論文 / 著書情報
Article / Book Information

論題	言語モデルを用いた看護師国家試験問題の誤答選択肢自動生成
著者	城戸祐世, 山田寛章, 徳永健伸, 木村理加, 三浦友理子, 佐居由美, 林直子
出典	言語処理学会第31回年次大会(NLP2025)発表論文集, pp. 1074-1079
発行日	2025, 3
権利情報	Creative Commons Attribution 4.0 International License(https://creativecommons.org/licenses/by/4.0/).

言語モデルを用いた看護師国家試験問題の誤答選択肢自動生成

城戸祐世 [1]　山田寛章 [1]　徳永健伸 [1]
木村理加 [2]　三浦友理子 [2]　佐居由美 [2]　林直子 [2]
[1] 東京科学大学 情報理工学院　[2] 聖路加国際大学 看護学研究科
{kido.y.ad@m,yamada@c,take@c}.titech.ac.jp
{rikakimura,miura-yuriko,yumi-sakyo,naoko-hayashi}@slcn.ac.jp

概要

本研究では，看護師国家試験問題における誤答選択肢の自動生成に大規模言語モデルを活用する．生成には日本語大規模言語モデルと，API を通じて利用可能なモデルをそれぞれ複数用い，過去の試験や予備校の模擬試験の問題をデータセットとして出力の制御を行う．生成した誤答選択肢を選択肢候補として看護師国家試験問題作成経験者に提示し，実際の問題作成における負担軽減や効率改善への寄与を分析した．その結果，大規模言語モデルによる誤答選択肢は有用であり，作問作業の効率改善の可能性が示された．

1　はじめに

現在，日本では資格試験におけるコンピュータ試験の利用が拡大している．資格試験は民間団体が行う試験と，医師や看護師，司法試験等の国家試験に分かれる．国家試験のうち，国が直接実施する試験ではコンピュータ利用が依然として遅れている．特に看護師国家試験において，コンピュータ利用の一つとして問題作成の自動化が挙げられるが，これらを実用的に行った例はまだなく，その可能性が期待されている．

多肢選択式問題で構成される看護師国家試験においては，選択肢の質が試験全体の信頼性に大きく影響する．そのため，問題文や正解選択肢と同様に，誤答選択肢の作成にも十分な注意が必要である．誤答選択肢の作成は，問題作成者にとって重要かつ難易度の高い作業である．問題作成経験者によれば，誤答選択肢には排除するための必要な知識の根拠が明確であることが求められる．また，誤答選択肢は受験者がすでに学習した用語や概念で構成されるべきであり，架空の用語や造語は認められない．さらに，選択肢の中に二律背反のものが含まれることは，片方を安易に排除できる観点から望ましくない．このような理由から，誤答選択肢の設計には慎重な配慮が不可欠である．質の高い問題を一定数作成するためには，試験問題として扱われる数以上の問題作成が必要である．現在は問題作成者が慎重に吟味して作成しており，これにより問題の質は高く保たれているが，多量の問題作成には多くの人的負担と時間を要している．したがって有用な多肢選択式問題を自動生成することと，問題作成のための労力や作成時間を削減することは重要な課題である．

大規模言語モデル (LLM: Large Language Models) の汎用性の高さ [1] から，問題生成においても LLM の利用例がある．GPT-3 を用いて問題の自動生成と生成された候補の評価を行い，LLM による効果的な問題生成の可能性が示されている [2]．さらに，ChatGPT が人間の専門家と同等の英語読解テスト問題を作成できるかを検証した結果，文章の自然さや表現については同等と評価された [3]．また看護学においては，看護師国家試験のための多肢選択式問題の自動生成に LLM を活用する試験的な取り組みが行われた [4]．

以上の背景を踏まえ，看護学における，多肢選択式問題の誤答選択肢生成タスクに焦点を当て，LLM の利用とその効果について検証を行う．

2　看護師国家試験

看護師国家試験とは，公的な資格である看護師国家資格を取得するために必要な試験である．試験の目的は，看護師が保健医療の現場に踏み出す際に，少なくとも持っておくべき基本的な知識及び技能を測ることである [5]．試験問題は多肢選択式問題で構成され，必修問題，一般問題，状況設定問題の 3 つに区分される．本研究ではこの中の必修問題について取り扱う．

必修問題は主に図 1 に示すような四肢択一問題か

日本の令和 2 年（2020 年）の死因順位の第 1 位はどれか。 ← **問題文**

1. 悪性新生物 ← **正解選択肢**
2. 脳血管疾患
3. 心疾患
4. 肺炎

} **誤答選択肢**

図 1　看護師国家試験の必修問題例.

