

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 年計画で進行する。

初年度の 2022 年度は、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、今後看護の国家試験に関するデータを強化したデータベースを使用した作問システム開発への示唆を得た（研究 1）。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・作成と、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。試験運用に向けた問題として 500 問収集し、動画や音声等のマルチメディアを活用した問題の作成にも着手した。インタビュー調査については、年度末に研究倫理審査委員会の承認を得て、次年度から調査開始の運びとなった。

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 准教授

研究協力者氏名 所属 職位

伊藤 圭 大学入試センター 准教授
西崎 祐史 順天堂大学 先任准教授
宮本 千津子 東京医療保健大学 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

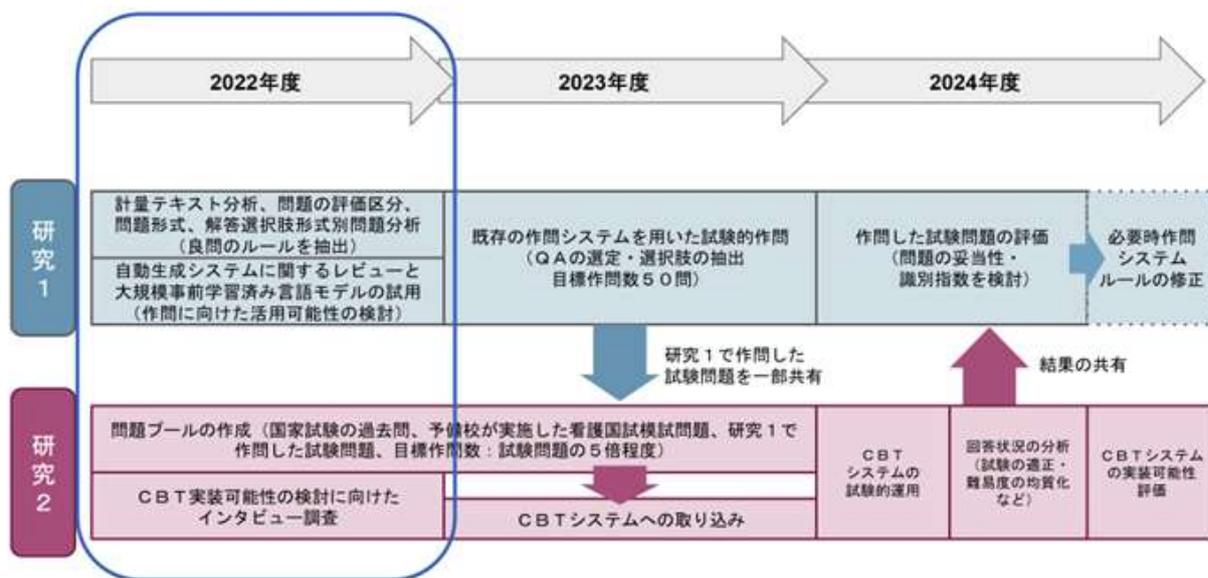
本研究は、保健師助産師看護師国家試験（以下「看護師等国家試験」）において、ICT・AI技術等を活用した具体的な作問システムを検討すること(研究1)、また看護師等国家試験におけるコンピュータを活用した試験（CBT：Computer-based Testing）の実装に向けて、CBTシステムの試用版を開発・試行し、試験問題の妥当性について難易度、識別指数、IRTスコア等から評価すると共に、出題形式、問題管理システム、受験者側の受容性に関する調査を行いCBT導入に向けた課題を明示すること(研究2)を目的とする。

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4年度は、研究1、研究2ともに図1に示す初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

担当者：林(研究代表者)、徳永、佐伯、佐々木、佐居(以上研究分担者 3 名)、研究協力者 5 名

1) 計量テキスト分析を用いた問題分析

(1) 分析対象

過去 10 年分の看護師国家試験必修問題
(問題文と選択肢)

(2) 方法

先行研究（令和元年度厚労科研林班）で示された基準に則り、正答率、識別指数等から過去問題を「良問」と「要改善問題」に分類し、問題文と選択肢をデータとして計量テキスト分析を行い、共起ネットワーク、抽出語を比較し、「良問」に特徴的なルールを試みた。分析には KH Coder 等を用いた。

