ORIGINAL ARTICLE

2023 Japanese national examination for rehabilitation professionals and Chat GPT: Chat GPT passed the national examination for speech-language-hearing therapists

Masatoshi TAKEDA, Kayo MATSUO, Yukito UEDA, Yasuko TAKAHASHI, Yoshimi TSUKAMOTO, Aoi ASHIZUKA, Hidetsugu WADA, Ryohei KONO, Hirotoshi UTSUNOMIYA, Kazuki NOMURA, Kunihiro UMAYAHARA

Osaka Kawasaki Rehabilitation University
Correspondence: Masatoshi Takeda, MD, PhD, Osaka Kawasaki Rehabilitation University, 158 Mizuma, Kaizuka City, Osaka, Japan. E-mail: masatakeda@ kawasakigakuwn.ac.jp
Disclosure: There is no conflict of interest to be declared.

Abstract

Background: The use of Chat GPT is rapidly expanding in Japanese society, and the educational community is considering appropriate uses of it. The medical education sector, including training schools for rehabilitation professionals, is also actively engaged in discussion of effective ways to use Chat GPT. The 25th National Examination for Speech Therapists and the 58th National Examination for Physical Therapists and Occupational Therapists were held in February 2023. The pass rates of Speech, Physical, and Occupational therapists were 67.4%, 87.4%, and 83.8%, respectively. Analysis of the results of answers by Chat GPT compared with those by actual students will hopefully facilitate discussion of how to use Chat GPT in the education of rehabilitation professionals.

Methods: We used Chat GPT 3.5 to answer the questions from the 25th National Examination for Speech Therapists, which had the lowest pass rate among the three exams for rehabilitation specialists, and because it was mainly based on language-based questions and not heavily affected by the use of figures and charts.

Results: Chat GPT had a 66% correct answer rate, which was above the level to pass the examinations. The scores were purely based on understanding of language-based questions, but Chat GPT had a comparable ability to our students in correctly answering questions from the exam.

Conclusions: This study is the first to show that Chat GPT has the capability to pass a national medical examination. AI is still in relative infancy, but our results show that there is wide potential for academic and education application beyond clerical work. Further studies will seek to elucidate more precise potential use in academic settings.

INTRODUCTION

Self-regenerative language artificial intelligence technology based on large language models has recently undergone remarkable development. Chat GPT (OpenAI, San Francisco, CA), which was first released in November 2022, can for example respond to questions entered by users in a natural, human-like dialogue format. Its high response accuracy has been widely discussed and it has been applied in many situations. Chat GPT is based on artificial intelligence (AI) that generates natural language in response to textual instructions. It is based on a large language model called generative pre-trained transformer and from which its name is derived. Chat GPT is capable of generating sentences with very sophisticated content, but can still return incorrect or inappropriate answers. Reasons for this include the source information being possibly outdated and because the AI creates sentences based on word frequencies and interrelationships rather based than on facts. The data to be learned does not contain up-to-date information, so answers to questions about recent events may be inadequate, and the AI may also fail to perfectly understand or clarify the intent of the given text and can give a different answer than expected.

Despite these disadvantages, Chat GPT is being increasingly used in the medical field. For example, it is being used to prepare medical certificates, letters

of introduction, and medical insurance claims. By using Chat GPT to assist in clerical work, for example, medical professionals can theoretically spend more time with their patients. In addition, the use of Chat GPT (and similar AI-powered language models) in the field of education for medical professionals is a hot topic, in part because it might eventually be used as a tool to help in passing national examinations and other qualification tests (Kung, 2023; Sharma, 2023).

Physical, occupational, and speech-language-hearing therapy are covered by Japanese national health insurance as rehabilitation services, which can be reimbursed when these rehabilitation services are applied to patients under the direction of medical doctors. Physical and occupational therapists have been officially registered as national qualified professionals by the Japanese Ministry of Welfare and Labor since 1965, and qualified speech-language-hearing therapists since 1998. To be officially registered, rehabilitation professionals have to pass the relevant national examination for physical, occupational, or speech-language-hearing therapists which is held in February or March every year. The accumulated number of registered physical therapists (*rigakuryohoshi*) is 192,327, there are 94,255 occupational therapists (*sagyoryohoshi*) and 38,200 speech-language-hearing therapists (*gengochokakushi*). The 25th National Examination for Speech-Language-Hearing Therapists was held on February 18, 2023, and the 58th National Examinations for Physical Therapists and Occupational Therapists were held on February 19, 2023.

