度 DV (distractor variation) も用意した。qi に対する DV_i は以下の式で定義される。

$$DV_i = 1 - \frac{\sum_{\{(s_j, s_k) | s_j, s_k \in S_i, j \neq k\}} \text{sim}(s_j, s_k)}{|S_i|}$$

DV は値が大きいほど互いに似ていない多様な誤答肢を生成していることを意味する。

以上の準備のもと、以下のように実験を設計した。まず、評価実験に使うデータとして厚労省から提供を受けた過去 10 年分の看護師国家試験の必修問題 500 問から、選択肢が名詞句あるいは文からなるもの 366 問を選択し、これを問題の中項目の分野分布を維持しながら訓練用、開発用、評価用として 193、87、86 問に分割した。

大規模言語モデルとしては以下の 3 つを使用した。

- ChatGPT (gpt-3.5-turbo-0613)
 - GPT-4 (gpt4-0613)
 - Japanese Stable LM instruct alpha 7B-v2
- 学習方法としては、問題と答の例をプロンプ

ト中に複数提示する few-shot 学習と訓練データを用いた微調整 (fine-tuning) をおこなった。ChatGPT と GPT4 は OpenAI の API 経由で利用し、JSLM は研究者管理の GPU サーバーで動作させた。実験時点では GPT4 では微調整の機能が未提供だったため、GPT4 は few-shot 学習のみ、また JSLM は微調整のみをおこなった。ChatGPT については few-shot 学習と微調整の効果の差を比較するため、両方の学習をそれぞれおこなった。

2. 【研究 2】看護師等国家試験への CBT 実装の可能性の検討

担当：(研究分担者) 米倉、宇佐美、佐伯、佐々木、佐居、(研究協力者) 伊藤、西崎、宮本、三浦、木村

1) CBT 実装可能性に関するインタビュー調査

(1) 調査対象

研究分担者・研究協力者の機縁により、日本国内の看護系大学 1 校、専門学校 2 校に協力を依頼し、各校で学生 20 名程度と教員 5 名程度の参加者を募集した。

学生は 4 年課程の学校においては 3 年生以上、3 年課程 2 年生以上を対象とした。

(2) 調査方法

① 学生を対象とした調査

自記式質問紙により、学年等の基本属性、学校内外での CBT 受験経験、パソコン等の情報機器の所持・使用状況を回答してもらった。次に、サンプル試験 1 として CBT システムの TAO にアクセスしてもらい、動画や音声を使用したマルチメディア問題 5 問、従来型の看護師国家試験の問題 5 問、合計 10 問を制限時間 15 分で解答してもらった。その後、サンプル試験 2 として、サンプル試験 1 のマルチメディア問題と同等の内容の問題

5 問、サンプル試験 1 で出題した従来型の看護師国家試験の問題と同じ問題 5 問、合計 10 問を筆記試験形式で制限時間 15 分で解答してもらった。

その後、5 名以内のグループを設定し、フォーカスグループインタビューを行った。インタビューでは、主に以下の内容を尋ねた。

- TAO 上で解答してもらったマルチメディア問題と、同等の内容の問題の解答しやすさや難易度等の感想
- TAO での問題への解答のしやすさ、筆記試験（紙ベースの試験）との比較
- マルチメディア問題を含めた CBT 形式での試験に向けた学習・試験対策に必要な教育内容やそれが現状の教育で提供されているか

②教員を対象とした調査

自記式質問紙により、年齢、性別等の基本属性、専門分野、教員および看護職の経験年数、国家試験対策の担当の状況に付いて回答してもらった。

次に、個別インタビューまたはグループインタビューを行った。インタビュー調査は、事前に学生を対象とした調査のサンプル試験 1 の問題を TAO 上で閲覧してもらったうえで実施した。インタビューにおける質問内容は主に以下のようなものであった。

- 看護師等国家試験が CBT で実施されることに対する賛否とその理由
- 問題を再利用するために過去問題を非公開とすることについての賛否、理由、非公開となった場合に試験対策や教育のために必要な情報・支援
- サンプル試験 1 のマルチメディア問題の難易度、試験問題としての適切性等についての感想、試験対策するにあたって必要な教育や支援

(3)分析方法

学生に対する調査で解答してもらった、サンプル試験の正答状況をサンプル試験 1 (CBT)、サンプル試験 2(紙)の対応する設問間で比較した。

インタビュー内容については、質的内容分析を行った。

(4)倫理的配慮

調査実施にあたり、聖路加国際大学研究倫理審査委員会の承認を得た（承認番号：22 - AC103）。

2) CBT システムへの問題の取り込み

本研究では CBT システムとして TAO を用いる。本システムに令和 4 年度研究にて看護国試の模擬試験問題の必修問題 250 問を取り込んだが、今年度は次年度の調査のためにさらに 100 問の取り込みを行った。

C. 研究結果

1. 【研究 1】ICT・AI 技術を活用した看護師等国家試験問題作問システムの構築

(1)過去問題の分析と作問提案

①過去問題の内容分析

問題分析では最終的に、全 16 の大項目、50 の中項目、269 の小項目（H30 基準）で構成される必修問題の小項目のうち、過去 10 年間で実際に出題された 200 の小項目に対して、258 の改善提案を作成した。

②作問ルールブックの作成

①の結果に基づき、258 の改善提案について、さらなる内容分析を行ない、200 の小項目に対して 147 の内容分析結果（ルール）が抽出された。

大項目 1. 平成 22/平成 26 健康に関する指標、平成 30 健康の定義と理解 ※は詳細非公開

(3) 既存の大規模事前学習済み言語モデル (ChatGPT) の活用可能性の検討

4 つのトピックについて 8 種類のプロンプトを用いて ChatGPT で生成した問題を (1) で過去問題を分析したメンバーにより評価した結果、問題として成立していると判断されたものは、問題文と正解を与えてモデルが誤答肢を生成した以下の 1 問だけであった。

(4) 言語モデルによる誤答肢の生成

実験に使用したモデルについて評価用データを用いて我々の提案した尺度で評価した結果を表 3 に示す。表中の gpt3.5-few、gpt3.5-FT はそれぞれ ChatGPT に few-shot 学習、微調整を適用したモデルである。また太字になっている数字は同じ行中の最大値を示す。

A. 細胞内液 (正解; 所与)
B. 血液
C. リンパ液
D. 細胞外液

(4) 言語モデルによる誤答肢の生成

実験に使用したモデルについて評価用データを用いて我々の提案した尺度で評価した結果を表3に示す。表中の gpt3.5-few、gpt3.5-FT はそれぞれ ChatGPT に few-shot 学習、微調整を適用したモデルである。また太字になっている数字は同じ行中の最大値を示す。

	gpt4	gpt3.5-few	gpt3.5-FT	JSLM
R	0.147	0.155	0.178	0.101
P	0.088	0.093	0.107	0.060
R _s	0.903	0.906	0.909	0.892
P _s	0.894	0.894	0.896	0.877
R _{CS}	0.901	0.903	0.905	0.886
P _{CS}	0.540	0.542	0.543	0.532
DV	0.116	0.116	0.116	0.122