2) 良問と要改善問題の評価基準の作成と良問ルールの抽出、ルールブックの開発

(1) 分析対象

1) に同じ（国試必修問題の過去問 10 年分）

(2) 方法

全 500 問の正答率、識別指数を元に良問と要改善問題の評価基準を再設定した。新たな基準により分類された結果を元に、出題基準の小項目ごとに良問と要改善問題の比較を行い、良問の特徴（良問ルール）を抽出することで、作問の手引となるルールブックの作成を開始した。なお、各問題の大項目・中項目・小項目の区分については、国家試験に関わる書籍情報に基づき、研究者らで振り分けた。

3) 既存の大規模事前学習済み言語モデルの活用可能性の検討

多肢選択肢問題の作成には図 2 に示すプロセスがあり、QA（Question Answering）システム¹⁾は、Sentence Selection、Key Selection、Question Formation に該当し、AIG（Automatic Item Generation）²⁾³⁾⁴⁾は、より広範に AI を活用したシステムと考えられる。そこで、QA、AIG を含め、既存の大規模事前学習言語モデルの適用可能性を検討した。

図 2：問題作成システムのフロー（例）

*文献 5) 図 3 より改編引用：多肢選択問題(MCQ)自動生成システムの一般的なワークフロー



4)既存の作問システムを用いた試験的作問

QA システム、AIG の他、既存の作問システムを検討し、本課題に適したシステムを選定し作問を試みた。作問システムとして、2022 年 12 月に一般公開された、ChatGPT を使用し、様々なパターンのプロンプト（指示文）を設定して作問を試みた。

2. 【研究 2】看護師等国家試験への CBT 実装の可能性の検討

担当者：米倉、宇佐美、佐伯、佐々木、佐居（研究分担者 5 名）、研究協力者 5 名

1)CBT システムに取り込む問題プールの作成

研究分担者・協力者をコアメンバーとして、作問チームを結成した。国家試験等の過去問、看護系大学共用試験 CBT の実用化と教育カリキュラムに関する先行研究も参照し、画像や音声も積極的に取り入れて問題プールを構築することとした。問題数は、実地出題数の 5 倍程度(250 問)とした。

2)CBT 実装可能性に関するインタビュー調査

(1)調査対象

看護学生と教員（看護系大学、専門学校）、各 20 名程度。

(2)調査方法

教育でのコンピュータ活用状況、CBT 実施状況、国家試験の CBT 化に対する意向、CBT 導入への準備等を半構造化インタビューで問い、結果に対して質的内容分析を行うこととした。調査にあたり、研究者所属機関の研究倫理審査委員会の承認を得た（承認番号：22 - AC103）。

3)CBT システムの検討と問題の取り込み

CBT システムとして既存のプラットフォーム（TAO、Classi 等）を検討し、決定後問題をシステムに取り込むこととした。

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

(1)計量テキスト分析を用いた問題分析

まず看護師国家試験過去 10 年分の必修問題 500 問のデータ（問題文、解答選択肢、正答、出題基準における該当する大項目、正答率、識別指数）を入手し、正答率・識別指数から、先行研究（令和元年度厚労科研林班報告書）の評価基準に則って良問等の評価を行った（表 1）。また、各問題の問題形式（what・how・why）・解答形式（名詞句・数値・文・図式）の分類基準を定め、500 問を基準に基づき分析した。

次に、問題文と解答選択肢のテキストデータを対象に、評価別（○、①、②、③、△の 5 分類）に計量テキスト分析（KH Coder を使用）を行った。

問題の評価ごとに共起ネットワークを作成し、語の出現度、語同士のつながりを図式化した（図 3～図 7）。円の大きさは語の出現頻度を示す。

分析の結果、評価ごと（5 分類）の問題数が限られること、また出題基準の項目ごとの問題数としては少数となることから、共起語の種類と出現数が限定される結果となり、良問（○評価）とそれ以外の問題（①、②、③、△）との相違の傾向はつかめなかった。

表 1 先行研究の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	90~96.15	0.2 以上	81
①	易しすぎる問題	99 以上	0.1 以下	69
②	良問だが難問	90 未満	0.22 以上	122
③	あまり適切ではない問題 識別指数 0 以下を含む	90 未満	0.15 以下	19
△	非該当問題			209
合計				500