らなる．1 つの問題は問題文と，1 つの正解選択肢，3 つ (稀に 4 つ) の誤答選択肢から構成されている．問題の選択肢は図 1 の例のような名詞句のほかに文，数値，図表のものがある．本研究では LLM の特性を活かしやすいという観点から，選択肢が名詞句，文である問題の誤答選択肢を生成対象とした．

3　誤答選択肢の生成

本研究では指示応答型の言語モデルに対して，誤答選択肢を生成させるようにプロンプトを与えて，生成する手法を採用する．複数種類の LLM を利用して生成し，出力の制御にはファインチューニングや few-shot 学習 [6] をモデルに合わせて利用した．

3.1　データセット

厚生労働省より提供された看護師国家試験問題の過去問題 500 問と看護師国家試験等対策予備校から提供された 2022 年度実施の模試問題 250 問をもとにデータセットを構築した．過去問題からは 366 問，模擬試験問題からは 210 問，計 576 問をデータセットの対象とした．

3.2　使用モデル

本実験で使用したモデルを以下にまとめる．

- Swallow-70b-instruct-v0.1 (Llama2-Swallow) [7] [8] [9]
- Llama-3-Swallow-70B-Instruct-v0.1 (Llama3-Swallow) [8] [9] [10]
- Llama3-Preferred-MedSwallow-70B (MedSwallow) [10] [11]
- GPT-3.5-turbo-0613 (GPT-3.5) [12]
- GPT-4-0613 (GPT-4) [13]
- GPT-4o-240513 (GPT-4o) [14]

3.3　生成実験

Llama2-Swallow, Llama3-Swallow, MedSwallow, GPT-3.5 に対しては，作成したデータセットを使ってファインチューニングを行った．このうち Llama2-Swallow, Llama3-Swallow, MedSwallow のファインチューニングには QLoRA [15] を採用した．GPT-3.5 は Microsoft Azure Open AI [1] の API 上でファインチューニングを行なった．以降，ファインチューニングを行った GPT-3.5 は行っていないものと区別して GPT-3.5-FT とする．

指示文:　N 択問題「ここに問題文が入る」で、正解選択肢が「ここに正解選択肢が入る」のとき、誤答選択肢を M つ考えてください。

指示文:　N 択問題「ここに問題文が入る」で、正解選択肢が「ここに正解選択肢が入る」のとき、誤答選択肢を 3 つ考えてください。
応答文:
- 誤答選択肢 $_1$:「ここに誤答選択肢が入る」
- 誤答選択肢 $_2$:「ここに誤答選択肢が入る」
- 誤答選択肢 $_3$:「ここに誤答選択肢が入る」

— 以上を 5 回繰り返す —
指示文:　N 択問題「ここに問題文が入る」で、正解選択肢が「ここに正解選択肢が入る」のとき、誤答選択肢を M つ考えてください。

(N は問題に依存，M は生成数を指定, 本研究では $M = 4$)

図 2　誤答選択肢生成プロンプト．上が 0-shot, 下が 5-shot.

ファインチューニングをした Llama2-Swallow, Llama3-Swallow, MedSwallow, GPT-3.5-FT では，応答例を与えずに生成をした．ファインチューニングをせずに，指示応答型言語モデルをそのまま用いた GPT-3.5, GPT-4, GPT-4o では，5 つの応答例を与えて生成をした．なおこの応答例はデータセットからランダムに選んだ．図 2 に誤答選択肢生成に用いたプロンプトを示す．本研究では各モデルが生成する誤答選択肢の数を 4 つとした．

4　LLM を利用した作問実験

本論文では問題作成者が試験問題の問題文と正解選択肢を見て，誤答選択肢を作成する作業を作問と呼ぶ．

4.1　作問の手順

本研究では，作問作業における LLM の効果を測るために，LLM により生成された誤答選択肢を作問者に提示するという方法をとった．問題文と正解選択肢のほかに LLM による誤答選択肢候補を提示する作問者グループ (グループ A: Assisted) と，問題