Details of these examinations are shown in Table 1. Candidates required 60% correct answers to successfully pass the examinations.

As a means of exploring the possibility of using Chat GPT in education for training rehabilitation professionals, we used Chat GPT version 3.5 to answer the questions from the National Examination for Speech-Language-Hearing Therapists, which had the lowest pass rate among the national examinations for rehabilitation professionals administered in February 2023. Understanding and evaluation of figures or charts was required in 60% and 42% of the practical questions for Physical Therapists and for Occupational Therapists, respectively, in the 58th National Examination. Considering the weighting of the questions, it is not appropriate to evaluate the ability of Chat GPT by having it solve only the questions without figures or charts. Therefore, we did not analyze whether the Chat GPT would pass the 58th National Physical/ Occupational Therapists Examination and the National Occupational Therapist Examination this time, and we focused upon the National Examination for Speech-Language-Hearing Therapists. We compared the percentage of correct answers with the response rate of students who took the exam at Osaka Kawasaki Rehabilitation University.

METHODS

The 25th National Examination for Speech-Language-Hearing Therapists had morning and after-

Table 1. Details of the 2023 National Examinations for rehabilitation professionals.

	Questions		Question point value	Total possible score	Passing score
25th National Examination for Speech-Language Pathologists and Hearing Therapists		200 questions comprised of: 139 x one-answer multiple choice, 50 x two-answer multiple choice, 11 x three-answer multiple choice. Two questions were answered using sound spectrum diagrams, so were not included in this study.	1 point	200 points	≥120 points (60%)
58th Physical Therapist and Occupational Therapist Examinations (100 of the 200 questions are the same in both versions)	Physical Therapist Examination	160 general questions (revised to 158 after 2 questions deemed inappropriate) Three questions are asked using figure or table	1 point	158 points	≥167 points (60%), including ≥43 points on the practical section.
		40 practical questions 24 questions using figures or tables 38 one-answer multiple choice 2 two-answer multiple choice	3 points	120 points	
	Occupational Therapist Examination	160 general questions (revised to 159 after 1 question deemed inappropriate)	1 point	159 points	≥167 points (60%), including ≥43 points on the practical section.
		40 practical questions	3 points	120 points	

noon sessions, both of which had 100 questions. Most of the questions were based on Japanese text, but one question in the morning session and one in the afternoon session required the use of a sound spectrum diagram. We used Chat GPT to answer the 198 questions that did not use diagrams. We compared these answers with the correct answers announced by the Ministry of Health, Labour and Welfare of Japan and the answers of all examinees from Osaka Kawasaki Rehabilitation University (*n*=15), from which the graduates of the speech-language-hearing course had a similar pass rate of 60-