図 3 共起ネットワーク：良問（○）81 問

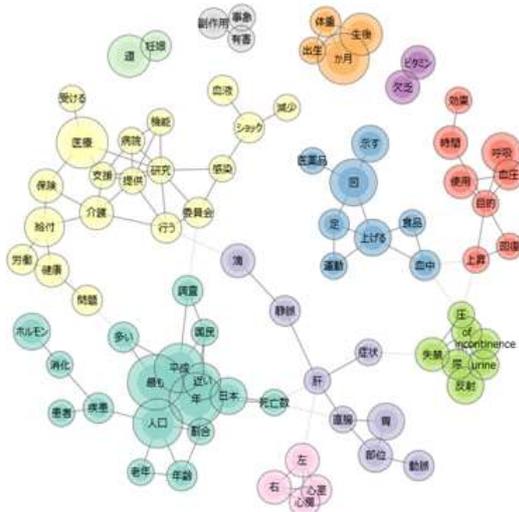


図 5 共起ネットワーク：良問だが難問(②) 122 問

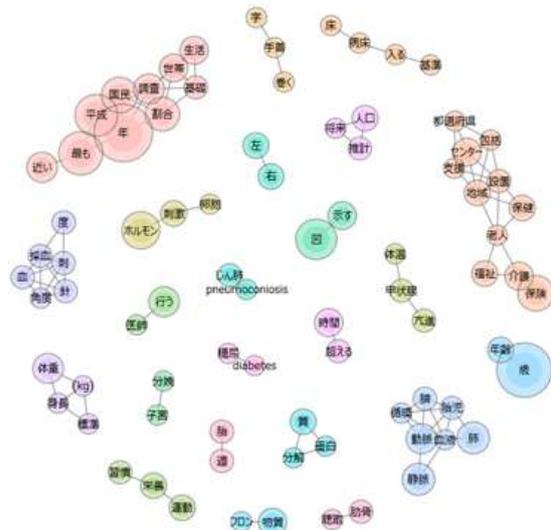


図 4 共起ネットワーク：易しすぎる問題 (①) 69 問

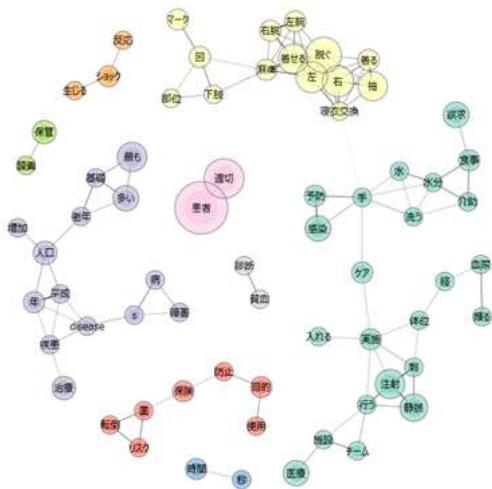


図 6 共起ネットワーク：あまり適切でない問題 (③) 19 問

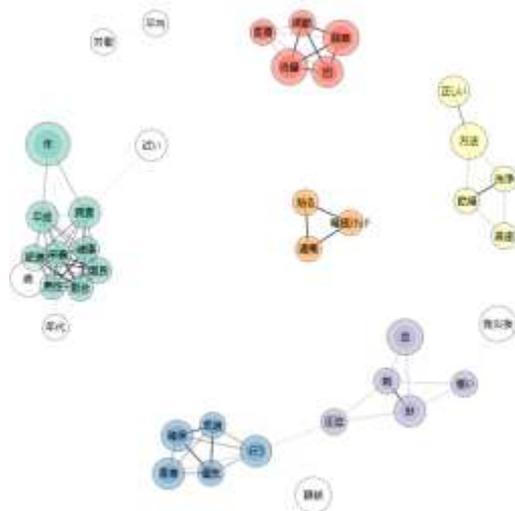


図8 過去問題 500 問の正答率・識別指数プロット (○、①~③ の詳細は表 1 参照)