- gpt3.5-few と gpt3.5-FT の結果を比較する

と few-shot 学習に比べて微調整の効果が大きい。今回、微調整に使用した訓練データが 193 問と小規模であることを考えると訓練データの量を増やすことによってさらに性能の改善が期待できる。また、gpt3.5-FT は gpt4 の数値も上回っていることは微調整のさらなる可能性を示唆している。

- JSLM の性能は一貫して他のモデルより低い、これはモデルサイズによるものと考えられる。OpenAI は ChatGPT や GPT4 のパラメタ数については公開していないが、JSLM の 70 億に比べると 2 桁以上は大きいと推測できる。自前のモデルを構築するにしても、より大きいモデルでの性能評価が必要である。
- 表 4 に評価データ中で選択肢が文であるような問題(11 問)のみの評価結果を示す。表からわかるとおり、選択肢が文の場合には、通常の再現率、精度はまったく機能していない。類似度を考慮することは意義あることだと言える。

表 4 誤答肢生成の評価結果 (文選択肢)

	gpt4	gpt3.5-few	gpt3.5-FT	JSLM
R	0.000	0.000	0.000	0.000
P	0.000	0.000	0.000	0.000
R _s	0.887	0.884	0.881	0.874
P _s	0.881	0.880	0.877	0.856
R _{cs}	0.885	0.882	0.877	0.862
P _{cs}	0.531	0.529	0.526	0.517
DV	0.115	0.111	0.088	0.114

- DV の尺度のみ JSLM は他のモデルより数値が大きい。これは他のモデルに比べて JSLM が生成した誤答肢は互いに似ていないことを意味する。JSLM の再現率、精度が他のモデルより一貫して低いことを考えると JSLM は過去問題の誤答肢を再現することは苦手だが、より多様性のある誤答肢を生成している可能性がある。ただし、それが誤答肢として適切かどうかを判断する

は専門家による評価が必要である。

2. 【研究 2】看護師等国家試験への CBT 実装の可能性の検討

(1) 学生を対象とした調査

大学 1 校 13 名、専門学校 2 校 14 名の学生が調査に参加した。対象者の学年は大学 4 年生が 13 名、専門学校 3 年生が 4 名、専門学校 2 年生が 10 名であった。学校内外でのコンピュータを利用した試験の受験経験がある者は 16 名(59.3%)、自分専用のパソコンを所持しているのは 22 名(81.5%)、パソコンの使用頻度が「毎日」の者は 16 名(59.3%)、自分専用のタブレット端末を所持しているのは 26 名(96.3%)であった。

表 5 にサンプル問題の正答状況を示した。全体として正答率に大きな違いはなかった。

表 5 サンプル問題の正答状況

	度数	CBT	紙	p ^a
		正答率	正答率	
合計	27	0.741	0.752	0.611
問1	27	1	0.89	0.083
問2	27	0.89	0.93	0.317
問3	27	0.48	0.59	0.18
問4	27	0.7	0.67	0.739
問5	27	0.37	0.41	0.796
問6	27	0.89	0.89	b
問7	27	0.89	0.89	b
問8	27	0.78	0.81	0.317
問9	27	0.56	0.59	0.317
問10	27	0.85	0.85	b

表 6 にマルチメディアを活用した問題に対する感想のまとめを示した。紙ベースの試験と比較して、「状況設定は文章よりも理解しやすい」一方で「必要な情報を得るのは文章題より

も時間がかかる」といった意見が挙げられた。また、マルチメディアを活用した問題が出題されることによる、学習や試験対策への影響としては、「臨床に活かせる形で学習できる」といったことが語られていた。

次に、CBT システムでの回答形式についての回答のまとめを表 7 に示した。「解答のマークシートへの転記が不要なので解答に集中できる」「クリックで問題に飛べる」といった点がメリットに感じられた一方で、「画面を見ることによる疲労」「時間配分がたてにくい」「メモ機能がないと不便」といった点がデメリット・不便な点として挙げられた。また、解答の入力や問題送り等の基本的な操作については説明がなくても問題なく対応できた一方、動画・音声の視聴可能回数や使用できる機能の説明、チュートリアル等がないと不安であるとする指摘もあった。

最後に、マルチメディアを活用した問題が含まれる試験のための学習に必要なものとして、シミュレータ・マルチメディア教材による学習を増やすことや、実習・演習で実際に体験することなどがより必要である、といった意見が挙げられた。(表 8)

(2) 教員を対象とした調査

教員を対象とした調査では、大学教員 8 名、専門学校教員 9 名から協力が得られた。教員経験は 9 ヶ月から 27 年、看護職の経験は 2 年 9 ヶ月から 35 年 3 ヶ月と幅があった。学生の国家試験対策担当をしている者は 6 名であった。

看護師等国家試験の CBT 化についての回答のまとめを表 9 に示した。

「マルチメディアを活用した問題を設定できる」「試験の複数回実施のハードルが下がる」「採点等の利便性が上がる」といった利点があり、概ね CBT 化には賛成との意見が多かった。一方、「マルチメディア問題による情報量の増

大の試験への影響」「作問負担の増大」「公平性の担保」「試験実施のためのインフラ整備」「実施費用の高額化」「問題漏えい等のセキュリティ」等の懸念も示されていた。マルチメディアを活用した問題については(表 10)、難易度は紙ベースの試験(筆記試験)と同等程度であるとの意見が多く、マルチメディアを活用することで能力を多角的に評価できるため、国家試験としてはより適切であるというような、肯定的な意見が多かった。一方で、解答時間や問題数は現行の紙ベースの試験と同じでは難しいであろうといった意見や、視覚・聴覚の多様性やジェスチャー等の解釈の違いなどから外国人受験者への配慮が必要であるといった課題が挙げられた。また、マルチメディアを活用した問題を作成する際の課題として、モデルとして出演してもらう人のプライバシーといった倫理的問題や、動画の解釈の多様性といったことが挙げられていた。次に、試験問題プール制のための問題非公開化についてについては賛否がわかれた(表 11)。「国際的には非公開が一般的」、「問題の質保証の体制が適切に整備されていれば問題ない」といった意見があった一方、「過去問対策も学習の一部である」ため公開が望ましい、「予備校等で出口調査等をして対策されるので機能しないのでは」といった意見もあった。また、非公開化時に必要となる情報・支援としては、移行初年度のサポートを手厚くすること、詳細な出題基準、サンプル問題、公式問題集のように出題される問題がイメージできるような情報公開が必要であるとの意見が出されていた。現状の教育におけるマルチメディア・シミュレーション教材の活用状況については、学習管理システム(Learning Management System: LMS)、ビデオ会議システム、動画教材、シミュレータ、模擬電子カルテなどが活用されていた(表 12)。LMS の機能として提供されている簡易 CBT 機能も授業各