[1] https://azure.microsoft.com/ja-jp/products/ai-services/openai-service

表 1 LLM により生成された誤答選択肢種類数

問題番号	生成数	重複数	候補数
1	13	4	6
2	28	0	14
3	25	2	11
4	20	8	10
5	24	4	10
6	27	1	13
7	18	6	9
8	25	3	11
9	8	5	5
10	27	1	13
合計	215	34	102

文と正解選択肢のみを提示する作問者グループ (グループ C: Control) に分けて実験を行い，作問に使われた誤答選択肢の分析，比較を行う．以下に 1 つの問題に対する，グループ A の作問の流れを示す．

1. 作問者に問題文，正解選択肢，LLM による誤答選択肢の候補を提示する．
2. 作問者は誤答選択肢候補から試験として相応しいものを 3 つ採用する．
3. 誤答選択肢候補の中に相応しいものが足りない，もしくはない場合は作問者自身で足りない分の誤答選択肢を作成する．

グループ C では誤答選択肢 3 つ全てを問題文，正解選択肢から作問者自身が作成する．

4.1.1 対象問題

作問実験の対象問題として，看護師国家試験等対策予備校から提供された 2021 年度実施の模擬試験問題 250 問をもとに，その中から 10 問を看護学を専門とする著者が選んだ．

4.1.2 誤答選択肢候補の選別

誤答選択肢候補は 3 項で示した 7 つの LLM (Llama2-Swallow, Llama3-Swallow, MedSwallow, GPT-3.5, GPT-3.5-FT, GPT-4, GPT-4o) を使用して生成した．利用方法を考慮すると，誤答選択肢候補には多様性が求められるため，複数の LLM から生成された誤答選択肢を統合して候補とした．以下に 1 つの問題における生成誤答選択肢の統合方法を示す．

1. 各 LLM で誤答選択肢を 4 つずつ生成する．
2. 複数の LLM により重複して出力された選択肢を候補とする．
3. 各 LLM に由来する誤答選択肢数がそれぞれ 2 個以上になるまで，生成された誤答選択肢から候補に追加する．

表 1 にて LLM が生成した誤答選択肢の種類数を示す．表 1 の左から 3 列目は生成数のうち複数の LLM による重複があった数，4 列目は実際に作問者に提示する統合後の候補数を表す．

4.1.3 作問者

作問実験を実施するために，16 名の看護師国家試験作問経験者に参加を依頼をした．作問者のうち 8 名は国家試験作問経験年数が 5 年未満であり，残りの 8 名は経験年数が 5 年以上である．これらをそれぞれ初心者と経験者と呼ぶこととし，グループ A の担当とグループ C の担当に均等に振り分けた．さらに各グループは 10 問の作問を行う際に，前半 5 問を担当する 4 名と後半 5 問を担当する 4 名に分け，こちらも初心者と経験者が均等になるように分けている．作問者は “A1N1” のように表記され，1 文字目がグループ A または C，2 文字目が問題の前半担当 (1) か後半担当 (2)，3 文字目が初心者 (N) か経験者 (V)，最後が 1 人目，2 人目としている．なお，作問者のうち C2V2 (グループ C の後半 5 問を担当していた経験者の作問者 1 名) が辞退したため，作問者は合計で 15 名 (グループ A が 8 名，グループ C が 7 名) である．

4.2 作問後の主観調査

LLM の利用による作問効率の変化を分析するために，各作問者に対してアンケートを実施した．アンケートは各問題の作問後にそれぞれ行われ，全 6 問で構成される．以下に設問の詳細を示す．

Q1: 誤答選択肢を 3 つ作成するのにどのくらい時間がかかりましたか (分)
Q2: 誤答選択肢の作成は、国家試験委員として誤答選択肢を作成したときと比べて負担感はどのくらいでしたか。
Q3: AI が作成した誤答選択肢の案は、誤答選択肢の作成にあたり参考になりましたか。
Q4: AI が作成した誤答選択肢の案は、ブレーンストーミングに役立ちましたか。
Q5: AI が作成した誤答選択肢の案は、ご自身の自由な発想を阻害しましたか。
Q6: AI が作成した誤答選択肢の案で有効だったもの、採用したものはどれですか。該当するもの

表 2　グループ A で作問された誤答選択肢の統計 (括弧内は LLM により生成された誤答選択肢の採用数).