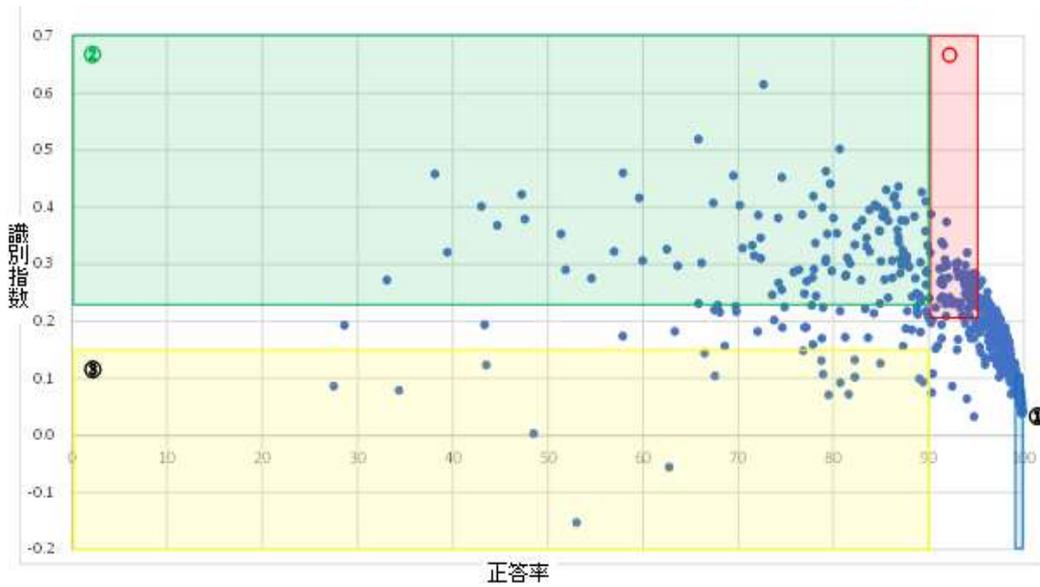
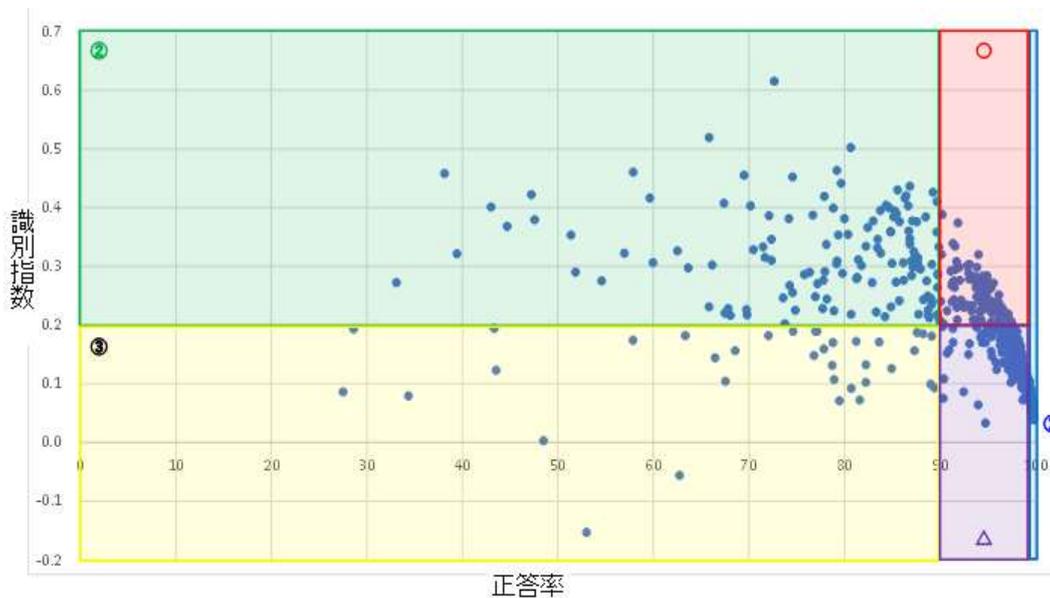


図9 過去問題 500 問の正答率・識別指数プロット (○、①~③、△の詳細は表 2 参照)



次に、出題評価基準の大項目別に評価数を集計した結果 (表 3)、良問の比率が高い (30% 以上) 大項目は、1.健康に関する指標・健康の定義と理解、3.保険医療制度の基本・看護で活