70% in previous consecutive national examinations.
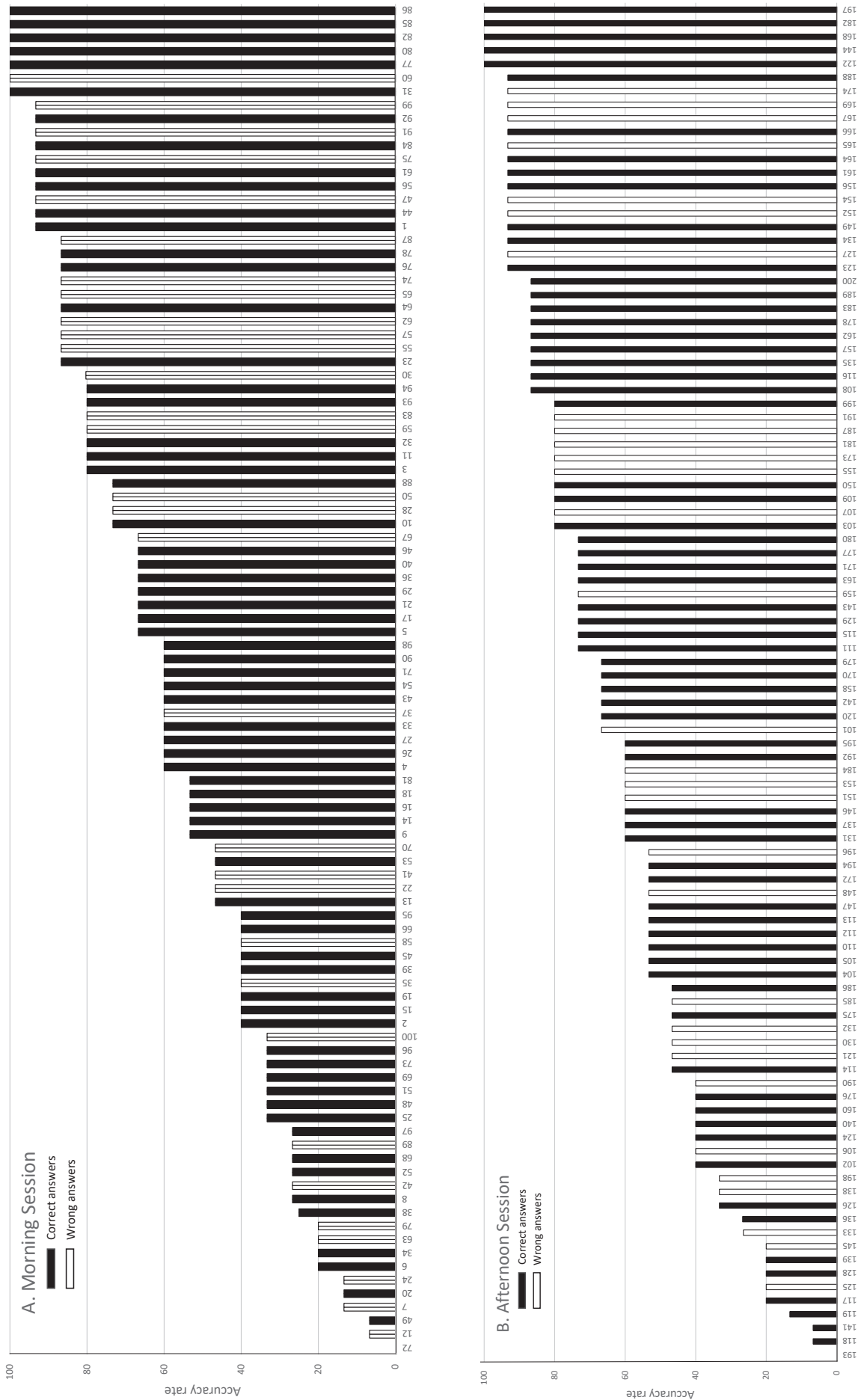
RESULTS

This year, the 200 questions in the examination were categorized according to question type (Tables 2, 3): 139 questions required the choice of one of five options, 50 questions required the choice of two options, and 11 questions required the choice of three options. For one question in the morning session, both options 4 and 5 were considered correct, because two options were considered possible

Table 2. Number of questions and number of correct answers by question type for the 2023 National Examination for Speech-Language-Hearing Therapists.

Correct answer rate by question type					
	Number of questions	Chat GPT correct answers	Chat GPT correct answer rate (%)	Our students' correct answers	Our students' correct answer rate (%)
One-answer multiple choice question	137	89	65	85	62
Two-answer multiple choice question	50	37	74	31	62
Three-answer multiple choice questions	11	6	55	7	65
Total	198	132	Mean = 65%	125	Mean = 63%

Table 3. Percentage of correct answers by question area.

Comparison of correct answer rates between ChatGPT and our university students by question field						
	Number of questions	ChatGPT correct answers	(a) ChatGPT correct answer rate (%)	Student's correct answers	(b) Student's correct answer rate (%)	difference(a)-(b)
Dysarthria (child)	4	2	50	3	83	-33
Swallowing/speech disorder	16	12	75	13	79	-4
Aphasia/higher brain function	21	11	52	16	77	-25
Dysarthria (adult)	8	5	63	6	75	-8
Language development disorder	20	9	45	13	66	-21
Introduction to speech and hearing	6	4	57	5	64	-7
Deaf studies	25	17	68	15	60	8
Basic medicine	17	14	82	9	58	24
Development	8	5	63	5	57	6
Social welfare/education	6	2	33	3	57	-24
Clinical Medicine	24	21	88	13	54	34
Sound/Language/Hearing	6	5	83	3	54	19
Speech/Language	20	14	71	10	51	20
Psychology	17	10	63	8	47	16

Figure 1. Percentage of correct answers by our students compared with those of Chat GPT
A. Questions of the morning session (Q1 - Q100), B. Questions of the afternoon session (Q101 to Q200)
The morning and afternoon questions are listed from left to right in descending order of percentage of correct answers by our students. (White bars: questions incorrectly answered by Chat GPT; Black bars: questions correctly answered by Chat GPT)

correct answers based on *a posteriori* evaluation. As mentioned above, two questions required the use of sound spectrum diagrams, so were removed from this analysis for a total of 198 questions.