回の理解度確認等に利用されていたほか、新型コロナウイルス流行時には単位認定試験にも活用されていた。こうした状況から、看護師等国家試験が CBT 化され、マルチメディア問題が出題されるようになったとしても、現状の教育から大幅な変更は不要との意見が多かったが、マルチメディアを活用した問題への対応のためマルチメディア教材・シミュレーション教材の重要度がより増大することや、試験対策として CBT システムの使用方法についての一定のトレーニングは必要であろうとの意見が示された。CBT に移行する場合の準備期間として、対象となる学生が入学する時点で周知されている必要があるとの意見が多かった（表 13）。

D. 考察

1. 【研究 1】ICT・AI 技術を活用した看護師等国家試験問題作問システムの構築

(1) 過去問題の分析と作問提案

問題分析については、前年度に定めた問題評価の基準に則って過去問題を分類し、小項目単位での良問のルールを抽出することで、今後の国家試験作問に向けた作問ルールブックとして即利用可能な資料の作成に繋がると考える。また、作問ルールの文言を、大規模事前学習済み言語モデルのプロンプトに適用する前提で作成することで、作問システム開発の一助となると考える。

(2) 既存の大規模事前学習済み言語モデル (ChatGPT) の活用可能性の検討

ChatGPT によって生成された問題を分析した結果、以下のような不適切な点があった。

- 明らかに誤答とわかる誤答肢
- トピックに関係のない選択肢
- 正解が複数になる問題
- 複数の解答を要求する問題
- 反意語を両方含む選択肢

- 造語を含む問題

(3) 言語モデルによる誤答肢の生成

評価実験の結果をふまえ、今後の検討課題として以下のような課題が考えられる。

- 多くの事例を用いて提案した自動評価尺度による評価と人手による評価を比較し、自動評価尺度の信頼性を確認する。
- よりサイズの大きなオープンモデルの性能の限界を確認する。
- 1)に示した作問のガイドライン（ルールブック）のような専門家の知識を言語モデルに取り込む方策を検討する。
- モデルが生成した誤答肢候補を作問の専門家が利用できる枠組み、および専門家からのフィードバックをモデルに反映する仕組みを検討する。

2. 【研究 2】看護師等国家試験への CBT 実装の可能性の検討

2023 年度は国内の大学、専門学校の学生、教員を対象に、看護師等国家試験の CBT 化に対する準備状況や課題について調査を行った。

看護師等国家試験の CBT 化に対する教員の意見としては、概ね好意的であり、マルチメディアを活用した問題を使用できることでより適切に能力を評価できることや、試験を複数回実施できるようになることへの期待が多く示された。一方、試験の複数回実施の基盤となる問題プール作成のための過去問題非公開化に対しては不安や懸念が示された。現時点では日本における看護師等国家試験の対策には過去問題の演習が大きな役割を果たしており、CBT 化のメリットの享受の前提となる問題プールの構築に対する理解を得るとともに、過去問題の演習に替わる試験対策の方法を普及していく必要がある。

次に教育に関しては全体として、CBT 化さ

れたとしても出題基準等に大きな変化がなければ、現状行われている教育を大きく変更する必要性などの影響は少ないと考えられた。しかしながら、マルチメディアを活用した問題への対策にはシミュレーション教材や動画・音声を用いた教材を使用できる方が効果的な学習ができることから、教育機関の財政状況により、こうした教材の導入に差異がある場合には、一定の配慮・支援が必要である。

学生側もコロナ禍を経て推進されたオンライン教育、シミュレーション教育の結果、CBTの試験形式やマルチメディアを活用した問題に対応する基礎的能力は備わっていると考えられた。

本研究では、機縁法により対象者を募ったため、準備状況の分布についての代表性には限界があるものの、CBT化を進めるにあたって検討すべき事項については概ね抽出できたと考えられる。教育現場による準備状況については、どのような形式で試験が行われるかある程度具体的に提示されなければ正確な回答が難しいことから、今後具体的な実施方法の方向性ができ次第、本研究で示された検討事項等について量的な調査を行うことで、より正確に準備状況を把握できると考えられる。

E. 結論

1. 【研究1】ICT・AI技術を活用した看護師等国家試験問題作問システムの構築

看護師国家試験の過去10年分の必修問題(500問)の分析結果を統合した、「作問に向けた改善提案(ルールブック)」の作成と、3種類の大規模言語モデル(ChatGPT3.5、ChatGPT4、JSLM)による4つの作問システム(gpt4、gpt3.5-few shot learning、gpt3.5-Fine Tuning、JSLM)を用いた試験的作問を行い、専門家による問題評価と各モデルの生成性能

を評価するための自動評価指標の作成を試みた。その結果、258の改善提案が提示され、提案内容のさらなる分析の結果、200の小項目に対して147のルールが抽出された。また大規模言語モデルの活用可能性の検討の結果、gpt3.5-FTが他と比べて再現性、精度の評価指標値が良いことが示された。

2. 【研究2】看護師等国家試験へのCBT実装の可能性の検討

国内の大学、専門学校の学生、教員を対象に、看護師等国家試験のCBT化に対する準備状況や課題について調査を行った。看護師等国家試験のCBT化に対する教員の評価は概ね好意的であったが、複数受験が可能になることなどのCBT化のメリットを享受するために必要な問題プールの構築に関しては時間をかけて理解を得ていく必要があると考えられた。学生のCBTの受容性に関しては、コロナ禍を経て推進されたオンライン教育、シミュレーション教育の結果、CBTの試験形式やマルチメディアを活用した問題に対応する基礎的能力は備わっていると考えられた。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

Yusei Kido, Hiroaki Yamada, Takenobu Tokunaga, Rika Kimura, Yuriko Miura, Yumi Sakyo and Naoko Hayashi, “Automatic Question Generation for the Japanese National Nursing Examination using Large Language Model”, Proceedings of the 16th International Conference on Computer Supported Education - Volume 1: AIG; ISBN

978-989-758-697-2; ISSN 2184-5026,
SciTePress, pages 821-829.
DOI:10.5220/0012729200003693.