用する社会保障、10.生命活動・人体の構造と機能、12.薬物治療に伴う反応・薬物の作用とその管理であった。

表3 大項目別の各評価数。下段青字は項目内の比率

各大項目名の数字（H22, H26, H30）は出題基準の改定年を示す

大項目	評価	○	①	②	③	△	計
1. 健康に関する指標 H22/H26 健康の定義と理解 H30		13 34.2%	4 10.5%	9 23.7%	2 5.3%	10 26.3%	38
2. 健康と生活 H22/H26 健康に影響する要因 H30		5 18.5%	2 7.4%	12 44.4%	3 11.1%	5 18.5%	27
3. 保険医療制度の基本 H22/H26 看護で活用する社会保障 H30		6 31.6%	1 5.3%	5 26.3%	-	7 36.8%	19
4. 看護の倫理 H22/H26 看護における倫理 H30		2 18.2%	2 18.2%	3 27.3%	1 9.1%	3 27.3%	11
5. 関係法規 H22/H26 看護に関わる基本的法律 H30		-	1 9.1%	1 9.1%	1 9.1%	8 72.7%	11
6. 人間の特性 H22/H26/H30		1 10.0%	2 20.0%	2 20.0%	1 10.0%	4 40.0%	10
7. 人間の成長と発達 H22/H26 人間のライフサイクル各期の特徴と生活 H30		11 22.9%	6 12.5%	19 39.6%	2 4.2%	10 20.8%	48
8. 患者と家族 H22/H26 看護の対象としての患者と家族 H30		-	1 33.3%	1 33.3%	-	1 33.3%	3
9. 主な看護活動展開の場と看護の機能 H22 主な看護活動の場と看護の機能 H26/H30		2 8.3%	3 12.5%	7 29.2%	2 8.3%	10 41.7%	24
10. 生命活動 H22/ H26 人体の構造と機能 H30		17 32.1%	5 9.4%	17 32.1%	5 9.4%	9 17.0%	53
11. 病態と看護 H22/ H26 主要疾患と看護 H22(大項目 12) 疾患と徴候 H30		15 19.7%	12 15.8%	30 30.3%	7 9.2%	19 25.0%	76
12. 薬物治療に伴う反応 H22(大項目 13) 薬物治療に伴う反応 H26 薬物の作用とその管理 H30		11 36.7%	2 6.7%	8 26.7%	1 3.3%	8 26.7%	30
13. 基本技術 H22(大項目 14) 基本技術 H26 看護における基本技術 H30		2 9.5%	6 28.6%	5 23.8%	-	8 38.1%	21
14. 日常生活援助技術 H22(大項目 15) 日常生活援助技術 H26/H30		6 16.2%	11 29.7%	2 5.4%	1 2.7%	17 45.9%	37
15. 患者の安全・安楽を守る技術 H22(大項目 16) 患者の安全・安楽を守る看護技術 H26/H30		1 4.0%	12 48.0%	4 16.0%	1 4.0%	7 28.0%	25
16. 診療に伴う看護技術 H22(大項目 17) 診療に伴う看護技術 H26/H30		8 11.9%	19 28.4%	14 20.9%	9 13.4%	17 25.4%	67
計		100 20.0%	89 17.8%	132 26.4%	36 7.2%	143 28.6%	500

また、各問題の問題形式（what：「何」を尋ねる、how：方法を尋ねる、why：根拠を尋ねる）・解答選択肢形式（名詞句、数値、文、図・写真）を調べ、形式別の評価数を集計した（表4、表5）。その結果、問題形式としては what を尋ねる問題では「良問だが難問（②）」、「非該

当問題（△）」が比較多く、how/why を尋ねる問題では「易しすぎる問題（①）」が多かった。解答選択肢形式としては、文を選択肢としている問題では「易しすぎる問題（①）」が多く、数値を問う問題では「良問だが難問（②）」が比較的多かった。

表4 問題形式別評価数（下段は形式内の比率）

問題形式 \ 評価	○	①	②	③	△	計
what	87 22.8%	52 13.6%	109 28.5%	25 6.5%	109 28.5%	382
how/why	13 11.0%	37 31.4%	23 19.5%	11 9.3%	34 28.8%	118
計	100 20.0%	89 17.8%	132 26.4%	36 7.2%	143 28.6%	500

表5 解答選択肢形式別の評価数（下段は形式内の比率）

問題形式 \ 評価	○	①	②	③	△	計
名詞句	71 22.5%	59 18.7%	75 23.7%	19 6.0%	92 29.1%	316
数値	22 21.6%	6 5.9%	36 35.3%	11 10.8%	27 26.5%	102
文	4 6.6%	22 36.1%	15 24.6%	4 6.6%	16 26.2%	61
図・写真	3 14.3%	2 9.5%	6 28.6%	2 9.5%	8 38.1%	21
計	100 20.0%	89 17.8%	132 26.4%	36 7.2%	143 28.6%	500

(3) 良問ルールの抽出と作問ルール提案

国家試験出題基準の小項目ごとに類似問題をまとめ、設問ごとに評価と問題形式、解答選択肢形式の組み合わせを分析した。その分析結

果をもとに、良問ルールの抽出と作問ルールの提案書（作問ルールブック）の作成を試みた。分析フォームを表6に示す。

表6 過去問題の改善案と作問提案の表形式

出題基準		過去問題データ					の タイプ	問 の タイプ	評 価	コ メ ン ト	改 善 案	作 問 例
中項目	小項目	出題年	問題番号	問題文	1 5 4 選択枝	正答						
							名詞句 数値 文 図式	what where how why	○ ① ② ③ △	・類似問題との比較 ・解答者数の分布傾向等		