As in previous years, the passing score was defined as correctly answering at least 120 of 200 questions correctly (≥60%). Across Japan, 1,696 out of 2,515 examinees passed the test, a pass rate of 67.4% across all examinees. Of our 15 students who took the exam, nine got ≥60% of the answers correct, for a pass rate of 60%. Chat GPT correctly answered 131 of 198 questions (66.1%), which is within the passing range.

By question type, the Chat GPT answered 89 of 137 one-answer multiple choice questions correctly (65%), 37 of 50 two-answer multiple choice questions (74%), and 6 of 11 three-answer multiple choice questions (55%). In contrast, our students correctly answered 62% of the one-answer multiple choice questions, 62% of the two-answer multiple choice questions, and 65% of the three-answer multiple choice questions. There was therefore no particular difference in the type of correctly-answered questions between our students and Chat GPT (Table 2).

By field, the highest percentages of correct answers by Chat GPT were in clinical medicine (88%) sound/ language/ hearing (83%), and basic medicine (82%) fields, while the lowest percentages were in the fields of social welfare/ education (33%), language development disorders (45%), and dysarthria (child) (50%) (Table 3). Compared with the University students, Chat GPT had higher percentages of correct answers in areas of clinical medicine (88 vs 54%), sound/language/hearing (83 vs 54%), basic medicine (82 vs 58%), and swallowing/speech disorder (75 vs 79%). Lower percentages for Chat GPT were in areas concerning aphasia/higher brain function (52 vs 77%), dysarthria (child) (50 vs 83%), language de-velopment disorders (45 vs 66%), and social welfare/education (33 vs 57%), in that order.

Based on the answers of our 15 University students, the percentage of correct answers for each question in the two sessions was calculated (Figure 1). The questions with the lowest percentage of correct answers were rearranged to the left and the questions with the highest percentage of correct answers to the right; the questions that Chat GPT answered incorrectly are shown in red. With the questions arranged by difficulty according to the student's answers, Chat GPT did not answer questions in the same way as the students, there were differences in questions that were answered correctly and incorrectly. In some cases, Chat GPT answered incorrectly for both high and low difficulty questions according to the student's results. In other words, the level of difficulty for Chat GPT is different from that of our students.

DISCUSSION

Training for rehabilitation professionals in Japan is based on a curriculum and with practical training in accordance with designated regulations at accredited universities (including junior colleges) and vocational schools. Graduation from an accredited training school (261 schools for physical therapists, 209 schools for occupational therapists, and 74 schools for speech therapists) is required to be eligible for the examination.

A total of 21,179 students took the national examination in February 2023, 17,801 of which passed, for a pass rate of 84%. Tables 4a and 4b show the number of examinees, the number of people who passed, and pass rates for the students nationally and from our university (Table 4). The pass rate for speech-language pathologists was low at 67.4%, compared with 87.4% for physical therapists and 83.8% for occupa-

Table 4. Number of overall examinees, successful examinees, and pass rates for the 2023 National Examination for Rehabilitation Professionals.

2023 National Examination for Rehabilitation Specialists

a

Nationally	Examinees	Passed	Pass rate
Physical Therapist	12,948	11,312	87.40%
Occupational Therapist	5,719	4,793	83.80%
Speech-Language Therapists	2512	1696	67.40%

b

Our students	Examinees	Passed	Pass rate
Physical Therapist	59	48	81.45%
Occupational Therapist	41	34	82.90%
Speech-Language Therapists	16	9	56.30%

tional therapists.

When we tasked Chat GPT with answering the questions of the National Examination for Speech-Language-Hearing Therapists, which had the lowest pass rate this year, Chat GPT answered 131 out of 198 questions correctly, with a 66% correct rate. Chat GPT thus passed this national exam.

The potential of Chat GPT use in the medical education field has been discussed, and the difficulties and moral conundrums associated with using Chat GPT in therapeutic settings have been pointed out, although patient inquiries, note writing, decision making support, trial enrollment, data management, and patient education are the area in medical science that Chat GPT can help with (Garg, 2023). Chat GPT generates largely accurate information to diverse medical queries from 17 specialties, which implies it is capable of doing some of the work of medical staff in helping physicians in clinical settings (Johnson, 2023). According to our results, the use of Chat GPT can also be considered in medical education regarding the diagnosis and treatment of patients.