問題番号	選択肢数	LLM 採用割合	アンケート Q6
1	4 (4)	1.00	2
2	10 (6)	0.60	4
3	10 (5)	0.50	3
4	11 (8)	0.73	4
5	9 (7)	0.78	4
6	10 (5)	0.50	3
7	8 (1)	0.13	0
8	10 (3)	0.30	0
9	8 (5)	0.63	0
10	12 (4)	0.33	2
合計	92 (48)	0.52	22

作問者	選択肢数	LLM 採用割合	アンケート Q6
A1N1	15 (15)	1.00	6
A1N2	15 (15)	1.00	1
A1V1	15 (4)	0.27	3
A1V2	15 (12)	0.80	8
A2N1	15 (2)	0.13	1
A2N2	15 (10)	0.67	1
A2V1	12 (8)	0.67	1
A2V2	15 (3)	0.20	3
合計	117 (69)	0.59	24

をすべてを選んでください。

Q1 および Q2 はグループ A, C の両方に対して聞き，Q3〜6 はグループ A のみに対して聞いた．設問のうち Q2〜5 は 1〜5 の 5 段階リッカート尺度を用いて調査した．なお数値が大きいほど設問に対してよく当てはまることを示す．Q6 は提示した誤答選択肢のうち，設問にあてはまるものを選ぶ形式である．

4.3　結果と分析

4.3.1　LLM が生成した誤答選択肢の有用性

グループ A で作問された 10 問についての誤答選択肢の統計を表 2 に示す．表 2 の上半分は問題番号ごとにまとめたもの (種類数) で，下半分は作問者ごとにまとめたもの (総数) である．左から 2 列目は作問された選択肢の数を表し，3 列目はそのうち LLM による誤答選択肢が採用された割合を示す．また 4 列目はアンケート Q6 で選ばれた (作問で採用された 3 つは除く) LLM による誤答選択肢の数を示す．表 2 の合計をみると，作問された全選択肢 92 種のうち LLM による誤答選択肢は 48 種 (52%) であった．総数で見ると 117 個の中には LLM による誤答選択肢が 69 個 (59%) 含まれていた．これらの結果から作問された誤答選択肢の半数以上が LLM 由来であったことが分かる．また採用された LLM による誤答選択肢 48 種とアンケート Q6 で選ばれた 22 種の和は 70 種であった．これは表 1 の本実験で提示した選択肢候補全 102 種のうちの 69%にあたる．以上の結果より LLM が生成した誤答選択肢が有用であることを確認できた．

表 3　アンケート Q1〜Q5 の結果 (問題単位のマクロ平均, 括弧内は標準偏差).

	Q1 (分)	Q2	Q3	Q4	Q5
A	7.0 (2.5)	2.2 (0.6)	3.6 (1.1)	3.1 (1.3)	2.0 (0.5)
C	21.0 (17.9)	2.6 (0.4)	N/A	N/A	N/A

4.3.2　LLM を利用した作問の効率

表 3 にアンケートの結果を示す．Q1 に関して，全 10 問の平均作成時間を比較すると，グループ A では 7.0 分，C では 21.0 分の作成時間を要した．次に，Q2 の負担感について 10 問の平均をとると，グループ A で 2.2，C で 2.6 であり，グループ A の負担感の方がわずかに小さいことが確認された．Q3 の参考度合いと Q4 のブレーンストーミングの貢献度合いは 10 問の平均でそれぞれ 3.6, 3.1 で，いずれも 1〜5 のリッカート尺度において評価の平均値である 3.0 を上回っている．これは LLM による誤答選択肢は参考になることを示唆している．さらに，Q5 の阻害度合いは 2.0 で 3.0 を下回っており，LLM による誤答選択肢は作問者の自由な発想を大きくは阻害しないと考えられる．