全 500 問について、表 6 に示す視点で分析を行った。各設問について、研究分担者と協力者が分析結果を記載した後（一次分析）、研究代表者がすべての問題を再分析し（二次分析）、その後研究代表者、分担者、協力者による複数

回のディスカッションを経て改善案（ルール）抽出の合意形成を行った。2022 年度末時点で 500 問のうち 450 問の一次分析を終え、200 問の二次分析、100 問の合意形成に至った。結果の一例を表 7 に示す。

表7 分析結果の例：大項目1. 健康に関する指標／健康の定義と理解

出題年	問題番号	問題文	選択枝 1	選択枝 2	選択枝 3	選択枝 4	正答	選択枝のタイプ	問いのタイプ	評価	コメント
●●●●	●●●●	日本の将来推計人口で 2020 年の 65 歳以上人口が総人口に占める割合に最も近いのはどれか。	15%	30%	45%	60%	2	数値	what	②	数値：良問だが難問、5年後の推定が難しいため数値間隔の設定が狭い可能性あり
●●●●	●●●●	平成29年(2017年)の日本の人口推計で10年前より増加しているのはどれか。	総人口	年少人口	老年人口	生産年齢人口	3	名詞句	what	①	名詞句：易問
●●●●	●●●●	平成29年(2017年)推計による日本の将来推計人口で令和47年(2065年)の将来推計人口に最も近いのはどれか。	6,800万人	8,800万人	1億800万人	1億2,800万人	2	数値	what	②	数値：間隔の設定が狭い可能性あり

【改善案】

- ・「増加している'人口区分'」を問うのは簡単すぎる
- ・将来人口推計に関する問いは具体的数値で問うのは難易度が高い。
- ・一方で、具体的数値を問うても正答率は7割を超えている
⇒数値を問う問題を出し続けることで周知され○となる可能性あり
- ・数値は明確な差をつけて選択枝設定

【作問例】

Q1:○年推計による日本の将来人口で、×年の将来推計人口に近いのはどれか。
Q2:日本の将来推計人口で○年の老年人口（年少人口、生産年齢人口のほうが難易度上がる予測）の割合で最も近いのはどれか。

2) 作問システムの検討

(1)問題の自動生成システムに関する研究動向調査

問題の自動生成に関するサーベイ論文⁽⁶⁾⁻⁽⁹⁾の調査に加え、最近の問題の自動生成に関する研究動向の調査をおこなった。その結果、従来から活発に研究されている QA システムのための大規模データセットを問題生成用に流用し、大規模事前学習済み言語モデルを利用した手法が主流であることがわかった。本研究課題で対象としている多肢選択問題においては、誤答肢の生成が主要な研究課題のひとつとなるが、先行研究で対象としている問題形式は穴埋め形式のものが多く、誤答肢の生成についても穴の位置になる語を言語モデルで推定し、その推定確率を元に生成する考え方が主流であった。本研究課題の分野において利用可能なデータ量で、先行研究で提案されているような手法が適用可能か、教科書データなどを入手し、引き続き検討する。

(2)作問システム開発のためのテキストデータ入手

作問システム構築に使用するため、出版社から看護系の教科書 54 冊分、1 冊あたり 1800 万字程度のテキストデータを入手し、データベースの構築を開始した。今後は一般公開されている医療系テキストデータとさらなる看護系教科書、関連する医学書、薬学関係の書籍、雑誌のテキストデータを入手しデータベースを構築する。

(3)ChatGPT による問題作成試行

AI を利用した大規模事前学習済み言語モデルである ChatGPT による看護師試験問題の作成を 3 つのパターンのプロンプト (A、B、C) で試みたところ、以下のような作問の結果が示された。(イタリックは ChatGPT の提示)

パターン A: キーワードを与え、問題文と 4 つの解答選択肢を作成

キーワード: (高齢者) (発症頻度)

問題文: 高齢者に多い疾患はどれか。

①心疾患 ②糖尿病 ③喘息 ④小児麻痺

正解選択肢: ①心疾患

解説: 高齢者に多い疾患は、心疾患であるとされています。高齢者の発症頻度が高い疾患としては、心疾患、糖尿病、高血圧などが挙げられます。したがって、①心疾患が正しい答えであると考えられます。