2. 学会発表

Yusei Kido, Hiroaki Yamada, Takenobu
Tokunaga, Rika Kimura, Yuriko Miura, Yumi
Sakyo and Naoko Hayashi, “Automatic
Question Generation for the Japanese
National Nursing Examination using Large
Language Models”, CSEDU 2024, Angers.

米倉佑貴, 林 直子, 木村理加, 佐居由美,
三浦友理子, 佐伯由香, 佐々木幾美, 宮本千
津子, 「看護師等国家試験の Computer Based
Testing(CBT)形式での実施可能性に関する調
査」, 第 44 回日本看護科学学会学術集会. 熊
本, 2024 年 12 月.

H. 知的財産権の出願・登録状況

該当なし

表 6 マルチメディアを活用した問題に対する感想(学生)

カテゴリ	サブカテゴリ	具体的例
文章問題との比較	状況は文章よりも理解しやすい	状況設定問題が文ではなくて動画の方がイメージしやすく、やりやすかった。
		字面で読むよりも何を問われているかが動画のほうが明確、伝わりやすかった。
	必要な情報を得るのは文章問題よりも時間がかかる	問題の内容によっては、文字だと大事なところだけを先読みしてポイントをつかんで速く読んで答えられるのが、動画だと最初から最後まできっちり見ないと答えが得られないので、時間が心配になる。 問題の難易度はそんなに難しくはないが、今までは紙だったら画像が4枚あってそこから選ぶから、ぱっと見てぱっと動画だったので、「おっと」って感じで巻き戻したりとかしてだったので、ちょっと時間はかかった。
試験対策、学習に関する影響	臨床に生かせる形で学習できる	国試の勉強を文字でしていると、国試にはいいが臨床にはいかせないなので、映像で見た方が実習の経験も役に立つので映像もあった方がいい。
		映像の方が実践ではいかせるが、人に伝える時に言葉で表さなければいけないので、紙で単語を覚える工程が大事だと思う。
		自分が働く現場を想定したら、コンピュータのほうが実際の現場に近い状態での音も聞けるし、そういうので分かりやすさでいったらコンピュータのほうがいいかなって思った。

表 7 回答形式についての感想(学生)

カテゴリ	サブカテゴリ	具体例
回答形式の違いに関して	転記が不要なので解答に集中できる	紙問題では解答を照らし合わせる時にミスしてないか不安になるが、CBT 試験の方がミスは防げると思う。 紙だと間違えて色塗ったりとか、塗ってなかったりとかがあると思う。コンピュータだと、記入漏れとか間違えて選ぶとかは、なくなっているのかなと思った。
	クリックで問題に飛べるので楽	スキップした問題を一覧で見えて戻ることができて簡単だった。
	メモ機能の必要性	問題に印をつけられる機能があればいいが、細かい機能が増えると使いこなせない人たちが出てくるであろうから難しいと思った。 紙のテストは印やメモを書きこむことができる点がいいが、CBT でもブックマークやメモの機能、戻れる機能があるか、手元にメモできる紙があればよい。 滴下計算とかあったら、ちょっと難しいかな。筆算したい。計算問題ができない。
	時間配分	時間制限ある中で動画の問題が続くと、(進みが遅く感じて)早くしてよって感じになる。 紙問題であれば最初に問題全体を見て計算問題はどこにあるか等を確認できるが、CBT ではどこに動画問題があるかわからないから時間配分がわかりにくい。
	試験形式の説明	動画は 1 回しか見られないと思った。リスニング試験のように仕組み(2 回読むなど)の説明がないと不安だ。試験をやりながらこれできるのだとわかってきた感じだった。 呼吸音の問題: 聞き取りにくかったが、PC 自体の音量が小さかったのだと思う。英語のリスニング試験のように、テスト前に設定の方法を伝えたり、チュートリアルとしてお試し問題や動画があったりした方が安心だ。
視認性について	紙の方が読みやすい	紙の方が慣れているため文が読みやすい。
	画面を見ることによる疲労	一定時間パソコンをずっと見るのがつらいので、紙の方が解きやすい。 長時間になると目がチカチカしたり、頭や背中が痛くなったりすると思う。 状況設定問題の長文がいっぱいある問題とかだったら、一遍に集中するから目疲れるかなっていうのはちょっと感じた。
マルチメディアについて	動画・音声の視聴形式	動画を倍速で進めることができれば助かる。 動画の解答の選択肢が 1 から 4 まで 1 本の動画なので、1 つだけ見返したい時には全部見るか、シークバーをこの辺りかと検討をつけて動かさないといけな。選択肢の動画が分かれていたらもっと見やすいと思う。 動画は最低限 2 回は再生できるようにしてほしい。再生回数の上限がなければありがたい。
	イヤホンについて	イヤホンを長時間つけると気持ち悪くなるので考えていただきたい。 イヤホンは耳に合うものと合わないものがある。自分は柔らかいタイプが合わないので、選べるといい。

表 8 CBT 形式になった場合の学習・試験対策に有用・必要だと思うもの(学生)

カテゴリ	サブカテゴリ	具体例
現状行われているもの	シミュレータ・マルチメディア教材による学習	急性期の実習時にシミュレータの人形で脈を測ったりモニターを取ったりしたこと。
		基礎のフィジカルアセスメントのホームページで音を聞いたこと。
	演習・実習	基礎看護の授業の時に見たり、自己学習したりした丸善の EVO の動画に似ていた。
		手袋は基礎看護で、BLS は授業で、小児の授業では実際に赤ちゃんと大人に練習した。
追加で必要だと思うもの		何度も練習した手技は流れを覚えているので、このような試験の問題を解くスピードに関わってくると思う。
		心電図のモニターを触って実際に動く波形を読むこと。
		今回わからなかった問題もマネキンや自分の体を使いながら教えてもらったら、答えられる問題が多いと思う。
		心臓の音とかがって、やっぱり授業で何回か聞いたぐらいしか印象がなかったから、腸蠕動とかもそうだが、あの辺は結構授業の中で繰り返し聴き慣れとかないと、しんどいところがあるかなと思った。

表 9 CBT 化の利点・欠点についての意見のまとめ(教員)