5　おわりに

本研究では，大規模言語モデルを利用して，看護師国家試験問題の誤答選択肢生成を行った．具体的には，指示応答型の言語モデルにプロンプトを与え誤答選択肢を生成した．生成した誤答選択肢を作問者に提示して，誤答選択肢の候補として利用することで，大規模言語モデルの効用を評価するという実験を行った．実験の結果，大規模言語モデルにより生成した誤答選択肢は，作問者からの評価では有用であることがわかった．また大規模言語モデルの利用による効率改善の可能性が示された．今回の作問者の評価による実験では言語モデルによる誤答選択肢の検証が十分とは言い切れないため，今後作問した問題を使って模擬試験を実施する予定である．

謝辞

本研究は厚生労働省（https://www.mhlw.go.jp/）および厚労科研 (課題番号：22AC1003，研究代表者：林直子) の支援を受けたものです．

参考文献

[1] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, Zihao Wu, Lin Zhao, Dajiang Zhu, Xiang Li, Ning Qiang, Dingang Shen, Tianming Liu, and Bao Ge. Summary of chatgpt-related research and perspective towards the future of large language models. **Meta-Radiology**, Vol. 1, No. 2, p. 100017, 2023.

[2] Xingdi Yuan, Tong Wang, Yen-Hsiang Wang, Emery Fine, Rania Abdelghani, Hélène Sauzéon, and Pierre-Yves Oudeyer. Selecting better samples from pre-trained LLMs: A case study on question generation. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 12952–12965, Toronto, Canada, July 2023. Association for Computational Linguistics.

[3] Dongkwang Shin and Jang Ho Lee. Can chatgpt make reading comprehension testing items on par with human experts? **Language Learning & Technology**, Vol. 27, No. 3, pp. 27–40, 2023.

[4] Yûsei 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. In **Proceedings of the 16th International Conference on Computer Supported Education - Volume 1**, pp. 821–829. INSTICC, SciTePress, 2024.

[5] 厚生労働省. 保健師助産師看護師国家試験出題基準 令和 5 年版,（2024-12 閲覧）. https://www.mhlw.go.jp/content/10803000/000958440.pdf.

[6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 1877–1901. Curran Associates, Inc., 2020.

[7] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models. **CoRR**, Vol. abs/2307.09288, , 2023.

[8] Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. Continual pre-training for cross-lingual llm adaptation: Enhancing Japanese language capabilities. In **Proceedings of the First Conference on Language Modeling**, COLM, pp. 1–25, University of Pennsylvania, USA, October 2024.

[9] Naoaki Okazaki, Kakeru Hattori, Hirai Shota, Hiroki Iida, Masanari Ohi, Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Rio Yokota, and Sakae Mizuki. Building a large Japanese Web corpus for large language models. In **Proceedings of the First Conference on Language Modeling**, COLM, pp. 1–18, University of Pennsylvania, USA, October 2024.

[10] AI@Meta. Llama 3 model card,（2024-12 閲覧）. https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md.

[11] Junichiro Iwasawa, Keita Suzuki, and Wataru Kawakami. Llama3 preferred medswallow 70b,（2024-12 閲覧）. https://huggingface.co/pfnet/Llama3-Preferred-MedSwallow-70B.

[12] OpenAI. Chatgpt,（2024-12 閲覧）. https://www.openai.com/.

[13] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report. **CoRR**, Vol. abs/2303.08774, , 2023.

[14] OpenAI. Gpt-4o system card,（2024-12 閲覧）. https://openai.com/index/gpt-4o-system-card/.

[15] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. **CoRR**, Vol. abs/2305.14314, , 2023.

A 付録: 実験設定

学習に用いたデータセットの内訳は訓練データ 518 問 (名詞句 427 問，文 91 問)，学習途中の最適なモデルを見つけるための検証データ 58 問 (名詞句 47 問，文 11 問) である．Llama2-Swallow, Llama3-Swallow, MedSwallow の学習に関わるハイパーパラメータは LoRA ランク 8, バッチサイズ 1, 学習率 10^{-4}, 学習回数 10 とした．また GPT-3.5-FT に関して，API 上で設定したファインチューニングに関わるハイパーパラメータはバッチサイズ 1, 学習回数 5 である．こちらもオープンソース LLM の学習と同様のデータセットを用いた．誤答選択肢の生成におけるハイパーパラメータは全モデルで temperature を 0 に設定し，GPT-3.5, GPT-4, GPT-4o では top_p を 0.95 に設定した．

B 付録: アンケート結果詳細

表 4 問題ごと，作問担当者ごとのアンケート結果 (括弧内は標準偏差).