パターン B: キーワードと正解選択肢を与え、問題文と不正解の選択肢 3 つを作成

キーワード: (医師の指示) (看護師) (業務)

正解選択肢: 静脈内注射

問題文: 医師の指示に従って、看護師が行う業務はどれか。

①静脈内注射 ②指導 ③診断 ④手術

正解選択肢: ①静脈内注射

解説: 医師の指示に従って、看護師が行う業務には、静脈内注射があります。静脈内注射は、患者さんの静脈に薬剤を注射することを指します。この業務は、医師の指示に従って看護師が行います。また、看護師は、指導や医療機器の管理など、さまざまな業務を行うことができますが、診断や手術などは医師が行う業務であり、看護師が行うことはありません。したがって、①静脈内注射が正しい答えであると考えられます。

パターン C: 問題文と正解の選択肢を 1 つ与え、不正解の選択肢を 3 つ作成

問題文: 患者とのコミュニケーションで適切なのはどれか。

正解選択肢: 否定的感情の表出を受けとめる。

問題文: 患者とのコミュニケーションで適

切なのはどれか。

①否定的感情の表出を受けとめる。

②空気を読む。

③期待感を持たせる。

④自分の意見を強調する。

正解選択肢：①否定的感情の表出を受けとめる。

解説：患者とのコミュニケーションでは、相手の感情を理解し、尊重することが重要です。患者さんが、治療や病気に対して否定的な感情を抱いている場合があります。その場合、看護師は、患者さんの否定的な感情を受け止めることが適切です。患者さんが話したいことを聞き、理解することで、患者さんが感じるストレスや苦痛を和らげることができます。したがって、①否定的感情の表出を受けとめる。が正しい答えであると考えられます。

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) CBT システムの試験的運用に向けた問題プールの準備

看護師国家試験の模擬試験を実施している会社に依頼し、過去に実施された模擬試験問題、解答、解説を計 500 問収集した。また、CBT 実装可能性の検討に向けたインタビュー調査で使用するため、動画や音声を用いた問題を 5

問作成した。

2) CBT 実装可能性の検討に向けたインタビュー調査

大学、専門学校等の看護師養成所の教員、学生を対象として、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進めた。教員を対象とした調査では、国家試験を CBT で実施することに対する考え、問題の形式や回答方法、問題の再利用等 CBT にすることによる変化の教育・試験対策への影響や必要な支援等に関する質問を設定した。学生を対象とした調査では、CBT のサンプル問題への回答を依頼し、併せて難易度等についての感想をたずねることとした。出題形式として、設問文を読んで選択肢から正解を選ぶ従来型のものに加えて、動画や音声視聴して正解を選ぶ、マルチメディアを利用した設問も含むサンプル問題を作成した。

3) CBT システムの検討と問題の取り込み

本研究で用いる CBT システムとしては TAO を用いることとし、試用を開始した。①で収集した模擬試験問題のうち 250 問を取り込んだ他、インタビュー調査のサンプル問題として作成した動画や音声を用いた問題 5 問の取り込みも行った。TAO 上でのサンプル問題は図 10 のとおりである。

図 10 TAO に取り込んだ動画サンプル問題の例



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) 良問の基準について

看護師国家試験必修問題の良問ルールの抽出のために行った計量テキスト分析では、結果の共起ネットワークから良問に特徴的なルールを抽出することが困難であった。これは、必修問題が広範囲に渡り、各出題項目に関する問題数（テキストデータ）が限られるため、比較的抽象度の高い共起語が抽出されるに留まったことが要因として考えられた。

また、分析対象 500 問すべての正答率・識別指数をプロットし、先行研究¹⁰⁾の結果と照合した結果、正答率 90～99%未滿かつ識別指数 0.2 以上を『○良問』と、しその他を①易しすぎる問題、②良問だが難問、③あまり適切ではない問題、△非該当問題（前述全てに当てはまらない問題）の 5 つに区分したことで、各問題の特性を明確にした。この基準に基づき、小項目、中項目ごとに類題を丹念に比較することで、良問とそれ以外の問題の相違点を見出すことが可能となったことから、本基準は国試必修問題の質評価の一指針となり、更に本作業を行うことでルールブック作成へと繋がるものと考えられる。

2) 作問システムの検討について

事前学習済み言語モデル（pre-trained model）として、公開直後の ChatGPT を活用し、様々なパターンのプロンプトを用いることで作問の可能性を検討した。本研究用のデータベース構築は途上であったため、あくまで一般の web 上のデータを元に試用したが、作問が技術的には可能であることが確認できたと考える。一方で限界の検証は必要であり、他社の事前学習済みモデル（大規模なものでは Google の Bard、Meta の LLaMA、小