カテゴリ	サブカテゴリ	具体的なコメント
CBT 化の メリット	マルチメディア 問題が使用でき る	紙で書いてあるものは、情報がすごい限られてる(見えない、聞こえない)が、雰囲気や表情も含めたビデオ教材から設問にすることは、学生さんたちはそれもアセスメントすることができる。文章で書くことは、それを、もうすでに出題者が書いてしまっているか書かないかというところが出てきてしまうから、より学生さんの能力をアセスメントする意味では、情報が豊かな分だけ、良いアセスメントが試験の目的として、できるのではないかと思う。
		いわゆる文章だけで知識を問うのは非常に難しくなっていることと、あとは学生の、どこまで理解しているかは、動画、例えばサンプルの滅菌手袋の問題があったが、あれは非常にいいなって思った。どうしてもやっぱ画像だけだと細かいところ伝え切れないが、動画だと「あ、ここが違うな」というのが、ほんとに知識が試される問題なので非常にいいなって思った。
	複数回実施が 可能になる	一番のメリットは、年 1 回ではなくなることが一番のメリットと思う。コロナの時に、もしかしたら試験自体ができないかもしれないと思った時に、何回かに分けてできるとか。ただ、そうなった時に、結局、CAT ではないんですよね。CBT。計算はしないで、難易度は計算するにしても、そうなった時に、じゃあ、複数やるなら、問題セットの難易度をどう調整しておくのかとか、そういうことは考えなきゃいけないけれど。
		将来的には、やっぱりアメリカみたいに何個か割と試験会場を多くして、その中で年に何回かやれるみたいなことをやるほうが、人の確保とか、1 回の受験で落ちちゃった人の再チャレンジとかにも効果的かなと思うので、コンピュータを利用した試験になることには賛成をしている。
CBT 化の デメリット		あとは多分これが稼働していくと、一回免許取った人たちの更新にもつながっていくのかなって思って。臨床の数値ってほんとに忙しい中でいろんな学習会開いたりとかしているが、基本的な知識に立ち戻ることもまた大事かなって思って。そういう意味では、更新制度っていうのも日本でどの程度進むか分からないが、そういうものにつながっていけば、看護師全体としての自己研鑽とかレベルが上がっていくことにつながっていくのかなって思っている。
	採点等の利便 性	基本的には紙ベースじゃないほうがいいというのは、その集計とか採点とかの利便性もあります
	マルチメディア 問題による情報 量の増大の試 験への影響	一方で、映像みたいなものから、たくさんの情報が与えられてくるわけで、それをどう捉えるかということが、均一・平等にできるのかということに関しては、難しいのかなと思う。
	作問負担の増 大	作問する側としては大変と思う。適切で答えがばらつかないような動画を作れるんだろうとか、一年一年で問題を作るわけですけど、その期間の中で動画を作らなきゃいけないのは、ほんと大変だなと思う。
		問題作りも大量にいいのを作らないと。全国の教員全員に 1 人 10 問とか、20 問とか、何でもいいから作って出せって。マストでやってくださいぐらいじゃないと無理かもしれない。
	公平性の担保 が難しい	看護師はやっぱり医師よりも N が多いということと、やはり看護基礎教育ですごく幅がいろんな大学から専門学校からいろんなルートで看護師になるところから考えると、看護基礎教育の期間もやっぱりインフラ・設備とかがさまざまある中で、一定の全ての受験生の公平性を保った準備を、インフラの準備ができるかなという点でちょっと疑念があって。
	インフラの整備 が難しい	看護師はやっぱり医師よりも N が多いということと、やはり看護基礎教育ですごく幅がいろんな大学から専門学校からいろんなルートで看護師になるところから考えると、看護基礎教育の期間もやっぱりインフラ・設備とかがさまざまある中で、一定の全ての受験生の公平性を保った準備を、インフラの準備ができるかなという点でちょっと疑念があって。

何か現実的に各一人受験生に PC をそろえてするっていうのが現実的に何ていうか、予算的なところであるとか、そういうのって設定は可能なのかはちょっと疑問はある。

指定された会場に集合して備え付けのディスプレイうんぬんというところでは、相当数の機材が必要になると思われて、その設置にかかる準備とか、配線とか、会場の配置とか確保、その辺りはどうなんだろうという、すごく懸念されると思った。

実施費用が高額になるのではないか

受験料金がこれをやるとしたら、お金がかかってきた時に、どのくらいになるのかなっていうのは、ちょっと懸念する。

漏洩等のセキュリティ

欠点はシステム上の問題で、どこまでいっても紙でも一緒だが、やっぱりセキュリティの問題とか、デジタルデータになるので一度流出すると分からない部分があること。一般的なデジタルのリスクは、ちょっとどれだけ仕組みでカバーできるかっていうところは課題かと思った。

紙でできることの再現に関する問題

今現在ペーパーで試験を受けている学生が、思考の整理をしながら線を引いたりとかして自分がどういう解答を導き出すかというのをマークシート以外に問題用紙に書いているので、そういう作業をパソコン上でできないのでメモ用紙とか持ち込みが可なのか、そういう問題をただ単にチェックするというのに慣れてない学生なので、現状は難しいかなと。

表 10 マルチメディアを利用した問題に対する評価(教員)

カテゴリ	サブカテゴリ	具体的なコメント
マルチメディアを利用した問題について	難易度に対する評価	現在の国家試験の難易度と比較して難しくなと思う。あの波形は見せてないですよね、心電図の。だから傾向と対策がちょっと変わるかも。音はない。
		難易度的には、そんなにものすごい高いっていうわけでもないと思って。逆に動画になったりとかしてるので、正確な形で問題としても作れますし、というのはとても感じた。
		動画に関してはやっぱり先ほども言ったように視点をどこに置いてるかというところで変わってくるかなと思うので、難易度はやっぱり高いのかなとは。看護の経験というか、そういった場面というものをやっぱり学生が知らないというか、映像とかそういったもので学習はしていくと思うんですけど、そういった経験というのがやっぱりないので難しいのかなっていうのは思います。
能力を評価する試験としての適切さ		動画とか音声とか、あと普通の4択の問題があったと思うが、難易度自体は簡単だったかなと思っているが、動画のほうは先ほども話があったが、着目するところがいっぱいあって、一体自分は何を見なきゃいけないのか、もう一回見てみようという、その時間がかかるので、解答に時間がかかって、もしかしたら終わるまでの時間がかかる、足りないことにもなるんじゃないかなと思った。
		多角的に判断するのは、今まで紙、文章で出せない問題をちゃんと出せて、いいと思った。国家試験としては今よりふさわしい。
		より実践に強い人が合格しやすくなるとして、すごく面白いと思いました。
解答時間・問題数		今の国家試験問題を踏まえて作成された動画を使った問題としては、すごく適していると思う。
		やはり状況判断とか、それから視覚的な情報から得たものの作問というのは、やはり文章では出せなかった問題だと思いますので、臨床実践により近いといえますか、実践内容を問うこともできるので、紙のベースよりも問う能力は広い範囲で問えるのではないかと思います。
		今みたいに1日やってたら大変かも。 今現在も、国家試験免許を与えるにはこれだけの知識量を取るという、そこは割と検証されて、あの時間の作問になったのか。全然関わったことがないので、やっぱり領域ごとに何問っていう設定で、それが結局あぁなってるのかなとは思いますが、そういう最適な時間というか、最も経済的な時間とか問題数とかっていうのが、コンピュータ化していくことによって進むのかなと思うとなるといいなと、でも、確かに、ずっと見るのは大変かもしれない。
動画の再生回数		動画に関しては流れていってしまうので、あれはもう一回見られるようになっているのか。何度でも見られるのであればいいかと思った。流れていくと、自分が注目すべき場所がどこなのかっていうのを、一瞬分かんなくなる人もいるから。
視覚・聴覚の多様性について		色覚多様性の人って、カラー問題がいくつかありましたよね、確か。あれ、どうしてるんだろうなと思う。マルチメディアが使えるようになると、音も聞こえにくい人の問題が出てくるから、音と色が見えない人へ、見え方が違う人への配慮をどうするかっていうのを考えなきゃいけないと思う。
外国人受験者への対応について		目が見えないんぬんでやると、外国の人が確かに難しいだろうなと、会話形式で出題した時なんかは、字幕付けるのか、でもこれぐらいは分かる人でないといけないうってやるレベルの会話だけにするのかという、そのあたりやっぱり合格基準の設定が改めて難しいと思った。 全部、平仮名振る感じなので、上に字幕を付けるっていう配慮をするのかしないのかというあたり。