		Q1	Q2	Q3	Q4	Q5	Q6
問題番号	グループ	作成時間 (分)	負担感	参考度	ブレスト貢献	阻害度	参考肢 (個)
1	A	4.0 (4.1)	1.0 (0.0)	4.8 (0.5)	4.5 (0.6)	1.8 (1.0)	2
	C	2.8 (1.7)	2.0 (1.2)				
2	A	9.5 (6.4)	2.5 (1.3)	3.8 (1.9)	3.8 (0.5)	2.3 (1.3)	4
	C	8.3 (3.5)	3.3 (1.3)				
3	A	5.0 (3.6)	2.3 (1.0)	4.5 (0.6)	4.3 (0.5)	1.5 (0.6)	3
	C	3.5 (1.9)	2.8 (1.0)				
4	A	5.5 (6.4)	2.0 (1.2)	4.5 (0.6)	4.3 (0.5)	1.3 (0.5)	4
	C	3.5 (1.0)	2.5 (0.6)				
5	A	5.5 (3.1)	2.3 (1.5)	4.8 (0.5)	4.5 (0.6)	1.8 (1.0)	4
	C	5.8 (3.0)	2.0 (1.2)				
6	A	8.3 (4.7)	1.8 (0.5)	3.8 (1.3)	2.8 (1.0)	2.8 (1.7)	3
	C	28.3 (12.6)	2.0 (1.0)				
7	A	6.0 (4.1)	2.5 (1.0)	1.8 (0.5)	1.5 (0.6)	2.8 (1.3)	0
	C	33.3 (25.2)	2.7 (1.5)				
8	A	8.5 (7.9)	2.5 (1.0)	2.5 (1.9)	1.5 (1.0)	2.3 (1.5)	0
	C	40.0 (17.3)	2.7 (1.5)				
9	A	5.3 (4.2)	3.0 (1.6)	2.3 (1.9)	2.3 (1.9)	1.5 (1.0)	0
	C	38.3 (44.8)	3.0 (2.0)				
10	A	12 (6.7)	2.5 (1.3)	3.5 (1.3)	2.0 (0.8)	2.3 (1.5)	2
	C	46.7 (37.9)	2.7 (1.5)				

		Q1	Q2	Q3	Q4	Q5	Q6
作問担当者	グループ	作成時間 (分)	負担感	参考度	ブレスト貢献	阻害度	参考肢 (個)
A1N1	A	3.4 (1.1)	1.2 (0.5)	4.6 (0.6)	4.6 (0.6)	2.2 (1.3)	6
A1N2	A	2.2 (0.8)	3.0 (1.2)	4.6 (0.6)	4.2 (0.5)	1.4 (0.9)	1
A1V1	A	6.0 (5.2)	2.6 (0.9)	3.8 (1.6)	3.8 (0.5)	1.8 (0.5)	3
A1V2	A	12.0 (2.7)	1.2 (0.5)	4.8 (0.5)	4.4 (0.6)	1.4 (0.6)	8
A2N1	A	14.0 (6.5)	2.6 (0.6)	2.2 (1.3)	1.8 (0.8)	1.4 (0.9)	1
A2N2	A	4.0 (2.5)	1.0 (0.0)	3.6 (2.0)	2.0 (1.7)	1.2 (0.5)	1
A2V1	A	5.8 (3.0)	3.4 (1.1)	2.6 (1.5)	2.4 (1.1)	3.0 (1.2)	1
A2V2	A	8.4 (4.4)	2.8 (0.5)	2.6 (1.3)	1.8 (0.8)	3.6 (0.9)	3
C1N1	C	5.0 (3.1)	1.8 (0.8)				
C1N2	C	3.6 (3.7)	3.0 (0.0)				
C1V1	C	4.4 (3.1)	3.6 (0.9)				
C1V2	C	6.0 (2.2)	1.6 (0.6)				
C2N1	C	68 (21.7)	3.8 (1.1)				
C2N2	C	27 (6.7)	3.0 (0.0)				
C2V1	C	17 (8.4)	1.0 (0.0)				