規模のものは多数該当）を使用し、構築中の看護分野のテキストデータを与えて更に fine-tuning をすることが今後必要である。

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) CBT システムの運用について

CBT システムの試験的運用にあたっては、豊富な問題プールが必要とされる。今年度は問題プールの蓄積に向けて、看護師国家試験の模擬試験の問題 500 問を収集し、当初目標として設定した 250 問を超えた。また、動画や音声などマルチメディアを活用した問題の作成にも着手しており、CBT システムの試験的運用の準備は順調に進んでいるといえる。2023 年度も問題の収集・作成を継続し、問題プールの充実を図っていく。

2) CBT 実装可能性の検討について

CBT 実装可能性の検討に向けた調査では、調査への回答にあたり、CBT をどのような形式で実施するかを定義しなければ、回答が難しいと考えられたため、想定する CBT の出題形式を定義し、サンプル問題を作成した。この調査を通じて、CBT ならではの作問や CBT システムの試用も行うことができた。このことから、調査により受験者側の準備状況や課題を明らかにできるだけではなく、作問や運営上の課題もリストアップすることが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

過去 10 年間の看護師国家試験必修問題 500 問について、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、国家試験問題の現状と各設問の課題を提示したことは、これまでの看護師国家試験の定量的評価に繋がるものとする。また、5区分の評価を軸に、問題形式、解答形式の観点で出題基準の項目ごとの問題分析を行うことで、項目単位での良問のルールを抽出することが可能となり、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、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試みたことで、今後看護に関わるデータを強化したデータベースを構築し、それを利用した作問システム開発への示唆を得ることができた。

2. 【研究2】看護師等国家試験へのCBT実装の可能性の検討

2022年度はCBTシステムの試験運用に向けた問題収集・作成と、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。試験運用に向けた問題は500問収集でき、動画や音声等のマルチメディアを活用した問題の作成にも着手した。インタビュー調査については、年度末に研究倫理審査委員会の承認を得て、新年度から調査開始の運びとなった。2023年度は調査を開始・継続し、CBT実装に向けての課題等を明らかにしていく。

F. 健康危険情報

該当なし

G. 研究発表

現状では未定

H. 知的財産権の出願・登録状況

該当なし

引用文献

- 1) Sheng Shen, Yaliang Li, Nan Du, Xian Wu, Yusheng Xie, Shen Ge, Tao Yang, Kai Wang, Xingzheng Liang, Wei Fan(2020). On the Generation of Medical Question-Answer Pairs, ArXiv, abs/1811.00681.
- 2) Mark J Gierl, Hollis Lai, Hollis Lai, Simon R Turner(2012). Using automatic item generation to create multiple-choice test items, Medical Education, 46(8), 757-65.
- 3) Mark J. Gierl, Hollis Lai(2018). Using Automatic Item Generation to Create Solutions and Rationales for Computerized Formative Testing, Applied Psychological Measurement, 42(1), 42-57.
- 4) Filipe Falcão, Patrício Costa & José M. Pêgo(2022). Feasibility assurance: a review of automatic item generation in medical assessment, Advances in Health Sciences Education, <https://doi.org/10.1007/s10459-022-10092-z>.
- 5) Dhawaleswar Rao CH and Sujan Kumar Saha(2020). Automatic Multiple Choice Question Generation From Text :A Survey, IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, 13(1), 14-25.
- 6) D. R. CH and S. K. Saha. Automatic multiple choice question generation from text : A survey. IEEE Transactions on Learning Technologies, 13(1):14--25, 2020.
- 7) C.-Y. Lu and S.-E. Lu. A survey of approaches to automatic question generation: from 2019 to early 2021. In The 33rd Conference on Computational Linguistics and Speech Processing (ROCLING 2021), pages 151-162, 2021.
- 8) G. Kurdi, J. Leo, B. Parsia, U. Sattler, and

- S. Al-Emari. A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 20:121-204, 2020.
- 9) L. Pan, W. Lei, T.-S. Chua, and M.-Y. Kan. Recent advances in neural question generation. arXiv:1905.08949, <https://doi.org/10.48550/arXiv.1905.08949>, 2019.
- 10) 令和元年度厚生労働科学研究「保健師助産師看護師国家試験における現状の評価及び出題形式等の改善に関する研究林」総括・分担研究報告書、林直子（研究代表者）他、2020年3月.