Chat GPT was used to answer the United States Medical Licensing Examination (USMLE) in 2023, and although it did not pass the exam, it was very close (Kung, 2023). The USMLE consists of Step 1, Step 2CK, and Step 3, and the Chat GPT had a correct response rate of 64.5%, 52.4% and 65.3%, respectively (Kung, 2023). Although the passing range varies from year to year, the average is 55-65%. Although Chat GPT was unable to pass the test, it was able to answer the logical questions correctly to a fairly good extent (Kung, 2023). In another recent paper on USMLE, Chat GPT was able to correctly answer 58.8% of the logical questions and 60% of the ethical questions (Sharma, 2023).

As for Japanese results, in this year's 117th National Medical Practitioners' Examination, the percentage of correct answers attained by Chat GPT was 55% (Kaneda, 2023). The passing criteria for Japan's 117th National Examination for Physicians was ≥80% for the required questions and ≥74.5% for the general and clinical practice questions, so Chat GPT was far short of passing the examination. The difference between USMLE and Japanese Medical Practitioners' Examination might be explained by the fact that Chat GPT accumulates much more information in English than in Japanese. The higher correct answer percentage (74.5-80%) required of Japanese Medical Examination may also explain the lower passing rate of Chat GPT in Japanese Medical Examination. There is a possibility that newer versions of Chat GPT will be able to answer high enough correct answers in Japanese Medical Examination in near future, considering the rapid development of AI technology. For example, Chat GPT-4 (March 14, 2023~) correctly answered 251 of the 396 attempted questions (63.4%), whereas Chat GPT-3.5 (March 15, 2022~) correctly answered 46.3% of 410 attempted questions orthopedic board-style questions, showing GPT-4's significantly greater accuracy than Chat GPT-3.5 (Hoffman, 2023).

In our study, Chat GPT correctly answered 66% of the questions on the 25th (2023) National Examination for Speech-Language-Hearing Therapists, indicating that it would be sufficiently competent to be certified as a speech-language-hearing therapist in Japan. To the best of our knowledge, this is the first paper to report on Chat GPT passing national medical examinations.

The pass rate in FY2023 was 87.4% for physical therapists and 83.4% for occupational therapists. Meanwhile, as mentioned in the introduction section, the 2023 version of the National Examinations for Physical Therapists and Occupational Therapists contained a large number of questions with figures and charts, so we could not analyze all of the questions this time. However, out of interest, we tasked Chat GPT with answering the questions given only in Japanese sentences and which did not require figures or charts. It gave >60% correct answers in both versions of the exams. Chat GPT would therefore likely do well on the National Exams for Physical Therapists and Occupational Therapists as well as the National Examination for Speech-Language-Hearing Therapists.

Chat GPT incorrectly answered 14% of the questions in the morning session that were answered 100% correct by our students. It also incorrectly answered 40% of the afternoon session questions that were answered 90% correctly by our students. Many of the questions that were comparatively easy for our students were thus answered incorrectly. This suggests that Chat GPT still gives incorrect answers for questions that are easy for humans, highlighting some of the issues with AI. The answers given by Chat GPT are sometimes completely different from the facts or irrelevant to the context, so there have been reports of so-called 'AI hallucination.' This is when an AI model generates incorrect information but presents it as if it were a fact. Even in such a qualification test, a final confirmation by a human being is necessary.

In conclusion, this study is the first to show that Chat GPT has the capability to pass a national med-

ical examination. AI is still in relative infancy, but our results show that there is wide potential for academic and education application beyond clerical work. Further studies will seek to elucidate more precise potential use in academic settings.

REFERENCES

Garg RK, Urs VL, et al. Exploring the role of Chat GPT in patient care (diagnosis and Treatment) and medical research: a systematic review. MedarXiv, 2023 doi:10.1101/2023.06.13.23291311

Hofmann HL, Guerra GA, et al. The rapid development of artificial intelligence: GPT-4's performance on orthopedic surgery board questions. Orthopedics 0(0), 1–5, 2023 doi:10.3928/01477447-20230922-05

Johnson D, Goodman R, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model. Res Sq, 2023 doi:10.21203/rs.3.rs-2566942/v1

Kaneda Y, Tanimoto T, et al. Can Chat GPT pass the 2023 Japanese National Medical Licensing Examination? Preprints.org, 2023 doi:10.20944/preprints202303.0191.v1

Kung TH, Cheatham M, et al. Performance of Chat GPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2(2), e0000198, 2023

Sharma P, Thapa K, et al. Performance of Chat GPT on USMLE: unlocking the potential of large language models for AI-assisted medical education. arXiv, 2023 doi:10.48550/arXiv.2307.00112