マルチメディア問題作成時の課題	モデルのプライバシー	子どもだと倫理的なところとかもありますし、今のテストだと目とかも全部写っている。通常消したりしてるはずなので、そのあたりどういうふうにするのかなって、SPさん、模擬患者さんのプロフェッショナルみたいな人がやってくれると上手かもしれない。
誤答肢の動画等を作成する労力がかかる	今、テキストには該当の動画いくらでもあるんだけど、あと3問、間違った場面を作らなきゃいけないのはすごく大変。だから動画があるから問題作りやすいんじゃないの？というのはちょっと違うと思う。素材はある。正しい素材はあるけど。誤った素材を準備するのが大かな。	
動画等の解釈の多様性に伴う問題	状況設定問題、動画でやろうかと思うと非常に難しいだろうなとか思って。細部まで気をつけないといけないから、危険だろうなとか思って。	
	紙で文章で作った問題よりも、動画のほうがいろんなふうに読めるから、より答えの幅が広がって、終わった後に、疑義がつきやすくなる懸念がある。そんなことが起こらないように、事前にしっかりと動画で求めることとか、こういう問題の場合は、こういうことを意図してるっていうことを、しっかりと受験生も、看護教育機関も、準備できるような情報提供が必要	

表 11 問題プール作成のための過去問題非公開化に対する意見(教員)

カテゴリ	具体的なコメント
学生の不安	過去問が非公開になることに対する学生さんの不安とかは、きっとあると思う。
予備校等による対策	結局は公開しないとなったら、多分学生が分担して記憶してみんなで持ち帰って、大学なり学校が教えるっていうのは、やると思う。 受ける人たちのための国家試験対策の観点としては、恐らく業者が頑張って、お金をかけてくと思うので、その辺でカバーはできるのかなと思う。
国際的には非公開が一般的	テスト学会系の先生たちとずっと話をしてて、そもそも日本が公開してることのほうがおかしいんだと、ずっと聞いてたので非公開で問題ないと思う。
過去問対策も学習の一部である	落とすための試験じゃなくて勉強させるための試験だからやはり過去問から何度も解いて復習して知識を蓄えることを、今は国家試験受験するために受験生はみんなやってるのを非公開にしたそれはできなくなるっていうことはやはりちょっと本来の趣旨からは外れて本末転倒かなと思う。 現在の国家試験対策は基本的には過去問での対策がやっぱりメインになっている。過去問がないとなると新たに国家試験対策の方法を考え直さないといけない状況にはなるかなと思う。
自己採点ができなくなる	自己採点をした結果を学校内で分析をかけていってどこの領域ができてないとか、その得点率とかを見ていって翌年に生かしていくんですけど、今回非公開になった場合に、合否だけではなく日本全国的にどう問題が得点率が悪いよとかいうのが出てきたらありがたいなって。 学生の自己採点とか振り返りが結果が出ないと分からないという形になると、それも手元に何も残らない状況で学生も不安になるのかなと思うので、試験を受ける時のことと、あと分からないというところで、どう問題に対応する力を教育していくのかというところが気になる点になります。
移行時のサポートの必要性	一番問題になるのは、試験の移行期の最初の学年だと思う。その人たちはすごい不公平だって、きっと感じる。2 年目は業者から過去問出る。切り替えに当たる人、あるいは前の学年で落ちそうで来年トライする人とか、要るじゃないですか。どの試験や制度でも変わる時に切り替えの人は移行措置を考えなきゃいけないのかもしれない。
問題の質保証の体制	ここで言ってるような質の保証っていう、第三者の目をどう入れるっていうあたりの仕掛けはやっぱりないと駄目だろうなと。あれだけ中でよくたいて出したのに、不適切問題が出てくるって、その仕組みをどう整えるかは必要だと思う
試験対策のために必要な情報	もし仮に非公開になった場合にどんな準備が必要かとなると、今でも出題基準は出てるので、この範囲から出題されますということはある程度、明示されているが、その具体的な詳細の内容は公開されてない。提示されてないのでそれをまるっと提示する必要がある。本 1 冊ぐらい書いて、この中から絶対出ますというくらいしないと、過去問を公開しないことに国家試験出題側の耐えることができないかと思うので、それぐらいしないと駄目かな。 そういうの(公式問題集)がいいと思います。恐らく国家試験の作問委員会、しかも厚労省とかが公式問題集みたいのを出して、それに準拠したものとかをもしかすると予備校さんとかもまねをして、そこから派生させてこういう予想問題みたいなを出したりとかっていうのはあり得るかもしれない。 いずれにしてもどんな問題が出るかを公式に出す必要がある。現在の出題基準よりももっと丁寧に出さないと準備できないと思う。そこら辺は多分他の既に非公開を実施している多分、試験とかでやっていることをやる必要はあるっていうところ。 アメリカとか非公開の CAT だけど、できているのは、検査センターが、「こういう問題って出せますよ」というのを公開、サンプル問題の公開はしているので、そういうふうにおけば。 試験問題の公開がないことで、どういうものが多く出題されてるかといった全体の傾向とか、難易度に関してのある程度の概要の情報公開はしてほしい 練習問題だったり予想問題だったり、あと模擬試験という形で何でもいいが、練習することができるようになっていると、過去問が公開されてない不安っていうのが学生も少なくて、準備もしやすいのではと思う。

表 12 シミュレーション教材等の活用状況(教員)

カテゴリ	具体的なコメント
LMS	学部では manaba で実習レポートの提出し、授業の資料を manaba に載せている。 ウェブ教材とか提供しているのと、コロナ禍で病院に行けなかったときは看護学実習、私は基礎看護学なんですけど、看護学の実習を学内で模擬患者さんのシミュレーション情報を作って、それをウェブを介して学生が参加して、それで実習をするというような授業の組み立てをした。 授業の案内は全て PC を通してとか、LMS を通して授業の案内。それから、成人看護 I では授業ごとに事前に LMS を通して授業資料を提供し、1 授業ごとに小テスト行って、確認テストと言われるものを行って。それも評価の中に入れるので、なしでは今はできない状況になっていると思う。
グループウェア	本校では Microsoft365、これを全員にアカウント配布してるので、全てのデータは学生と共有する、全てとは言わないけど、できる。なので、365 を使った授業とか、Teams で課題の提出や返却、授業資料の配布とか、Forms で授業アンケートの実施だとか、あとは授業前に「この動画見といてね」って学生に流して、それをもって授業に臨むみたいな事前学習的なことに使ったり
ビデオ会議	海外の人と直接会えない講義の時は Zoom
シミュレータ	人形を使って、あたかも人間のように技術を学ぶってことはしています。時々にはコンテキストとか、その経過とかを設定して、それに合わせた看護の提供の仕方、看護という技術になるが、主には、これに合わせた技術の提供の仕方を検討し、それをやることを試すこともあります。 シミュレーションの教材、看護系の教材を利用して、多機能の人体モデルがしゃべったり、瞳孔反応が出たりして、心電図波形出たりっていう高機能シミュレータと、あと胸の音、胸部の異常音と肺、呼吸音の異常音とか聞けるシミュレータとかがあるので、そういうシミュレータを用いた演習とかを授業で計画することもある。それをウェブ経由で聞くこともある。 コロナ中の実習に行けなかった時に、エルゼビアの vSim というのを買わせてもらって、ロールプレイングゲームみたいになって、CAT ではないけれど、「今、患者さんの目の前にいました。何とかが起きてます。どうしますか」、ピコって選んで、最終的にすごくできると 100 点満点で、スコアが付くみたいな。あと、患者さんが「痛い」ってしゃべりだしたり。あまり時間がかかると「早くして」って怒ったりっていうようなものは使ってみました。 (使ったのは)4 年生の総合実習。たったの 4 人とかのためにすごい高かった。だから、もう使えないってなったけれど。最初のころは、エルゼビアさんが、すごい使ってほしいから「どうぞ、無料で」みたいな感じで使わせてくれて。でもあまりにも高かった。 成人看護学ではこのコロナ禍になってから業者さんから SCENARIO をお借りしてた時期があったが、どうしても実習に行けなかったのもその SCENARIO を使って演習をしたり、本物の患者さんに見立てた事例として使ってた時期はあったが、学校のものではなくてリースしてもうお返ししたので今現在そういったものはない。 シミュレーション教材を使って、僕、クリティカルケアの授業も担当しているんですけども、いわゆる SCENARIO 君って商品名の人形があって呼吸音聞けたりとか、不整脈ってやると、モニターがあってモニターにその不整脈 Vt みたいなのが出たりっていう、血圧も数字として出たりっていうのを使っている。 なので使い方としては、まず学生に普段の事例を提示して、そこで「この患者さん、何観察するか考えときなさい」ってまず提示する。で、実際に SCENARIO 君が「人形が置いてある部屋に行って観察をしなさい」って。ちょっと事例、例えば不整脈になってたりとか、そういった仕掛けがあるが、それを用いて「じゃ、何観察した？」って。「何か普段と違ってたね。じゃ、それ何でかね。じゃ、どうということアセスメントすればいいかな。何観察すればいいかな」っていうのを考えさせる授業をする。おおむね学生にはすごく好評で、やっぱり紙面の事例で展開するよりも、そういったものがあつたほうがすごく分かりやすいっていうのは、学生の声として上がっている。
模擬(電子)カルテ	実際に救急カルテを使って、そこからどうやって情報を取ったらいいのかってことを学生自らちょっと学んでもらうように、そういった事例を作って、その事例をカルテに落とし込んでさらに。
動画教材	高齢者の状況であるとか、あとは認知機能が低下した人って、どういう状況なのかとかをやっぱり動画を使ったりとか、その動画を見てこういう状況になるんだっていうのを理解してもらったりというので、そういったメディアのものを使ったり。

学部定期試験 以外での CBT 利用	大学院では使用している 小テストに使用
コロナ流行時に 定期テストで利 用	テスト起動で。時間を決めて、よいスタートで。見るなよって言っても見るかもしれなかったので結構問題量を多くして、見るといってもある程度調べないとできないような問題とかすごいいっぱい作った。ちゃんとばらついた。全員が多分見れる環境にいたと思うが、パソコンの前に座ってたので、見る、アクセス時間もちゃんと見て、2022 年ですかね、コロナ始まったときにやった。 単純に manaba で選んでもらうってだけのやり取り。その時は、もう一斉にアクセスしてもらって、時間決めてというやり方。 その時、ちょっとやりづらいとか、ほんとに公平にできてるのかとか、カンニングとか採点とか、どうすんだとか。カンニングをしようがないって思って、何か見ていいことにはして、見れば答えられるって問題ではなくしてみたり。 コロナ禍でも確かに試験はした。終講試験を PC で行ったが、これについてはやはり今すぐ問題があるといったらおかしいが、まだ今課題が多いと思う。試験についてはやはりコロナ禍においては学生の善意というか、良心を信じて試験は行ったが、各部屋、自らの責任においてやってもらったが、そこは本当に公平だったかどうかの確認は取れないという意味においては、だいぶ課題はあったと思う。 学生は家でやった。その時間に時間は合わせたけれども、LINE しながらでもできるから。そういう環境だったので少し課題は多いと思う。集合型で国試でやられてるようなスタイルで考えるのであれば、それは問題ないと思う。

表 13 CBT に対応した教育の準備状況、移行に必要な期間についての意見(教員)

カテゴリ	具体的なコメント
現状の教育で対応できる	国際看護で出題するものが、すぐこの方式かと言われると、逆に、あまりないような気もする。例えば、今、テレビでやってる SDGs で、「SDGs は何個ゴールがありますか」とか。「前の MDGs とどう違いますか」とか。見せる問題、あんまりないと思った。 教育そのものが、すぐこれによって変わんなきゃいけないのかっていうのは、ちょっとイメージがあんまり、すいません、各論、今、担当してないので。
マルチメディア問題への対応	動画は授業でいっぱい使う必要があるだろう。対策を考えるのであれば、国家試験の一部が動画になるとすれば練習問題として授業の中にいっぱい入れてったほうがいい。(No.3) カリキュラム組めないからそれこそ動画、何パターンも模擬試験で準備してもらってそれもひたすらやるから。あとは動画で出る出題の範囲をある程度細かく出してもらおうとした上で、教員が想定動画を作っておくか。そういう問題を売られるかもしれないが想定動画を作って準備しておくか。
学校・学生の経済的格差についての懸念	やっぱり明らかに動画を見る機会も少ないので、そういった動画を見る、見て判断する機会をつくるために、そういった動画ってすごい高いので、なかなかうちも手が出ない。なので、もう少し安価なもので、小規模の看護学校とかでも購入できるような、今でいう状況設定を動画化したものとかそういうものがあると、非常に実習の事前学習にも使えるかなと思うが、そういったものがあることで、もうちょっと読み取る力は上がっていくのかなと思うので、そういう動画類とか。 今の学生の状況とか学校の経済状況とか、そういうところで懸念があるなどは考えていた。コンピュータを利用した試験の練習が例えば経済的に裕福な家庭とか、学校によって、事前に練習ができる、できないにばらつきができてしまうんじゃないかというのはある。特にうちの学校、市立の学校なので経済的に余裕がない家庭もあって、家にパソコンがなかったり、そういう子とかのことを考えるとちょっと。 どういふに国家試験対策させればいいのか、それこそ経済の格差が生まれるんじゃないかなと思う。経済格差っていうのは、結局予備校に通ってる人が強いみたいな。予備校はもう仕事なんだし、僕らも仕事なんですけど、予備校はもう全力で対策してきて、それなりの動画も準備してくると思う、きつこうなった場合は。 けれども、そういった通うお金もないっていう子たちは、じゃ、何見て勉強すればいいのかっていうのが路頭に迷っちゃうっていうふうに思う。
CBT システムの操作のトレーニング	音声問題だったらここで、動画問題だったらここみたいな、操作のことは一定の慣れが必要なのかもと思って、音を聞く時、どこを押すんだっけって感じで迷ったりしたのがあったので、画面の操作のやり方は、慣れがちょっと必要かなと思う。 パソコンを使った国試の問題ってなると、今回初めて問題を解かせてもらったが、選択肢の選び方だったり、次の問題に行き方だったり、そんなに難しくない画面ではあったが、国試の本番にこれを見たら、学生困るかなと思った。なので、国試の問い方に合わせた問題の答え方の練習とか、パソコンの操作の仕方を事前に練習しておいて緊張度を下げたほうがいいと思った。
不足している内容準備期間	それは、もう圧倒的に演習の時間が足りないなと思っている。やっぱり知識ベースというか、出題範囲。国試の授業を組み立てる上でも、国試の範囲っていうのは一応意識はしますんで、それらもカバーしようと考えたら、基本的な知識や看護を行うために必要な知識の部分というものの確認する時間が、授業時間の中でも結構占める割合が高いので。やはりこの判断力を問うような問題を解けるようにしていくためには、さらにもっと演習時間を増やすっていう必然性はあるんじゃないかと思っています。 ああいう質問、もちろん臨床に行くと恐らく求められる力なんだなって思うので、もう少し実習の中で、すごい大変だが、今は看護課程中心にやっている、ほんとにその場その場での判断とか状況を読むっていうところで、今いわれてる臨床判断とか推論とかをもうちょっと実習の中で強化していかないと、あれには対応できないなって思ったのと、そこに、実習にそういうものを入れるのであれば、学内での講義とかももう少しいろんな知識をつなぐような、もちろん今もやってはいるが、なかなか成果が出てなくてあれなんですけど、例えば解剖とかの知識、病態とかの知識をつないで、看護とつなぐところを、授業の中でももう少し意図的にやっていかないと、実習での臨床判断とか推論とかの実習には持っていけないんだらうなと思うので、全体的にもう一回カリキュラムを見直さなきゃいけないのかなと思った。

その場面を見て状況をつかむ能力とか、あとこれがどうなるんだろうって予測するような能力っていうのは、やっぱり教育内容としては必要なのかなと思うが、現状それはあまりできていない。

国試がコンピュータを使ったものになる場合、学生の国家試験対策の教育面での準備は、3年あればいい。

学びのほうに関しては、やっぱり2年か3年あるほうが良いのかなって、学部が入って、せいぜい2年生ぐらいにはそういうことになるんだっていうのが分かりつつ、準備をするぐらいの期間があるほうがいいのかと思う。

ただ、もちろん来年なると言っても、やればするんじゃないかとは思いますが、でもちょっと乱暴な感じ。あと、機器を使うほうの準備は、やっぱり1年くらいあればいいのかなと。専門学校とかはそういう練習するかもしれない。

入学のときから必要かと思うので4年間ぐらいは必要と思う。大学1年に入学したときから今うちの学校でも国家試験を意識させて4年生の2月に受ける国家試験の合格率が上がるようにとやっているの、それを考えるとやっぱり4年前かなと思う。2年生になってから3年後、CBTだと言って言われても、ええーみたいな感じなので。やっぱり入学時からあなたたちはCBTですっていうことを分かってたほうが良いと思うと、その子たちが入学する前から教員は準備しなきゃいけないとなると、学生としての準備期間は4年間一緒だと思います。

教員としては5年前ぐらいじゃないと。新入学生の国試対策を準備するには4年かかと思うと5年ぐらい前から告知されないと準備できない。

3年かな。と思ったのは入学した学生が、「あなたが卒業する時は、CBTなんです」というのを知って、テストしてもらいたいから。少なくとも3で。5年ぐらいはかかるのかなって思う。ほんとに、何の根拠もなく。

それこそ、パソコンに慣れるみたいなどころから始まることになったら、「いい、普段から、パソコンで、もの書くのよ」とかから始めなきゃいけない人たちがいる。

結局、彼ら(国試向けに)1年しか勉強しないので。問う知識は(紙とCBTで)同じですよ。だからといって、例えば、今までだったらシナリオだけでずっとやってたものが、画像になる。だから、演習も全部できるだけ画像でやってとか、だから4年必要とか、そういうことはあんまり思わない。少なくとも試験に出るようなことは、多分全て教材、電子教材になってるので、それを使って勉強するっていう感じになると思う。1年前だとあれかもしれないんで2年前ですかね。2年前にアナウンスして、学生たちは1年前から心して試験対策するという、そんな感じかなと思います。

3年間で国家試験を受けるようになるので、入学時からこういう試験になってるということで3年間準備をすればということかなと思う。

授業計画とか計画を立てるための1年前になると、その受ける学年の在籍数プラス1年は必要なんじゃないのかなと思います。そうしないと、学生にきちんとした教育プログラムで学習するというのは難しくなるんじゃないのかなと思います。
